University of São Paulo – USP
School of Communications and Arts – ECA
Postgraduate Program in Information Science – PPGCI

Erika Alves dos Santos

FRBRizing bibliographic references structures: an approach from in-text reference pointers and SPAR Ontologies for describing bibliographic metadata

Corrected version

São Paulo
Julho de 2021

University of São Paulo – USP
School of Communications and Arts – ECA
Postgraduate Program in Information Science - PPGCI

Erika Alves dos Santos

FRBRizing bibliographic references structures: an approach from in-text reference pointers and SPAR Ontologies for describing bibliographic metadata

Corrected version

Thesis submitted to the University of São Paulo (PPGCI-ECA-USP), as part of the requirements for achieving the PhD degree in Information Science.

Concentration area: Culture and information.

Research line: Organization of Information and Knowledge.

Advisor: Professor Doctor Marcos Luiz Mucheroni (University of São Paulo)

São Paulo
Julho de 2021

Cataloging in Publication

Erika Alves dos Santos – CRB8/7110

Santos, Erika Alves dos.

FRBRizing bibliographic references structures : an approach from in-text references pointers and SPAR Ontologies for describing bibliographic metadata / Erika Alves dos Santos. – Corrected version. – São Paulo, 2018.

308 f.: il. color.

Thesis (Doctorate) – Postgraduate Program in Information Science – School of Communications and Arts – University of São Paulo.

Advisor: Marcos Luiz Mucheroni.

Includes glossary and bibliographic references.

1.Bibliographic references. 2. Bibliographic metadata. 3. Information representation. 3. Citing metadata. FaBiO Ontology. I. Mucheroni, Marcos Luís. II. Título

CDU 020

Erika Alves dos Santos

FRBRizing bibliographic references structures: an approach from in-text references pointers and SPAR Ontologies for describing bibliographic metadata

Corrected version

Thesis submitted to the University of São Paulo (PPGCI-ECA-USP), as part of the requirements for achieving the PhD degree in Information Science.

Concentration area: Culture and information.

Research line: Organization of Information and Knowledge.

Advisor: Professor Doctor Marcos Luiz Mucheroni (University of São Paulo)

Date of the defense: July 6th, 2021

Silvio Peroni
Doctor in Computer Science
University of Bologna

Ricardo César Gonçalves Sant'Ana
Associate professor in Management Information Systems
Doctor in Information Science
São Paulo State University - UNESP

José Eduardo Santarém Segundo
Associate Professor in Information and Technology
Post-Doutor degree, Western's Computer Engineering University/Canadá
Masters and Ph.D. in Information Science
University of São Paulo – USP / São Paulo State University - UNESP

Giovana Deliberali Maimone
Doctor in Information Science
University of São Paulo – USP

Janus

Universidade de São Paulo

ATA DE DEFESA

Aluno: 27151 - 9064227 - 2 / Página 1 de 1

Ata de defesa de Tese do(a) Senhor(a) Erika Alves dos Santos no Programa: Ciência da Informação, do(a) Escola de Comunicações e Artes da Universidade de São Paulo.

Aos 06 dias do mês de julho de 2021, no(a) realizou-se a Defesa da Tese do(a) Senhor(a) Erika Alves dos Santos, apresentada para a obtenção do título de Doutora intitulada:

"FRBRizing bibliographic references structures: an approach from in-text reference pointers and SPAR Ontologies for describing bibliographic metadata"

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Nome dos Participantes da Banca	Função	Sigla da CPG	Resultado
Marcos Luiz Mucheroni	Presidente	ECA - USP	APROVADA
Ricardo Cesar Gonçalves Sant'Ana	Titular	UNESP - Externo	APROVADA
Giovane Deliberali Maimone	Titular	ECA - USP	APROVADA
Jose Eduardo Santarem Segundo	Titular	FFCLRP - USP	APROVADA
Silvio Peroni	Titular	UNIBO - Externo	APROVADA

Resultado Final: APROVADA

Parecer da Comissão Julgadora *

Eu, ______________________________, lavrei a presente ata, que assino juntamente com os(as) Senhores(as). São Paulo, aos 06 dias do mês de julho de 2021.

P/ Ricardo Cesar Gonçalves Sant'Ana

P/ Giovana Deliberali Maimone

P/ Jose Eduardo Santarem Segundo

P/ Silvio Peroni

Marcos Luiz Mucheroni
Presidente da Comissão Julgadora

* Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

A defesa foi homologada pela Comissão de Pós-Graduação em ______________ e, portanto, o(a) aluno(a) ______________ jus ao título de Doutora em Ciências obtido no Programa Ciência da Informação - Área de concentração: Cultura e Informação.

Prof. Dr. Mário Vidoira
Presidente da Comissão de Pós-Graduação

With all my love, I dedicate this work to my greatest loves. My parents, Luiz and Carmem, and my brother Everton.

ACKNOWLEDGEMENTS

To God, who in his infinite greatness and goodness, makes it impossible for me to assign him any adjective, at the risk of committing the unforgivable injustice of reducing him to the poverty of the meaning of a term. God, I will be forever grateful to you, for your love, for the gift of life restarted every dawn, for my intelligence, for my family, for our health and sanity, and for everything else that was freely attributed to me, even though I am so small and flawed, in view of your greatness. For everything, thank you!

To my parents, Luiz and Carmem, and my brother Everton, my foundations, I thank you for being part of my life and for being always by my side, holding my hand, by choice, even during the most severe storms. I am forever indebted to you for giving me the opportunities and experiences that have made me who I am. Accept this homage as a simple demonstration of all the love that you awaken in me and a simple "retribution" for the long months deprived of our coexistence in favor of the construction of these writings. I Love you all so much!

To my aunt Maria das Graças and my uncle Chico (*in memoriam*), I think of you as my second parents. Thanks for being what you are for me. I love you!

To my advisor, Marcos Luiz Mucheroni, who took me carefully along the paths of knowledge, demanding from me what I did not even think I could give. To you, Professor, to whom I dedicate great admiration and respect, thank you very much for your enthusiasm for the project, for your support, for your friendship, encouragement and patience!

To Professor Silvio Peroni, who informally co-supervised this work, I extend my deep thanks for the partnership, for the friendship, for the competence materialized in the suggestions incorporated into this work and for having believed in my potential, even before meeting me. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. For everything, thank you very much!

To Professors Silvio Peroni, Ricardo César Gonçalves Sant'Ana, José Eduardo Santarém Segundo and Giovana Deliberali Maimone for your dedication on examining this study and for your valuable contributions.

To Alessandra Rodrigues and Lianna Borelli, my dear friends who accompanied each leg of this journey, thank you for always being there!

To Tarcísio Buschinelli, thanks for your friendship and care.

To my sister-in-law Cristiane Peixoto, for your friendship and support.

To my coworker (and current immediate boss) Karina Penariol Sanchez for your support, friendship and patience along this incredible journey.

To Fundacentro for giving me the opportunity of being a doctoral student, and for all the support and structure along this 16 year of my professional journey, which provided me the professional development and growth which made all this possible, and for all my coworkers, thank you!

To the School of Communications and Arts of University of São Paulo and all my PhD colleagues for the stimulating discussions, and for all the fun we have had in the last four years and a half. Also, I thank my friends in the University of Bologna, in particular, Marilena Daquino for your invaluable advice and feedback on my research and for always being so supportive during my stay in Italy.

To the Department of Classical Philology and Italian Studies of the Università di Bologna, to its kind staff, and to my doctoral colleagues, thanks for accepting me as a visiting doctoral student for one year!

To the Library Service of the School of Communication and Arts (ECA) of University of São Paulo and the Library of the Department of Classical Philology and Italian Studies of University of Bologna for the contribution in providing part of the articles that composed the sample of this study.

Finally, I could not have completed this thesis without the support of my family, my aunts and uncles, my cousins, and my godchildren. My deep and sincere gratitude to my whole family for their continuous and unparalleled love, help and support.

For all my friends, specially Diogo Galhardo, Andrea and Georgina Mustafa, Wanderley Santos, Thais Pimenta, Veronica Bahdour, Ione and Ivone Carvalho, Carolina Menez, Maria Laura Bolognesi, Viviana and Maurício, with whom I have shared moments of deep anxiety but also of big excitement. Their presence was very important in a process that is often felt as tremendously solitaire. Thank you all, guys! (sorry if I forgot someone, you are so many!)

I saw that one enquiry only gave occasion to another, that book referred to book, that to search was not always to find, and to find was not always to be informed.
Samuel Johnson (1709-1784), 1753

ABSTRACT

Descriptive representation is passing through a huge conceptual restructuring whose first main results materialized in the publication of the Functional Requirements for Bibliographic Records. From this perspective, citing and referencing metadata, which were considered as facets of descriptive representation, were approached towards a reflection on what is, and what are the trends on the role of bibliographic references and citation metadata in the citation network scenario and what are the possible impacts of these changes in bibliographic metadata representation and citation network. The main objectives of this study are to discuss the role assumed by bibliographic references in the citation network, considering them as a facet of descriptive representation and point the possible impacts to be expected in bibliographic metadata representation and citation network fields, considering the descriptive representation restructuring. Starting from this, the usage of the FaBiO Ontology on describing bibliographic metadata is pointed as a facilitating instrument for bibliographic metadata management in the context of the elaboration of bibliographic references. The method involved qualitative and quantitative analysis of citing and referencing elements (i.e., bibliographic references, mentions, quotations, and respective in-text reference pointers) identified citing and referencing habits within disciplines considered in the SCImago Journal & Country Rank and errors occurring during the transcription of citing and referencing metadata withing scientific articles over the long term, as stated by previous studies now expanded. Future expected trends of information retrieval from bibliographic metadata were gathered by approaching these referencing elements from the FRBR Entities concepts and the use of semantic resources, specifically the use of FaBiO Ontology on bibliographic metadata processing as an instrument of strengthening the citation network and an enabler of the automatic gathering of bibliographic metadata by reference managers, as a means of saving researcher's time on normalization tasks, in addition to the provision of more clear, efficient, and consistent bibliographic metadata within bibliographic references. The findings showed that reference styles do not fully accomplish with their role of guiding authors and publishers on providing concise and well-structured bibliographic metadata within bibliographic references. The analysis of the dynamics of the relation between mentions, quotations and its respective bibliographic references showed, first, long stand errors in citing and referencing metadata, which contribute to the lack of standardization (or is worsened by it). Second, Reference styles do not provide broad and clear coverage of all aspects concerning bibliographic metadata description in the form of bibliographic references. Third, publishers do not demonstrate great efforts to improve bibliographic metadata included in the articles' bibliographic references lists. The improvement on the way which bibliographic metadata is provided within bibliographic references depends on a joint effort between the Scientific community (here understood as the authors and the publishers – including all the agents involved with the editorial process), besides the Information Science and Computer Science community.

Keywords: Bibliographic references; Information representation; Bibliographic metadata; Citing metadata; FaBiO Ontology.

LIST OF ILLUSTRATIONS

GLOSSARY

LIST OF CHARTS

LIST OF GRAPHICS

CONTENTS

1. THE BACKGROUND

Science is a collective construction that results from an empirical discover. This represents three characteristics of the scientific communication that also applies to scientific information that is to be fragmented, derivative and editable. Fragmented, in the sense of when each scientist produces a fragment of knowledge, a piece of the puzzle of Science is unveiled. Derivative, in the sense that one research is based on the previous ones, be for supporting, proofing, or refuting. Editable, in the sense that an author's considerations are submitted to the opinion (evaluation) of editors and of the scientific community (ZIMAN, 1979; RODRIGUES; LIMA; GARCIA, 1998). The combination of these three characteristics, is responsible for lapidating of scientific knowledge and Science advances.

In this context, no researcher is supposed (or ethically allowed) to make a statement within the scientific contextualization, that is, the theoretical foundation that configures the background for the construction of new ideas and discoveries. But maybe there is one exception in this context that allows anyone to affirm, without a previous scientific grounding, that the way of producing, storing, representing, retrieving, and using (scientific) information has changed along the years and, that this transformation occurred quite fast. Each two days it is created the same volume of information that was ever created until the year of 2003. 90% of the total amount of existing data in the world in 2013 has been generated along the two predecessor years (EIS, 2016).

Such scenery, added to technological advances, demanded transformations in the way that information[1] is produced, stored, retrieved, and used, especially considering the electronic information approach. The management and manipulation of all these data also had to get adapted to this, to make it possible to all this amount of existing and newborn information to be (and stay) searchable and retrievable. This is the point where Information Science (IS) and Computer Science get an intersection as complimentary disciplines to manage and make all this information retrievable somehow.

[1] It is not the aim of this discussion to point out a definition of what is information, considering that this term itself may denote a range of significances that would be subject for a whole encyclopedia. So, in the present context, information is considered all intellectual production made available by scholars and the whole scientific community. Please find the detailed definition in the glossary attached.

Considering the Information Science approach, it is important to keep in mind that although information corresponds to an immaterial concept, it is richly endowed with meaning and context. And this is something to really care about because the value and application of information may completely change depending on how these two aspects are dealt with. That is to say that individualizing a FRBR work, a FRBR manifestation or even an FRBR item, aiming to make it (and mainly its content[2]) searchable and retrievable in multiple ways of access, through structured metadata is one of the Information Science concerns which also coincide with one of descriptive representation's purposes. Lancaster (2004) defines descriptive representation as a process in which it is identified authors, titles, sources, and other bibliographic elements.

Providing bibliographic information to make documents searchable and retrievable through several access paths using structured metadata corresponds to one of the purposes of descriptive representation. Descriptive representation is a process in which we identify data related to the editorial production of documents, such as their authors, titles, publishers, years of publication, number of pages, sources, and other bibliographic elements (GALVÃO, 1998; LANCASTER, 2004). Although descriptive representation is still considered the physical description in manual catalogs (MAIMONE *et al.*, 2011), it is also a way to materialize and make evident information's meaning, context, and relationships (LANCASTER, 2004) between documents. One of the products of this type of information representation is the bibliographic reference (GALVÃO, 1998), considered by Baptista (2007) as the elaboration of records containing the descriptive representation.

This fosters the understanding that representing information is more than just converting an abstract tool to metadata. It is also a way of somehow, materialize and make evident its meaning, context, and relationships, in some cases. This background demands Information Science's efforts on representing information in a trusty way regarding to the content represented, be it a conceptual instance (or FRBR work, which is the information itself) or a physical FRBR item (e.g., a single hardcopy of a book), what is called descriptive representation of information. This activity consists of

[2] and here begins the paradigm of what exactly must researchers refers to in their bibliographic references: the work or the expression?

a decision process in which the researcher, or another professional who are running it (usually a librarian), defines and register a metadata pack to compose a bibliographic record (or reference) to describe an information or the physical characteristics of a document on the most fidelity way as possible. Descriptive representation may be presented in the form of a bibliographic record, which is more commonly seen in library catalogs or, in the form of bibliographic references, which are more commonly used at the end of scientific texts. In a FRBRized approach (although here referring particularly to bibliographic references), representative description shares some generic tasks that are performed by users when searching and making use of bibliographic catalogs, once it (also) aims to identify a resource, an agent (e.g., an author, publisher, translator) or both. In some cases, it also allows the reader to select a resource according to its content (whether it attends to one's information needs or not), or even obtain access to the entity described.

That is to say that representative description in the form of bibliographic references, allows the reader to

a) identify all the background information that have supported an author's ideas conception, or the information which reading is being recommended by him as a complimentary content of his own ideas;
b) use the data contained in the bibliographic reference to select an entity that is appropriate to one's needs;
c) find and obtain access to the entities described (IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS, 2009).

By referring to "other bibliographic elements", Lancaster's definition to descriptive representation suggests a wider approach to it once the level of accuracy adopted in a description of a FRBR work or an entity may correspond to another facet of descriptive representation: the bibliographic reference. Baptista (2007) defines a bibliographic reference as the elaboration of records that **content the descriptive representation** of any supports, and complements that it begins to be made by different professionals, according to their own interests, and from multiple locations: scientists, artists; enterprises; negotiators; publishers; libraries; archives; museums, etc. This represents (or at least it should) a factor of concern to Information Science. Once bibliographic references writing is being carried out by professionals from other areas than Information Science, and being this an important tool for establishing links among scientific FRBR Works, which also supports the citations networks work, this scenery leads to three questions: do these professionals (researchers) have the required skills or even the access to the guidelines to proceed this? Are bibliographic styles accessible and written in a clear and comprehensive language? And, finally, what should be the Information Science's posture facing this scenery?

Ziman's (1979), and Rodrigues, Lima and Garcia's (1998) discourse on human knowledge construction and development, suggests that knowledge is a cooperative construction based on the interaction and experiences sharing among scholars, which is the input for stoning and advance knowledge. This interchange occurs in a conceptual sphere of scientific world and materializes in the discussions lead in scientific events and FRBR works, through citations and bibliographic references. Such scenery constitutes the background for citations networks work that makes information string attached in a conceptual approach. In such context, the proper description and links between mentions, quotations and bibliographic references may be considered as supporting instruments for citations networks work, once it favors the identification and consequently, the access to other related FRBR works according to the researcher's needs. This understanding evidence the importance of providing metadata on the most faithful and consistent and complete way, to provide all the information requested by the researcher while retrieving the information of interest. This statement may instigate questions on: what is the most appropriate bibliographic reference structure and what kind of metadata must be registered to favor information retrieval?

At last, this briefly introduction leads to the following considerations:

a) descriptive representation is a tool that attributes the characteristic of being searchable and retrievable to information and which has at least two facets: cataloging and bibliographic references writing;
b) citations networks are formed by conceptual link that attaches relationships and interconnections among scientific contents. These links are indirectly evidenced by mentions, quotations, in-text references pointers and their respective bibliographic references registered in scientific publications;

c) since descriptive representation is a decision process and once this process is being leaded by other professionals than librarians, it is necessary to make clear all the definitions, objectives and requirements for writing bibliographic references, including the ideal bibliographic reference structure.

Parallel to that, Information Science has been carrying out researches and discussions in order to adapt descriptive representation to the various ways in which information is produced and made available nowadays. Considering that the democratization of technological resources has extended the possibility of publishing to anyone who has access to an electronic device connected to the Internet, blogs, open repositories and softwares are some examples of "recent" tools that unlinked the obligation of a publishing company to make public the record of the intellectual and/or artistic production produced by any person. This was one of the factors that leaded to the conception FRBR, whose guidelines, in a very general approach, recommends describing the information itself (content and meaning approach), and in this context, it is formed a cycle, where the author is considered an entity, which produces an information (FRBR Work), that may be published in many ways (FRBR Expression): press, electronically, audio, video and so on, and each of these supports of information (FRBR Manifestation) may be exemplified and distributed all over the world (FRBR Item). In this context, an author's ideas (intellectual or artistic production) are comprised as subjective issues which correspond to the main essence of a bibliographic record. In this context the physical support (media) that holds the knowledge registry is admitted as a secondary instance (entity), while the FRBR Work itself is the protagonist.

In complementation, bibliographic references should be recognized as a facet of descriptive representation (which aims to describe documents in an unmistakable way). Besides, once the new approach of descriptive representation, according to FRBR guidelines, tends to consider not the document (FRBR Manifestation and FRBR Item) itself, but a FRBR Work content, the actual relationship among mentions, quotations and bibliographic references seems to get into a collision course. While mentions and quotations generally refers to and represents an immaterial concept, that is, the idea of an author (FRBR Work), bibliographic references instead, generally focus on FRBR Items description, that is, the physical registry of information, which is a single and physical exemplar of a publication (FRBR Item). Whereas the same FRBR Work can be (and actually is) published in different media and formats and also, that most of scientific production is available at Internet (freely available sometimes), describing a specific journal issue or volume, book edition or any other FRBR Item (media) in a bibliographic reference can hinder or configure an impediment to access information. Considering a FRBR perspective, in-text citations generally refer to the entity (FRBR Work) while bibliographic references commonly refer to its embodiment (FRBR Manifestation or FRBR Item).

Reflections on these issues are aligned to Information Science's accomplishment of its mission, that goes back to the Five Laws of the Library Science proposed by the Indian mathematician and library science scholar Sr. Shiyali Ramainrita Ranganathan (1892–1972) in 1931, whose first, third and fourth Laws are directly connected with citation matters:

1. Books are for use;
2. Every reader, his/her book;
3. Every book, its reader;
4. Save the time of the reader;
5. The Library is a growing organism (ZABEL; RIMLAND, 2007, p. 24).

Referring to the first Law, Zabel and Rimland (2007, p. 24) mention that "library materials and services should be accessible to users […] this law might seem trivial in our day and age". However, "The digital age does have new repercussions for this first law" and so, providing electronic access is not always the only way to go for better access (ZABEL; RIMLAND, 2007, p. 24). Therefore, when a user asks a reference question, he usually has already proceeded with any kind of previous (and unsuccessful) search on his own. Besides, the access to scientific publications usually demands financial investments from the reader, or from the institution he might be linked to, and this boosts the publisher's responsibility on providing accurate bibliographic metadata, as a demonstration of their commitment and respect for the reader, by saving their time and efforts while retrieving such information.

From a bibliographic reference's perspective, these arguments support the understanding that chronological, geographical, linguistic, and technological aspects should never configure issues to information identification and retrieval. Ways of accessing information resources should be guaranteed to users (readers and researchers included) and since this requires the proper identification of a particular publication, bibliographic references should provide users with enough metadata which allow them to quickly, clearly, and unambiguously identify a publication and retrieve it by their own (Ranganathan’s fourth Law), expanding its access pathways (Ranganathan’s third Law) regardless the means of accessing it (electronic or not), or the information support it has been published or stored (Ranganathan’s first Law).

Referencing and citing issues are complex, especially considering the current descriptive representation’s revision context. Referencing and citing are huge and multifaceted activities, and, because of this, this study is not intended to fully deal with all aspects involved in these matters and, should, therefore, be taken as a starting point for a major discussion on ways of becoming citing and referencing tasks less laborious and time-consuming.

1.1. The content organization

The first chapter presents a brief contextualization of the theme within Information Science (IS) domains. Specifically, an overview on information representation places the approaches on citing and referencing metadata within such universe and suggests the respective connections with the FRBR context.

Chapter 2 addresses the research’s problematization and evinces the research object of this study. Given the complexity of the theme and to evince each of the questions this research aims to answer, the investigated problem is represented by 9 research questions (RQ) which are further addressed along the discussion, both to evince the findings concerning each question and to evince the discussions partially or fully answering them.

Chapter 3 provide the justificative supporting this study and evinces the importance of addressing the issues considered in this study. The wide range of reference styles and the obstacles they represent to researchers and to the processing of bibliographic metadata in a standardized way are highlighted as encouraging points for the joint efforts of Information Science and Computer Science towards the management of bibliographic metadata, specially considering citing and referencing metadata.

Chapter 4 provide a brief overview on ambiguities involving the term information, since it is the main object of descriptive representation. Next, the discussion advances towards to information approaches from the descriptive representation perspective and introduces the topic approached in the following chapter: the metadata.

Chapter 5 discusses the role played by metadata within Information Science universe considering its main facets and typology matters.

Chapter 6 deals with the methodological procedures adopted both for data gathering and for data analysis. This chapter is divided according to the phases in which this research was proceeded. First, it considers the methods for selecting journals for sample composition. Next, the selection procedures for issues and articles are explained. In sequence, the methods used for analyzing articles, bibliographic references and in-text reference pointers are fully explained.

Chapter 7 is dedicated to report the findings of the study. This chapter is divided into three parts. The first part provides the results of the analysis considering the analysis of data collected from the Medicine and Social Sciences journals and articles sample. The second part expands the results of the analysis proceeded with the Medicine and Social Sciences to the 27 subject areas considered by SCImago database. As for the third part provides the findings on the approaches on the descriptive elements considered within citing and referencing metadata analysis (i.e., bibliographic references, mentions, quotations and respective in-text reference pointers), which determined the most used bibliographic elements within bibliographic references across subject areas for different types of publications. Chapter 7 ends with the findings from the data cross-checking between the most used bibliographic elements within bibliographic references and the FaBiO Ontology, to detect its suitability on describing such bibliographic metadata.

Chapter 8 contains the discussions concerning the answers to the RQs 1 up to 9 based on the analysis of the findings provided in Chapter 7. The content is also divided into three parts which are directly connected to the subdivision in which Chapter 7 is presented. Therefore, the findings showed in Session I in Chapter 7 are discussed in the Session I in Chapter 8. The same logic applies to the Sessions II and III in both Chapters.

Chapter 9 provides a broad overview of the study and suggests further ramifications.

Lastly, all cited works are indicated in the bibliographic references list, composing the Chapter 10, which is followed by Chapter 11, which contains a glossary of terms used in this research. The same content also can be found at Santos, Peroni and Mucheroni (2020a).

2. THE RESEARCH PROBLEM

Citing and referencing are probably one of the less appreciated tasks among scientists. And not without reason, considering the huge volume of data they are expected to deal with to produce a single piece of information or, an improvement of a previous concept, facing the also huge number of rules they are expected to observe to make their publications in an "scientifically acceptable and publishable format". Searching, selecting, gathering, reading, and writing are, therefore, everyday tasks of researchers and so, for all of this to be possible, the representation of the information is a mandatory task for everyone involved in the information production process.

In this scenery, the posture, the form of action and the level of involvement of authors and publishers may be different considering, overall, the degree of automation of information collection and standardization tasks. But the focal point of this discussion on this panorama is, to what extent the descriptive representation is human readable, considering bibliographic standardization aspects. And, more than that, what can be done to assist, mainly authors, dealing with the tasks involving bibliographic issues, like bibliographic seeking and normalization aiming to allow the researchers to use the time which would be dedicate to the standardization tasks in the effective production, improvement, and registration of knowledge, instead. So, efforts should be devoted on finding out the best way to apply the technological resources to make machines able to deliver the same bibliographic products (e.g., a bibliographic reference) as they should be delivered by a human, or something as close as possible to it, mainly concerning to the bibliographic metadata normalization matters, which are scientific tasks which demand (much) time and specific technical skills to be perfectly carried by the researcher.

Bibliographic references are a facet of descriptive representation. They correspond to a specific kind of description that acts like a reference that points the reader to an original source of information cited by an author in the text body of a work. Each bibliographic reference contains the textual representation of a minimal set of descriptive bibliographic metadata which enables the identification of a publication, a speech, a piece of information, or anything else that may be citable, to locate and retrieve it (ABNT, 2018; CUNHA; CAVALCANTI, 2008; ISO, 2010).

Baptista (2007) argues that bibliographic references are now being written by different professionals, according to their interests – scientists, artists, enterprises, negotiators, publishers, libraries, archives, museums, etc. – coming from multiple locations. Bibliographic references are one of the tools that establish links between scholarly works and enable the creation of citation networks. Writing bibliographic references correctly demands previous background and some Information Science skills. It requires that both reference and citation styles guidelines are presented in comprehensive language, with a wide scope of the bibliographic universe, especially considering Baptista's statement. Information Science should have a primary role in this activity. However, since researchers in all the disciplines usually deal directly with this aspect without involving experts, one can introduce mistakes in the bibliographic references that prevent the clearly and unambiguously identification of the represented (i.e., cited) works.

Citation error is an old issue from the XVII Century (SWEETLAND, 1989), and started before the publication of the first reference style manuals – the Hart's Rules for Compositors and Readers (1893) (HART, 1893) and the University of Chicago Manual of Style (1906) (O'CONNOR, 1977). Despite the publication of those style manuals, "citation errors continued to appear, as did an increasing number of complaints about them" (SWEETLAND, 1989, p. 293).

A study carried out by Sweetland (1989) highlighted the functions of bibliographic references and style manuals and the errors in the reference lists and in-text citations, that represent a crucial issue for the accomplishment of the citations' functions. Sweetland's findings pointed that the great variety of formats for referencing to cited articles added to the lack of agreement among journals or authors increases the chances of misunderstanding referencing guidelines which, consequently, contributes to the high rates of errors in bibliographic metadata description. The study also pointed other causes for bibliographic errors: the lack of commitment of publishers with the normalization of citation metadata, the diffusion of responsibility in the publishing process, the lack of training in the norms and purposes of the bibliographic citation, the misleading of citation rules (i.e., the reference styles' contents), the misunderstanding of foreign languages, the human inabilities to correctly reproduce long information strings and, the failure to examine the cited documents as addressed in Chart 1.

Chart 1 - The main citing and referencing errors pointed by Sweetland' study (1989)

Core citing and referencing errors pointed by Sweetland	**Ramifications of core errors pointed by Sweetland**
Lack of standardization in citation formats	• While based on particular standard formats, multiple versions of references styles seem to have made little improvement in citation accuracy. • Errors are made in the first place because the "complexity" or lack of standardization. Given the variety of formats of citation and the lack of any real agreement among journals or authors, the chance of misunderstandings is high. • Lack of standardization, noting the variants in authors names; spelling tendency to invert vowels and number pairs; miscited page numbers, incorrect and misleading journal titles and wildly misspelled author's names, incorrect and incomplete citations; lacking or incorrect work's titles, use of the same abbreviations to refer to two different journals; cite only one reference when two or more were listed under a single number in the cited article, miscopy numbers. • Complaints about lack of uniformity among authors or librarians.
Diffusion of responsibility in the publishing process	• The role of citations is not taken very seriously by the scientific community. • Errors are not discovered and corrected before publication.
General lack of training in the norms of citation	• General human inabilities to reproduce long strings of information correctly. • Tendency to invert vowels and number pairs. • Misunderstanding of foreign languages.
Failure to examine the document cited	• Citers have not actually seen the original work. • Dishonesty: fake experimental data or references to inexistent papers.

Although Sweetland's study was published decades ago, his arguments remain updated and most of his conclusions are remarkably like the findings of our study suggesting that neither the publication of reference styles nor the technological advances were effective in solving those bibliographic issues.

Contextualizing and supporting the tracking of historical approaches is one of the functions played by citations (KARCHER; ZUMSTEIN, ca. 2018). As regulation and mediation instruments between citations and readers, reference styles directly influence the way we cite and read, and indirectly the way we follow back on author's thinking and research. An appropriate use of reference styles grants the accomplishment of these purposes by providing clear guidelines under which cited works should be formatted to be correctly retrieved.

This study also approaches on bibliographic references and the other contextual entities – namely mentions, quotations, and in-text reference pointers, summarized in Figures 1 and 2 – from several perspectives.

Figure 1 - Visual representation of the main bibliographic elements involved in citing and referencing author-year date system – i.e., when in-text reference pointers referring to mentions and quotations include author's surnames and year of publication of the cited works described in a particular bibliographic reference

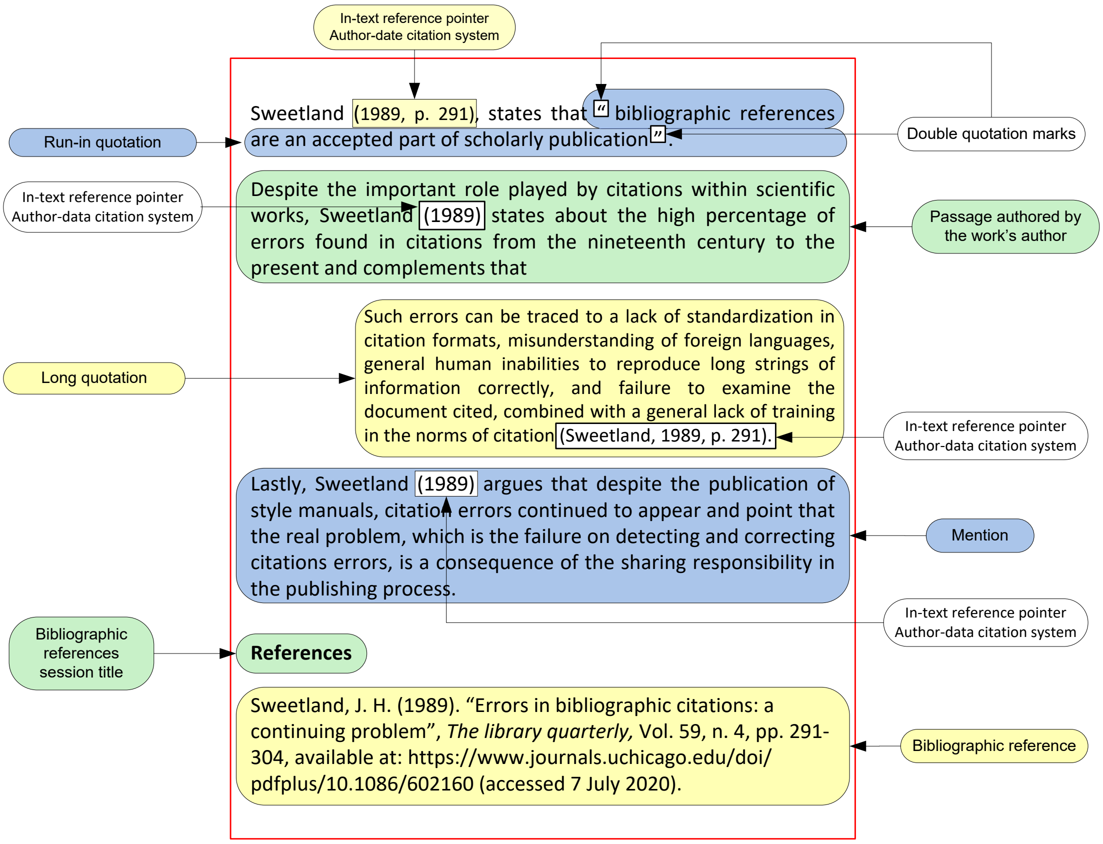
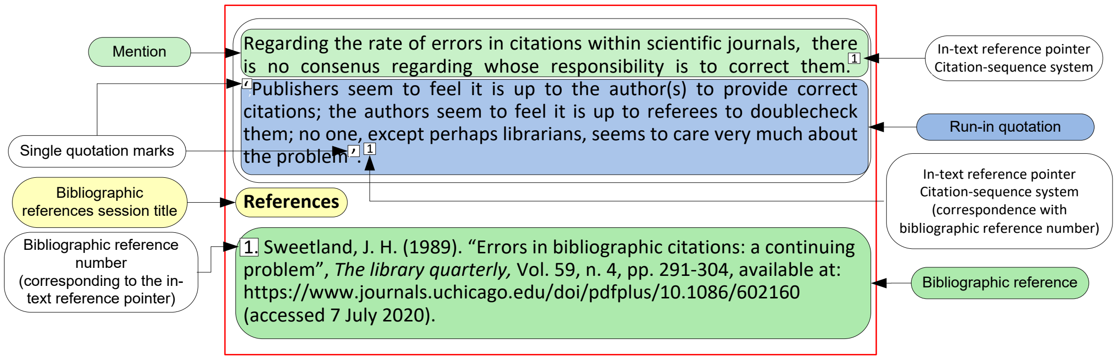


Figure 2 - Visual representation of the main bibliographic elements involved in citing and referencing citation-sequence system – i.e., when in-text reference pointers referring to mentions and quotations are specified using a number corresponding to a particular bibliographic reference in a bibliographic references list arranged in ascending numerical order



Starting from the premise that the problems referred by Sweetland still stand, the purpose of this study is to reflect on how recent journal articles published worldwide in all the subject categories of two main academic disciplines (i.e., the Medicine and Social Science) organize their citation apparatus. We focus on how bibliographic references are defined in published articles and the way they are denoted within the article text through the related in-text reference pointers to support the author's argumentation.

This approach leads towards a reflection on what is, and what will be the role of bibliographic references in the citation network scenario, considering this as a facet of descriptive representation, which is passing through a huge conceptual restructuring, and what are the possible impacts of these changes in bibliographic metadata representation and citation network. This problem may be represented in nine main questions:

- RQ 1 – Considering current bibliographic tools like reference styles and reference manager softwares, were they effective on fully addressing the issues pointed out by the study made by Sweetland in 1989?
- RQ 2 – Are there other possible causes for errors in citing and referencing other than those specified by Sweetland?

- RQ 3 – What impacts are to be expected by readers on retrieving information from citing and referencing metadata, considering the current descriptive representation revision and the potential differences between the level of description adopted by bibliographic catalogs and bibliographic references?
- RQ 4 – Considering the changes in the production, storage, retrieval, and use in the information's universe, do Sweetland's (1989) claims remain updated?
- RQ 5 – Are there current possible causes for citing and referencing errors other than those pointed by Sweetland's study?
- RQ 6 – Which are the possible impacts of the FRBR approach regarding descriptive representation on information retrieval from citing and referencing metadata perspective?
- RQ 7 – Which are the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations, and bibliographic references, considering different types of cited works?
- RQ 8 – Is there a common metadata set used across the disciplines for describing cited works within bibliographic references?
- RQ 9 – Do SPAR Ontologies comply with bibliographic elements composing bibliographic references, particularly the "starred metadata set"?

Considering that Science is guided by empirical processes and that, therefore, the citations network that links related works and, in a way, the logical reasonings that support the construction of scientific thinking in each of the areas of knowledge, our research assessed how clear the relationship between citations and bibliographic references is, from the reader's point of view. The discussion that follows is based on four pillars: first, the identification of the types of publications cited by scientific articles published in 2019, second, the way in which the metadata representing such publications are presented to readers, third, how the relationship among bibliographic metadata elements (i.e., in-text reference pointers referring to mentions and quotations and bibliographic references) takes place, considering the citation and references habits and forth, how the technology can help in this process.

2.1.The premises

The initial premises of the study is that the relations between citing statements (i.e., mentions and quotations) and bibliographic references not are not always clear to readers and not always it is possible to properly identify the works cited by an article, from the bibliographic metadata provided in the bibliographic reference writing. This configures an impediment to the accomplishment of one of the main purposes of referencing a particular publication which, in the ambit of this study, is understood as the accurate identification of the cited publication, from the metadata provided in the bibliographic references, without the need to consult external information sources. The importance of such issues in the scientific environment and, the responsibility attributed (mainly) to the authors in ensuring that such relations would happen in a clear and harmonious way, represent implicit challenges to Information Science, which demand the proposition of solutions that favor, on the one hand, the descriptive representation in its clearest and most objective form possible, allowing the identification and supporting the subsequent search and retrieval of information and, on the other hand, the guidance and support to authors, so that they can properly manipulate and record bibliographic metadata in order to favor the fulfillment of the purpose for which they were conceived and registered in a particular work. The eventual confirmation of this premise suggests that the correspondence among mentions, quotations and bibliographic references might not be clear among all the agents involved in the publishing processes, including the authors, or might be being neglected by them or, might not be clearly described in the normative instruments that guide the scientific publications, reference styles included.

The second premise of this study suggests that the Computer Science is no longer a discipline out of the Information Science context. On the contrary, considering the increasing volume of information and respective metadata, the need to properly register and establish relations among them to assure their retrieval in the proper time, context and space has become a considerable challenge to Information Science, which becomes more evident when factors like the variety of formats and supports (embodiments) in which information may be registered. In this context it is possible that Information Science is no longer the only discipline required to manage information and respective metadata and considering this, the second premise of this study suggests that Computer Science might be the discipline which can offer the missing complements to Information Science to better manage information organization in the current scenario, especially considering the need to establish relations between bibliographic entities to whom may be attributed several embodiments of the same information, for example. The main premise in this context indicates that Computer Science not only may contribute to Information Science in this area but, already has developed tools in ambit, like the FaBiO Ontology, which is probably an efficient instrument for describing bibliographic metadata in the semantic context, which can properly be considered by Information Science.

3. THE OBJECTIVES

Facing the importance of bibliographic references to the development of Science and the conceptual restructuring which descriptive representation has been submitted, the main objectives of this study is to discuss the role assumed by bibliographic references in the citation network, considering them as a facet of descriptive representation and point the possible impacts to be expected in bibliographic metadata representation and citation network fields, considering the descriptive representation restructuring. Starting from this, it is aimed to suggest the application of the FaBiO Ontology as a facilitating instrument for bibliographic metadata management in the context of the elaboration of bibliographic references.

3.1. The specific objectives

The specific objectives of this study are:

a) To verify whether the issues pointed out by the study made by Sweetland in 1989 still remain updated in nowadays scientific articles;
b) To verify whether there are other favoring factors for the occurrence of errors in bibliographic references, beyond those suggested by Sweetland in his study from 1989;
c) To identify the possible impacts of the FRBR approach regarding descriptive representation on information retrieval from citing and referencing metadata perspective;
d) To identify the descriptive elements which are most commonly used across the disciplines, considering the description of the same types of publications;
e) To verify the compatibility of FaBiO Ontology on describing the most commonly used descriptive elements considered across the disciplines, considering the description of the same types of publications.

4. THE JUSTIFICATION

In a very informal context, it is common to hear researcher's complaints regarding to normalization of quotations and bibliographic references. And it is hard to agree that they have concrete reasons for that. There are around 9000 bibliographic styles all over the world and, although around 7500 of them are derived from the same main sources (KARCHER; ZUMSTEIN, ca. 2018), each one has its own particularities and differences. On the other hand, editors are used to request the application of their own bibliographic styles' guidelines in their editorial production. Usually, they do not provide an expert exclusively dedicated to proceeding with a critical standardization review on the submitted papers and, therefore, such responsibility rests with the authors.

Since researchers may submit papers to different publishers simultaneously, they may be required to deal with several bibliographic styles at the same time. Notwithstanding, researchers could be devoted to other scientific productions at the time they are dedicated to bibliographic normalization issues (which, usually takes a long time). So, having this wide range of bibliographic styles and, assigning (only) to researchers the responsibility of applying them to their own bibliographic productions may not be the most efficient way to fill their intellectual production time.

The dynamism in which things happen in the universe of information may difficult some processes, and the bibliographic normalization is one of them, indeed. Once bibliographic normalization is being presented as a facet of descriptive representation, it is inevitable to consider the adaptations that have been carried out in the ambit of cataloging, what includes the conception of FRBR which started a series of other adaptations to better suit to the information dynamics in the actual technological scenery. The understanding that changes on the ways of producing and accessing information has a direct impact on the forms of representing data, is something to be considered also by the normalization universe, even to understand the dynamics of these changes and the possible impacts and necessities of adaptations from the bibliographic references view.

Such argumentation partially justified the execution of this research that certainly will not solve the whole issues concerning the use of quotations and bibliographic references in scientific documents (and these are not even our pretensions, in fact), but also may represent a starting point to a discussion that could already have been carried out by Information Science, especially considering that Computer Science has already been concerned about metadata interchange regarding scientific information and respective metadata management and information access fields.

Since this subject is not an exclusive matter of Computer Science, so that It is essential that Information Science takes its place in these discussions, to find the best possible way of managing and make scientific information (and metadata) strings more consistent, evident and understandable by researchers in general.

The discussions supporting the answers to the research questions are based on the FRBR principles – and not the conceptual models that followed it, including IFLA LRM – as it is considered the materialization of the initial framework of the solutions proposed by Information Science for the adequacy of the descriptive representation and the management of bibliographic metadata to the current scientific context in which the diversification in the forms of ideological expression, communication and use of scientific information have taken on quite expressive proportions, which requires adjustments in the way of managing and offering such bibliographic products to the scientific community, in order to meet their needs and expectations in full. In addition, the SPAR Ontologies are based on FRBR model.

The citations network is the instrument that links scientific thinking, forming a chain of documents related to each other and highlight the research trends in each discipline. As being an abstract concept, the citation network is noticed in scientific publishing by mentions, quotations and bibliographic references and, therefore, it is important that such elements are clearly identified and properly related to each other. Starting from those aspects, the discussion considered them through the observation of what are the bibliographic elements mostly used for identifying citations and for describing bibliographic references and how they relate to each other, seeking to identify whether such connections are clear and evident in the analyzed texts. The results of this specific part of the research evinced the importance of providing clear bibliographic metadata in scientific works, so that the citation networks can satisfactorily and efficiently contribute to the intellectual exchange among researchers.

Such results pointed that the presentation of bibliographic elements identifying mentions, quotations and bibliographic references within the articles composing our sample can be improved. This evinced first, that authors and publishers could be more careful on normalization issues and, second, that technology has a lot to contribute on this matter. So, it is suggested the use of SPAR Ontologies, particularly the FaBiO Ontology, for describing bibliographic elements referring to citable contents, in order to take advantage of the semantic web resources to strengthen the citation networks links and support authors dealing with such bibliographic elements and present them within their works in a more clear and effective way.

5. CONSIDERATIONS ON THE TERMINOLOGICAL CHARACTERIZATION OF INFORMATION IN THE CONTEXT OF DESCRIPTIVE REPRESENTATION

Conceptual reflections on “what is information?” are countless. This is a concept so richly endowed with semantic, taxonomic, etymological, and epistemological aspects, among several others, that makes it impossible to establish a universal concept. It is well known that the definitions are not true or false but, more, or less productive. In a way, people are free to define terms as they want, but in reality, [these definitions can present] problems (CAPURRO; HJØRLAND, 2007).

Despite information has been traditionally considered as the main object of study of the knowledge area which not coincidentally also bear its name, Information Science, holds a trans and interdisciplinary concept that run through all subject areas. This possibly explains the existence of so many different conceptual definitions for the term information and the fact that the arguments of each definition approaches it from the perspective of the area of knowledge from which it comes from. And it could not be any different. Health, Social, Human, Exact, Biological and Technological Sciences are distinguished from each other not only because of their objects of study, but also because of the way they search, retrieve, interpret, and use information and this is reflected in the products derived from these interactions in each of the areas, without calling into the question what is considered information according to the perspective of each of those areas of human knowledge.

The understanding that the authors are mistaken when considering that there should be moderation in the efforts used to define the term information is not true, even considering that the actual use of terms may differ from their more formal definitions. The ordinary use of a term like “information” may assume different meanings from its formal definition, according to the context in which is being introduced. This suggests that conflicting theoretical views may arise between the explicit scientific definitions and the implicit definitions in common use. As a result, we must not only compare different formal definitions, but also consider the meaning of a word as information, as used in relation to other terms, for example, the search for information, information systems and information services (CAPURRO; HJØRLAND, 2007).

The argument about the concept of information is so complicated, subjective, and complex, that even if an entire encyclopedia discussed on this subject, the risks of impoverishing the discussion regarding the semantic and conceptual relations inherent to the term “information” would not be extinguished. However, this does not suppress nor distort the need for contextualization in each discussion, considering, above all, the comments of Capurro and Hjørland in this regard, in the sense that the actual use of terms may differ from their formal definitions, as the definitions according to the Information Science specialized dictionaries.

Facing the countless approaches that the term “information” may assume, the convergence is a not an unanimous characteristic among the different definitions and concepts coming from different disciplines perspectives which, including, may not even be complementary to each other, even because there are so many perspectives from which the concepts of information can be approached, that establishing a universal concept corresponding to the specificities of all areas of knowledge, simultaneously, remains unfeasible.

Capurro and Hjørland (2007) criticize the attempts on defining the term information, and even consider it as a way of rising up the status of Information Science and impress people, what they call persuasive definitions. Regarding the function and nature of scientific theories on Information Science, the authors consider that the focus on conceptualizing information may have misdirected Information Science and that greater attention to concepts such as signs, texts and knowledge may provide more satisfactory conceptual models for the types of problems that Information Science tries to solve. When we use the term information in IS, we should always keep in mind that information is what is informative for a given person. What is informative depends on the individual's interpretive needs and skills (although these are often shared with members of the same discourse community) (CAPURRO; HJØRLAND, 2007).

Well, the understanding that information is restrict to “only what is informative” is too subjective, restrictive and risky, for the following reasons: first, the language used by the sender of a message (the registration of information), must be common to the understanding of the recipient (reader or end user). In this perspective, any information recorded in a language that is not commonly understood by its author and the reader for which it is designated (sender and receiver, respectively), can unviable the understanding of the message intended to be transmitted.

For instance, the documents published in Braille, which prevent the decoding of their contents by individuals who do not master such language, among which there may even be visually impaired people, and to whom the content registered in documents of a pictorial character is also intelligible. Similarly, the hearing impaired are deprived of decoding information recorded in audio. And besides, physical limitations also can configure barriers to access certain buildings and natural environments that may, for example, contain the registration of information of architectural, historical and geological nature. Such considerations confront Capurro and Hjørland's assertion that “information is what is informative for an individual person”, even because, what is not informative for one person can be for another, or even for the same person, in another context or moment. Thus, the valuation of the information contained (or represented) in any document (information support) or record of knowledge based on the expected interpretive needs and skills of each individual is, at the very least, questionable.

Since the world is a world, information has always suggested to have fostered feelings of greed. Perhaps because it is related to power and status, in addition of being an essential input for the solution of eventual problems and decision making. The social empowerment through the information access has been evident since antiquity, when libraries were primarily (if not exclusively), intended to be used by the nobility. Long before the appearance of technology (particularly referring to the resources that involve Informatics), at least in the format in which it is currently known, the humankind has already expressed its concern on its history registration (that is, information registration). A clear example of this are the inscriptions of primitive men in the caves, from which the forms of production, registration, storage, retrieval and use of information have been reinvented numerous times, until reaching the format in which it is currently known. However, this does not mean that the end of the line has been reached, on the contrary, the innovations involving information, especially considering the use of technologies, indicate that there is still much more to explore and advance on this matter.

5.1.What information is not

At this point of the discussion, the complexity involved in assigning meaning to the term information has already proved to be real and evident. If so, perhaps the application of the fundamentals of reverse psychology might be useful, after all, prior to formulate a definition of what is information in the context of descriptive representation, it is necessary to clear what information is not.

Information is not knowledge. It is not intended here to deepen the discussion on the conceptual differences between information and knowledge, even because, those topics have already been widely explored in the scientific literature. However, it is important to clarify that information is the main input for the construction and transfer of knowledge. Davenport (1998), despite highlighting the difficulties and tenuousness between the conceptual limits among data, information, and knowledge, refers to knowledge as the most valuable information and, consequently, the more difficult to manage. It is valuable precisely because someone has given information a context, a meaning, an interpretation; someone thought about knowledge, added their own wisdom to it, considered its broader implications. The term also implies the synthesis of multiple sources of information (DAVENPORT, 1998). The author still complements that knowledge is difficult to be structured and captured by machines. Usually, it is tacit (not expressed) and difficult to categorize, locate and transfer. Pessoa (2014) also adds that the information is cumulative, while knowledge is selective.

Information is not training (in the education sense). Training implies creation, construction, and the acquisition of specific skills for the development of an intellectual activity or practice. Such results are usually the fruits of the interaction between information and the individual's previous knowledge.

Information is not power. And this is a controversial and questionable aspect. In fact, Pessoa (2014) refers to information as an essential source of power, which allows changing the facts and the course of History. From this perspective, consider that: the information is liable to cause changes in an individual's understanding and posture upon the facts (and that in itself is already a new fact) or a person’s relationship with a particular object. However, such changes are the result of the interpretation of the information received by someone, based on the individual's prior knowledge, not exempting his/her beliefs, customs, social position and convictions. So, in fact, information is not power, but it might promote decision-making and attitudes that can constitute a criterion of favoring for such.

Information is not absolute truth. On the contrary, all the information is subject to question and discussion. Inclusive, empiricism is one of the characteristics of scientific thought. And that is exactly what allows the advancement of knowledge in all subject areas. If this were not the case, there would be no sense on research developments, publications and alike. Furthermore, consider that Science has an infinite perspective, in the sense that a study is always capable of continuity (deepening), questioning or even refutation. Facing this, the theoretical construct that governs the development of human thought is, in fact, the foundation that sustains Science, and this configures the reason that reinforces the need for constant basal renewal and strengthening.

Information is not deformation. Such statement inspires caution due to the double interpretation that the term can lead. Considering deformation as the alteration of the original form of something, be it a custom or a thought, the information assumes a positive character, assuming that every changing process demands a critical sense with the intention to conduct analyzes and assessments about the possible directions for which such a change can lead. The panorama is similar in terms of forming opinions. In contrast, the information takes on a negative character, considering deformation as a change for the worse, which is perfectly possible, depending on the form of interpretation and use of the correct information, in the proper thematic, temporal, and spatial context.

Information is not differential. And again, the difference is not the information itself, but the product that its influence can generate.

Information is not a decision making. But in contrast, making any (wrong) decision may configure a negligence and result in catastrophic results. It is evident that the correct information with correct guidelines in time and space favors the reduction of risks and the increase of security in decision making.

Information is not a data set. Assuming this definition as a true, can be a simplistic, reduced and even dangerous argument, which makes real and imminent the risk of arguing in a reduced way, especially concerning semantic and terminological aspects. Setzer (2015) refers to information as an informal abstraction (which cannot be formalized through a logical or mathematical theory), which is in someone's mind, representing something meaningful to that person. The author, however, opportunely points out that, this is not a definition, but a characterization, since the terms "something", "meaningful" and "someone" are not well defined. For instance, the phrase "Paris is a fascinating city" is an example of information – as long as it is read or heard by someone, as long as "Paris" means the capital of France for that person (assuming the author of the phrase wanted to refer to that city) and "fascinating" has the usual and intuitive quality associated with that word.

Shannon correlates information - that is, the number of possible choices to create a message - and uncertainty. The greater the freedom of choice, the greater the uncertainty, that is, the information. The concept of information seems, as Weaver notes, disappointing and bizarre – disappointing, because it has nothing to do with meaning and bizarre, because it deals not with a single message, but with the statistical character of an entire set of messages instead, and bizarre because in these statistical terms the two words information and uncertainty are partners. (SHANNON; WEAVER, 1972, apud CAPURRO; HJØRLAND, 2007). The irony of total immersion in information and the central role it plays in our economic, social, and cultural life is that, for the most part, we do not have a clear understanding of what exactly information is. Information is not a simple and straightforward concept, but a very slippery notion, used in many different ways. Linguistically and grammatically, the word information is a noun, but, in fact, it describes a process and, therefore, it is like a verb (LOGAN, 2012).

But what is most interesting about it after all, is that information is not a phenomenon, neither a concept nor a property. Information is not knowledge, it is not training, it is not power, it is not absolute truth, nor is it deformation. And in the end, information ends up being everything, even “being nothing”.

5.2.The role of information signification in the construction of knowledge

In 2010, Eric Smith[3] stated that every two days it was created the same volume of information that that was created until 2003. In May 2013, Smith pointed out that 90% of the total of the data in the world was generated over the two predecessor years (EIS, 2016; SIEGLER, c2021). In this scenario, thinking about dissociating Information Science from Information and Communication Technologies is a step backwards. However, even though technologies are essential for the performance of Information Science, it is prudent to maintain the understanding that machines, despite being capable of performing fabulous functions and tasks, are absolutely limited with regard to the interpretation of the data stored by them. Computers cannot think because they lack our semantics (SEARLE, 1991). However, the data structures and semantic correspondences of the Web (information architecture) are tasks that cannot be accomplished without human intervention.

As an alternative to mitigate possible obstacles to the recovery of information resulting from this scenario, Information Science has instruments to reduce the degree of unintelligibility of information by machines. It deals with descriptive and, above all, thematic representation. In fact, the essence of the semantic web and information architecture is based on variants of thematic representation, which are the ontologies, which enable the virtual correlation between the information decoded into data, which culminates in the data and metadata structure.

The research on the production, storage, dissemination, appropriation and use of information, among other factors, is one of the reasons of existence of Information Science. Starting from this and, from the perspective of descriptive representation, considering metadata as the primary elements for the management and organization of information, leads to the understanding that the metadata, and not the information itself, are the main object of study of Information Science (disregarding social and cultural aspects of information). The approach may seem shocking, but it makes sense from considering the following points:

[3] Eric Emerson Schmidt is the president and former chief executive officer (CEO) of Alphabet Inc., formerly Google, the main subsidiary of the newly created company. He worked at Novell and holds a Ph.D. in Computer Science from the University of California at Berkeley and a Bachelor of Science degree in electrical engineering from Princeton University (CRUNCHBASE, c2021).

a) The information is immeasurable. There is no science or unit of measurement capable of quantifying the volume of information under a single person's domain. In contrast, Information Technology has specific computational measurement units to quantify data, not only in terms of storage, but also considering transfer (data traffic)[4] matters. (BERNERS LEE, 199 ?; GLEICK, 2013).

b) Information is priceless. Despite being the object of desire of many, mainly publishers of scientific content, the valuation of information is another aspect for which there are no quantifiable parameters. Information is not subject to sale, donation, lease or loan. Such an argument reinforces the invalidity of the trivialized statement that "information costs expensive" since, what really costs expensive and what is sold, in fact, is the access to the means of storage of information records, and not the matter itself.

c) Information cannot be stored. Information is dynamic, in the sense of integrating an individual's knowledge. Davenport (1998) points out that knowledge is the most valuable information and, consequently, the most difficult to manage. It is valuable precisely because someone has given information a context, a meaning, an interpretation; someone thought about knowledge, added their own wisdom to it and, considered its broader implications.

d) It is endowed with meaning. Starting from the definition of information by Drucker (1998), according to which information is defined by data with relevance and purpose, Davenport (1998), complements that even when a computer automatically turns a cost sheet into a more informative graph, like the commonly used 'pizzas', someone has to choose how to represent that design. People turn data into information, and that is what makes information administrators' lives difficult. Unlike data, information requires analysis. And, as simple the information entity (price, taxes, consumer, year) may seem, there will always be someone disagreeing with its definition.

The contextualization and the attribution of meaning to information (be it scientific or not), is not the priority of Information Science, but the treatment of data, under a systematic and phenomenological approach, through which the objective is to establish relations between the meaning of information as an object, and the

[4] The widely used units of measurement for measuring data in Computer Science are: Bite (B), Quilobyte (KB), Megabyte (MB), Gigabyte (GB), Terabyte (TB), Petabyte (PB), Exabyte (EB)), Zettabyte (ZB), Yottabyte (YB)

information needs as human beings. The descriptive and thematic representation of information and knowledge are the main instruments that Information Science has to connect data to readers. Now, such procedures, especially the thematic representation, deal with the codification of information (meaning) in descriptors whose intention is precisely to contextualize and reflect the approach (interpretation) intended by the author. On the other hand, the information is capable of being represented in the form of data, which in turn can be stored, processed and retrieved, to enable its access. Although such data tends to be assumed by whoever accesses it as information, it is important to emphasize that the information architecture is based on the semantic interconnections of the data. And in the end, from a pragmatic approach, Information science has the systematic management and organization of data, as its object of study and work, considering them as a means of access to information so that they can reproduce the information they represent, in the reader.

According to Drucker (1998), information is a set of data endowed with relevance and purpose. Such attributions are the result of the action of the psyche and human cognition, which interpret the data and give it meaning, interpretation and contextualization. It should be noted that this understanding is peculiar to each individual and, depending on the way it is conceived, it can set up obstacles for the transfer of information. Capurro and Hjørland (2007) corroborate that information is not a pure observable element, but a theoretical construct. It is an interpreted data.

In addition, it cannot be ignored that the registration of information transcends the limitations of the physical support and the generation and storage of data: signs also represent encoded information, such as the traffic light that controls the vehicles flow in an urban road. Searle's allegory suggests an example that can further clarify these concepts. Suppose a table with three columns, containing city names, months (represented from 1 to 12) and average temperatures, in such a way that the column titles and city names are in Chinese. For someone who knows nothing about Chinese or its ideograms, the table is made up of pure data. If the same table were in Portuguese, for those reading this article it would contain information. Note that the table in Chinese could be formatted, the lines ordered according to the cities (given an alphabetical order of the ideograms) or months etc. – examples of purely syntactic data processing. (SEARLE, 1991; SETZER, 2015).

In this sense, Davenport (1998) points out one of the information characteristics, which is being much more difficult to transfer with absolute fidelity and exemplifies with Chinese whispers children's play. Also, the case of the book Raízes do Brasil[5], authored by Sérgio Buarque de Holanda clarifies this panorama. Although the work deals with Brazilian historiographic and sociological aspects, a misinterpretation of the title may induce the work to be indexed, in a wrong way, in Botany instead of History.

It is understood, therefore, that the concept of information, in addition to representing much more than a set of data, necessarily requires the human action with regard to the attribution of sense, meaning and contextualization. Such argumentation leads to the understanding that data, as raw information, is the representation of quantified or quantifiable symbols, which are converted into information through decoding and human interpretation, as refined, treated and / or interpreted data. Then, if the symbols are to be equally interpreted, then the information can also be understood as the decoding and interpretation of the signs and symbols that can be represented by the data, or not.

The conversion of data into information necessarily takes place through human mediation. The interconnected, interpreted, and internalized information results in knowledge which, by its turn, is the result of reflection and contextualized synthesis, which culminates in insight. Knowledge and insight are tenuously distinguished and the correlation between the two terms occurs in a degree of abstraction such that their dismemberment would yield a rich theoretical-philosophical discussion, which will not be addressed here because it is not the focus of this discussion.

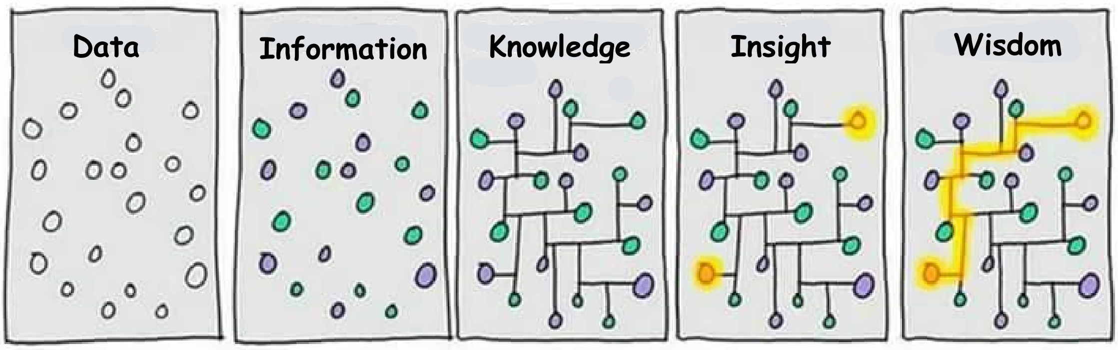
More commonly applied to psychology, psychoanalysis and psychiatry, insight is linked to the concept of understanding, interpretation, idea, clarification and self-knowledge. It deals with the understanding of a certain fact or problem, based on the knowledge previously acquired. Wisdom, in turn, is the ability to identify and link insights in specific contexts, to promote the practical application of knowledge in understanding situations and facts or, in solving problems. Wisdom is the summit, the height of human knowledge, and can be communicated (externalized) through written, spoken, visual, audiovisual or tactile language, in digital or analog media. And it is

---

[5]In Portuguese, the term “Raízes”, isolated, can be understood both as “the roots of a plant” or as “historical roots”.

precisely in this context that appear the documents, here considered as secondary inputs of Information Science and enabler of the Information's primary input, the information itself, contained in an information support. Unless new technologies are developed that allow the reading and decoding of the human mind Knowledge, insight and wisdom can hardly be machine readable, as they are products of thought, based on data and information, which generate new knowledge and insights and culminate in the increase or specificity of wisdom, in a continuous cycle.

Figure 3 -The construction of wisdom



Source: Extracted from a social media. Authorship unknown.

The macroanalysis of Figure 3, leads to reflection for the questioning about the construction of knowledge, considering the path to be taken from the conception of an idea, its trajectory marked by the coding and decoding in data, information, knowledge, and insight, up to its maximum maturation stage, the wisdom.

This fosters the understanding that scientific communication implies the encoding of information into data, and its recoding into information and again into data and again into information, successively by different individuals, each one from his own perspective. In such a scenario, information, as it was conceived, is submitted to different interpretations, influenced by cognitive, social, economic, geographic aspects. And in this context, there is no need to talk about fidelity in the transfer of information, first because what is being transferred, in fact, is the representation of information (data and metadata) and second, the information is intimately and conditionally linked to understanding, configuring a fruit of the human psyche. This is precisely what Setzer (2015) referred to when mentioning that “any text constitutes data or a sequence of data". In view of this, wouldn't it then be a utopia to consider fidelity in the transfer of information?.

5.3.The Shannon's communication model[6]

Capurro e Hjørland (2007) associate the concept of information with "communicated knowledge", in this context it is important to consider that the communication of knowledge occurs through its registration (the communication itself), through oral, visual, auditory, tactile, or written ways. This means stating that the communication of knowledge takes place through its documentary record.

Brookes (1981), reflects on an anomaly in the use of the term information retrieval as usually considered by the bibliographic databases. The author reports that in response to an eventual search, the researcher obtains a list of references, possibly with summaries, through which he hopes to locate the referenced material in a library. Once located, it is up to the researcher to read the publications' contents, searching for the desired information, without any guarantee that it will be found in the publications selected by the used search engines, nor that the retrieved publications are all relevant to their research. The author also points out that although the techniques of information retrieval have been improved over the years, the criticism of information retrieval systems remained valid (at the occasion of his statement, in 1981), as they remain nowadays. Brookes emphasizes the practicality of information retrieval systems; however, he criticizes the arbitrariness in indexing techniques, since indexing is nothing more than recording the interpretation of the significance of the data recorded in a particular information support, under the vision of the documentalist / indexer (who is influenced by the most diverse socio-cultural and other factors). Considering that such professionals are not usually experts on the subject reported in the publication being described and indexed, the chances of misinterpretation of its content are key factors for the increase in the retrieval of documents which do not meet the search criteria (parameters) pre-established by the researcher within the search engines. In this perspective, Brookes (1981) stresses that, from the Information Science perspective, the research in information retrieval systems offers only a

[6] The Shannon's Mathematical Theory is available at Shannon (1948).

theoretical 'cul-de-sac'[7], which leads nowhere, and adds the computers information handling processes used by information retrieval systems, their storage capacities, their data inputs, and internal information transmissions, are measured in terms of Shannon's theory – in bits, megabits per second, and so on. On the other hand, in theories of effectiveness in information retrieving, such processes are measured in what Brookes call physical measures – that is, publications (or their substitutes) are counted as relevant or not relevant and just the proportionality of these numbers is used - the documents (or their substitutes) are counted as relevant or non-relevant and simple ratios of these numbers are used). Subsequent probabilistic calculations are made as if the documents were physical things (of course, they are in part), even though the initiative, as a whole, is called the information retrieval theory. So why, Brookes asks, the logarithmic measures of information used in the theory of the machine and linear or physical measures of information in information retrieval theory? [even because, bits and bytes and other measures of Computer Science are not units of measure of information, but of volume of data]. If the information retrieval theory were called document retrieval theory, the anomaly would disappear. And the theory of document retrieval would be considered as a component of Library Science, which is similarly concerned with documents. But this is a very simple idea. Those who work with information retrieval theory explicitly claim to work with information, not documentation. Brookes therefore abandon the simple explanation of incorrect terminology. Finally, he assumes that information retrieval theorists mean what they say – that they are contributing to Information Science. But, are they? (BROOKES, 1981).

Brookes demonstrates a terminological and conceptual confusion established within the scope of recording, processing, measuring and retrieving information. Capurro and Hjørland (2007) disagree with Brookes on the argument that it is not such a simple idea to state that the information retrieval theory is, in fact, the document retrieval theory and, therefore, is intimately linked to Librarianship. And it really is not, mainly because it is related to the Information Science's conceptual arsenal. The authors also add that Brookes' argument is incorrect and has provoked an endless speculation

[7] Cul-de-sac corresponds to an expression of French and other Romance languages origin, whose literal translation is “sack bottom”. In the context approached by Brookes it is used in the sense of “dead end”.

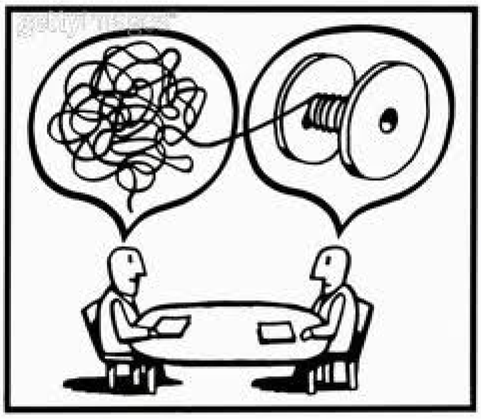
about the nature of the information, which has not contributed to the understanding of the problems of information retrieval (CAPURRO; HJØRLAND, 2007).

Despite the complexity and interdisciplinarity between Information and Computer Science, there are no inconsistencies in Brookes' arguments. In fact, Ellis (1996, p.188), goes further and complements that "even if the differences between objective and subjective knowledge are considered valid, it is still far from being clear how this can be operationalized for measurement purposes". Still in this aspect, it is necessary to consider that more important and complex than quantifying information, is measuring the relevance degree and the impact (results) that it provokes in its target public, considering, above all, that the alteration or the complementation in an individual or group’s behavior is the information’s reason of being.

Noises in the communication between the sender and the receiver of the registered knowledge, are established from the conceptual considerations of Information Science, and remain in the treatment of information as argued by Brookes. The documentalist's cognitive and interpretive interference in the thematic treatment of information can affect (positively or negatively) its retrievability levels. Brookes also referred to “arbitrary indexing techniques”. Some institutions, especially those with expertise in particular subjects, adopt too specific descriptive and thematic representation methods, even to guarantee the representativeness of the specificity of the documentary collection they hold. However, such a posture, while reducing dispersion, can restrict access to information by individuals who do not have complete mastery of the adopted documentary languages. Shannon stresses that the fundamental problem of communication is the exact or approximate reproduction of a message at a point other than that of its origin, considering its meanings and relationships with physical and conceptual entities, and highlights that this is not a problem of and for technology. In fact, it is up to the Information Science to guarantee the minimization or, in an ideal scenario, the transmission of the semantic content of the information in its essence, free from the influence of external noises.

However, even with such nuances and particularities, it should not be disregarded the fact that, regardless of the form, all available information is only searchable and accessible, thanks to the efforts of some professional groups dedicated to the representation of human knowledge. And not only that: in addition to the proper treatment of information, it is required a management and recovery instrument which is sufficiently robust to manage the information and overcome the limitations of the human brain. This scenario makes essential the electronic systems for managing and retrieving information, especially considering the huge volume of information recorded on the most diverse media on a daily basis, as stressed by EIS (2016). And was it not for those system, how would information be retrieved, even considering the restrictions which are inherent to manual labor? How would the management of the registry of the human knowledge take place, considering its current proportions, without using the resources of the electronic systems for managing and retrieving information? And concerning the information delivery: how is it perceived and understood by its target audience? What are the distortions influencing information on its way from production to reception by the final reader, and how much is it exempt from the author's own perceptions or, to what extent is information impartial and true to reality? The Figure 4 illustrates, in a satirical tone, the role of Information Science in relation to all these issues and represents the area's commitment to Science: the attribution of clairvoyance in the process of archiving, searching, and retrieving the records of human knowledge.

Figure 4 -The role of the Information Science on the information representation



Source: Getty Images

One of the competences of Information Science is the transposition of the “tangle of information” encoded and dispersed in the digital universe, into an intelligible language, not only by machines, but mainly by human cognition. In this scenery, information representation is one of the main agents responsible for information retrieval. This justifies the improvement practices in the elaboration and management of data and metadata in the various forms it can be used in the Scientific universe (bibliographic records and bibliographic references included).

5.4.Information contained (represented by) in data or data containing (representing) information?

Data can be understood as quantified or quantifiable symbols. Quantifiable means that something can be quantified and then reproduced without one even noticing the difference from the original. Therefore, a text is a data. In fact, the letters are quantified symbols, since the alphabet, being a finite set, can itself constitute a numerical base (the hexadecimal base generally used in computers uses, in addition to the 10 decimal digits, the letters A to E). Photos, figures, recorded sounds, and animation are also data, as they can all be quantified when inserted into a computer, to the point of eventually having difficulty distinguishing their reproduction from the original. It is very important to note that, even if incomprehensible to the reader, any text constitutes data or a sequence of data. (SETZER, 2015). Cunha and Cavalcanti (2008) define data as the smallest conventional and fundamental representation of information (fact, notion, object, proper name, number, statistic, etc.) in analog or digital format which can be submitted to manual or automatic processing.

The understanding of the data concept requires the previous comprehension that the conceptual approach to this term is restricted to representative aspects. There is no known of approaches associating such definition with any cognitive, interpretive, or semantic aspect. Davenport (1998, p. 19), understands the concept of data as the “observations on the state of the world”. As an example, Davenport suggests the statement that “there are 697 units in the warehouse”. The observation of these raw facts, or quantifiable entities, can be done by people or by appropriate technology. From an information management perspective, it is easy to capture, communicate and store data. Nothing is lost when represented in bits, which certainly comforts Information Technology (IT) people (and not just the IT people, but also the IS people).

However, what comforts IT people, does not necessarily comfort IS people. At least not in the same proportion. Considering the example suggested by Davenport, in terms of technology, the 697 units in the warehouse correspond to 697 entries in a database, capable of promptly returning any search for such a volume of units. As from the Information Science perspective, the 697 units in the warehouse correspond to 697 records through which each of the units is expected to be properly and systematically described and individualized, so that through the database containing such record it is possible to identify not only the quantity total units, but how many of these are blue, how many are green and how many are yellow. Which manufacturers, dimensions, and mass (physical unit of measure), purpose and validity, among other aspects, which can also be considered as variant forms of access to such data.

This leads to the understanding that, from a technological perspective, data corresponds to the representation of information. On the other hand, for Information Science, the data has the mission to inform about the existence of an object, but they must go further: provide as many descriptive and thematic details (not to mention the correlations) as possible about it. In this context, data can be considered information, in the sense that the provision of a more detailed description about an object enhances the possibilities of the selection X information retrieval relationship.

The characteristics inherent to data, as a mathematical entity capable of representing and quantifying, and the fact of being machine readable, enable data storage, treatment, and processing by computers, which are undoubtedly fabulous machines for data management, in any instances. And (be justice done), with the volume of data generated at each moment, Information Science would certainly lack “arms and legs” to treat and make them searchable and recoverable by itself, without the use of the information and communication technologies. (ICTs).

While ensured the exempt from the technology influence, Santos and Ribeiro (2012) define data from the perspective of Information Science, as the standardized representation of facts, concepts or instructions in order to allow communication, interpretation or processing by human or automatic means.

There is an infinity of data types from which, the most relevant for this discussion being the following:

a) the bibliographic data, which deal with the “set of elements (author, title, place of edition and other data used in the bibliographic description that represents a particular document)” (CUNHA; CAVALCANTI, 2008);

b) the cataloging data, which Cunha and Cavalcanti (2008) equate to bibliographic data;
c) the thematic data, corresponding to those that have the structure of a classification and whose main objective is to represent the subject to which a particular content is about and, finally;
d) the machine-processable data, which Cunha and Cavalcanti (2008) equate to computer-readable data and define as data recorded in the most diverse formats that can be entered into the computer for processing. They can be stored in magnetic tapes, floppy disks or even be read by optical readers as mechanically readable information, machine-readable information or mechanizable information.

There is also another type of data, which in addition to being representative, also has characteristics that favor the organization, location, and retrieval of information. This is metadata.

Data and mainly metadata are the determining agents of management and organization of information concerning its storage and retrieval which have a direct influence in the hit-silence[8] relationship in an information search process. This justifies the evident Information Science's concern on the establishment of instructions that guide the creation of increasingly clear and consistent metadata.

Such considerations reinforce the understanding that the Information Science's object of study and work is not the information itself, but the (meta) data. The information is in such a degree of abstraction that its access is conditioned to the reader's decoding, understanding and cognition capacities and, therefore, outside the governance zone of librarians (and other IS experts), technology and IT professionals, and any other professionals and / or areas of knowledge. And it is precisely there that the greatest challenge for Information Science and their professionals is established: promoting

[8] Hit – relevant information that is selected after a search. [...] in computerized databases, the comparison of two items that meet predetermined characteristics [...] number of citations, records or units of information obtained in a search. (CUNHA; CAVALCANTI, 2008).
Silence - The non-recovered data, understood as silence in the information retrieval, correspond to the "absence of pertinent documents, which were excluded from the list provided by an information system, due to the system itself failure" (CUNHA; CAVALCANTI, 2008).

access to information, considering all the determining factors for that and which are outside their zone of governance.

The processing and retrieval of information involves several IS activities, including cataloging, classification, indexing, bibliographic references writing and other normalization activities, ordering documents in physical collections and structuring databases and research interfaces in digital environments. Even academic researchers have been studying how to index, store and retrieve bibliographic references, calling this discipline ‘information retrieval’, and not ‘retrieval of references’. Regardless of the nomenclature, the retrieval of references can represent obstacles that do not always reflect most of the real problems related to the retrieval of information, which demand a solution considering the contemporary information explosion. For instance, business analysts, journalists and scientists hardly need bibliographic references for developing their work. Most of the time they need facts, that is, the direct information about the problem in which they are involved and, generally, they have no interest or time to follow references, search for texts in libraries and read articles (FREI, 1996). Such a scenario reaffirms and highlights the essentiality of dynamism concerning the treatment of information and, considering even the daily demands, not simply of the professionals mentioned by Frei, but also of scientists and researchers in general, it is understood that metadata, in addition to identifying and facilitating access to the publications they represent, they also provide the optimization of time spent on the search and retrieval time.

5.5.Conceptual observations concerning information to the descriptive representation

Capurro and Hjørland (2007) relate the difficulty in attributing meanings to concepts to the fact that they do not represent a specific and well-defined function in the context of the theory in which they are inserted. They highlight the use of terms with different approaches in different disciplines and that the actual use of terms may differ from their formal definitions (CAPURRO; HJØRLAND, 2007). Peirce (1905 apud CAPURRO; HJØRLAND, 2007), mentions that the meaning of a term is determined not only by the past, but also by the future. Therefore, the question: “what is information?” cannot be done without reference to a situation (CAPURRO;

HJØRLAND, 2007) and in view of the points above, any attempt to answer this question cannot fail to consider the promising aspects concerning technology, which in fact, pose themselves as challenges to Information Science, demanding changes or even breaks in paradigms and adaptations in their ways of managing and making information available.

The word information – and its combinations like information retrieval and information center – has definitely contributed to raising public opinion about the work of the library and documentation, which is generally considered to be uninteresting, dusty and distant from what is really happening. in society. Perhaps it would be wise to leave the word information there, were it not for the fact, already mentioned, that several attempts have been made to define information as a formal term, related to working with documentation and information and even to define it as quantity measurable, corresponding to questions such as: how much information was retrieved by the search? (CAPURRO; HJØRLAND, 2007).

Starting from the argument that the actual use of terms may differ from their formal definitions, also referred by Capurro and Hjørland (2007), it suggests that the conceptual panorama in Information Science has questionable nuances: For instance, we did not find explicit manifestations about the concept of descriptive representation in the specialized dictionaries in Information Science. Including, Santos and Ribeiro (2012), conceptually equate the terms descriptive representation, bibliographic description, and physical description. In addition, the authors indicate the term cataloging as the “see” remissive type, for the three terms, defined on the entry “cataloging” as: Cataloging 1. is a conventional set of information determined by examining a document and intended to provide a unique and accurate description of such document. It is the first stage of the intellectual treatment of a document from which is extracted the information described in accordance with fixed rules. 2. Phase of the cataloging process related to the identification and description of the works. It serves to establish the author access points and provide adequate bibliographic information to identify a publication; 3. Descriptive representation. (SANTOS; RIBEIRO, 2012).

Considering cataloging as a librarianship tool aimed at individualizing, recording, and maximizing the ways of accessing a particular document, with sights to managing and organizing information through catalogs formation, it is understood as being an integral part of a larger process called descriptive representation, as a facet, even disregarding its subdivisions: cataloguing-in-publication and cooperative cataloging.

Descriptive representation can be understood as the set of information that identifies the physical, authorial and editorial characteristics, in addition to the main access points (non-thematic) of a work. Although the differences are tenuous, it should not be said that cataloging constitutes the universe of descriptive representation, but that it is part of this process, instead.

On the other hand, Santos and Ribeiro (2012), thus define the term bibliographic reference: Bibliographic reference - 1. It is a set of bibliographic data that identifies a document or part of it, following a specific norm; 2. Transcription of data that allows the identification of the documents, in whole or in part. It is worded according to existing standards in a large number of countries. 3. Data set that allows the identification and location of documents cited as research sources, carefully ordered according to specific Bibliographic Referencing rules.

Before commenting on this, it is necessary to briefly clarify the combination of the terms "bibliographic" and "references", since this is not the focus of the present discussion. The prefix biblio- refers exclusively to books, so that the use of the two terms (reference and bibliographic) combined gives a reductionist aspect to the bibliographic references, in the sense that it restricts them to the universe of printed publications. Thus, it cannot be said that the expression “bibliographic reference” is correct or not, without a further accurate evaluation on this matter. If the reference refers to a printed book, there is no problem using the term “bibliographic”. However, the scenario changes if, for example, the publication being described is a book in digital format or any other source of information published in non-paper format. Added to this, the fact that the use of electronic publications is increasingly expressive in the scientific universe. Technology is so intrinsic and so impregnated in scientific production that it became almost impossible to develop an argument exempt from the use of technological resources and, thus, there are rare cases in which the use of the term “bibliographic references” is used correctly, mainly as a session title within a publication. Starting from this, the tendency is (or at least it should be) that the use of the term "bibliographic references" is gradually discontinued in favor of the term "references". However, neither the cataloging entities and/or metadata producers nor the standardizing institutions comment on this regard and this constitutes an obstacle to the change of posture in the academic-scientific community. However, considering such aspects, the understanding of the use of the biblio- prefix is also a non-consolidated issue in Information Science.

Having made these clarifications and considering the definition of bibliographic reference according to Santos and Ribeiro (2012), the comparison between the definitions for cataloging and references highlights common goals between both techniques. Two examples of this are the identification of a document and the use of guidelines determining the way which metadata should be registered. However, the bibliographic references are not aimed at forming catalogs. Even sharing some similarities, while the cataloging refers to the registration, control and access to a certain collection, bibliographic references are dedicated on communicating about the existence of a work, about a collection of documents selected according to specific criteria (subject, author, delimitation geographic or chronological, etc.).

Another issue to consider is that bibliographic references not always allow the location of the referenced work. In cases where written documents are available on the Web, it is recommended that the author of the work indicate the URL where the referenced work can be found, for favoring its location and further access. However, in cases where the publication is not available for consultation on the Internet, the location of the work is usually restricted to consulting the libraries' bibliographic catalogs.

Although the differences between bibliographic description, descriptive representation, bibliographic references, and cataloging are marked by tenuousness, it is important to have clarified the object of interest of each subject. Consider, therefore, that the cataloging and the elaboration of references are fragments of descriptive representation, while the bibliographic description is one of the elements composing cataloging. Starting from this, the premise that the intentionality of each of these practices refers to Ranganathan's five fundamental Laws is true:

1. Books are for use
2. Every reader, his/her book
3. Every book, its reader

4. Save the time of the reader

5.The Library is a growing organism (ZABEL AND RIMLAND, 2007).

However, despite the shared purposes among the bibliographic description, the descriptive representation, the bibliographic references and the cataloging, they should not be confused nor treated as synonyms, first because, in fact, they are not and, second, because the approaches on the completeness and the ways of registering the descriptive metadata, and the prescriptive documentation (e.g., the standards, the conceptual models and the rules), differs among them.

Information is what is understood versus what generates information. For the purposes of Science, "information" had to mean something special, said Gleick (2013), referring to Shannon's investigations that led to the elaboration of the theory for information. [...] We have autonomy, we are specialists, and that is why we see information in the foreground. But she was always there.

Information is the record of knowledge, which is not measured, not priced, not transferred, but can be treated, accumulated, and disseminated. Considering the information as the meaning that a set of data represents for one or more individuals, it is evident the understanding of the concept of information from the descriptive representation, since it considers as information all and any characteristic of the treated work, without dwelling to aspects of meaning (semantics). This means stating that the descriptive representation is focused on the physical description of a work, so that its characteristics and access points regarding responsibilities and other access points not related to the thematic treatment are primarily considered.

From the perspective of descriptive representation, the relationship established between information and its meaning is peculiar. Traditionally, the concept of information is linked to aspects of context and meaning. However, the descriptive representation presupposes that the registration of the information must be presented in a clear and easy to understand language, in addition to speaking for itself and making itself intelligible without the need for prior contextualization or knowledge for this.

The issues referring exhaustiveness in the bibliographic description is likely to generate discussions and even controversies in Information Science. A good documentary description should be detailed enough to identify and individualize the described document, while maximizing and enhancing its forms of access in a way to send the researcher precisely to the requested piece of information and, at the same time, such description should be succinct enough to omit the metadata considered unnecessary for the proper identification and retrieval of information. Finding the balance between these two aspects is one of the challenges imposed on the documentalist which varies, according to the needs of the public to whom a description is being directed.

The intention of this brief reflection was not to elaborate a universal concept for the term information, but just to make a few comments on how the information is approached in this study. However, this reflection is important for better understanding the bibliographic universe and how information is being approached in different scenarios and perspectives. For instance, considering the bibliographic references' view of information, it can be said that they register data representing a publication, which, in fact contain the registration of the information which represents the knowledge of its author. Dealing with those aspects may be a complex task, primarily for professionals other than those from Information Science. Breaking down such barrier, is a task for which there is still a long way to go.

6. METADATA IN THE UNIVERSE OF DESCRIPTIVE REPRESENTATION

Understanding what information is helps to understand the data encoding and decoding process. Angeloni (2003), states that the communication process is a sequence of events in which data, information and knowledge are transmitted from a sender to a receiver. Davenport (1998), suggests that one of the characteristics of information consists in the difficulty of its transfer with absolute fidelity and, since knowledge is information with value, consequently, its transmission is even more complex. Thus, the process of reading, decoding and interpreting the descriptive record representing a document is also a communication process for which special care should be taken, so that it is reduced the risk of noise interference upon the metadata decoding and in the identification of the described work.

The identification is the focal point of descriptive representation, considering that its main objective is describing an item in an individualized way, with maximum precision. In addition, the identification of the item being described is the basis of the organization of information, considering its constant and exponential growth that showed pharaonic proportions, especially after the Internet's advent. It is evident that electronic information is not a single protagonist in the scope of Information Science. However, it cannot be ignored the fact that the habits of production, search, retrieval, storage and use of information has undergone drastic changes over the decades and this has affected not only the data treatment and management procedures, but also has fostered a multitude in the forms of presenting such information and in the types of media containing its registers, so that a work can be expressed through the most diverse information resources, be it analog or electronic, and these variations must be treated by the descriptive representation's processes in an individual, clear, detailed and careful way, to highlight all the facets and particularities of the document being described, as well as the support of information in which it is stored, which correspond to the practice of the documentary description itself.

The incorporation of technological resources by the academic-scientific universe imposed on Information Science the challenge of organizing the dense and diversified mass of documents available on the Web. By the way, Breitman (2007) comment that here is currently an enormous amount of information available on the Internet. For many users, the biggest advantage of the network is the number of services that can be accessed from their homes and offices. Today's Internet provides access to financial information, queries to large databases, stock purchases and sales, books, household appliances, auctions, weather information and tickets and hotels reservations, among many other options. The possibilities seem endless, but the technology leaves a lot to be desired, considering a crucial point - information about information is lacking. For instance, Breitman points that the results for a search for websites for booking hotels, in addition to the desired websites, shows a series of "junk", which do not interest the user who really wants to make a hotel reservation. The author’s conclusion is that to improve situations like that, it is necessary to index the resources of the Internet; in other words, add elements that inform what kind of information or service is provided by those pages. What is needed is metadata.

Since information is an abstract asset (or an intangible phenomenon) of difficult measurement, which is contained and contains several social, cognitive, and scientific aspects besides a multitude of contexts, Breitman's discourse leads to the understanding that information is not self-retrieving, nor self-organizing. In such a context, metadata can be considered as the raw material for the organization of information and assume a fundamental and indispensable role in the identification, description, contextualization, and location of information, without which, the Web would lose its importance and meaning, and would be nothing more than a receptacle for a heap of irrecoverable and out of context data.

Any information, regardless of its presentation format, has three elements in common: the content, the context and the structure. The content refers to what is contained in the document or the subject it deals with and is intrinsic to the information object. The context indicates the aspects related to who, what, why, where and how, associated with the creation of the object at hand, and is extrinsic to the to the information object. The structure refers to the formal set of associations within, or between objects of information and can be intrinsic to the information object, extrinsic, or both (GILLIAND, 2008).

On the other hand, the stripped information, mainly of context and structure, results in the data, which corresponds to the smallest part of the information. However, the lack of some properties, especially of human interpretation, reduces the information into data, which although can be quantified, is not enough for the reduction of the uncertainties regarding the organization of the information. Data are signals that have not been processed, nor correlated, integrated, evaluated or interpreted in any way, and, in their turn, represent the raw material to be used in the production of information (RUSSO, 2010). In turn, data must be strategically and systematically organized under a logical structure that makes it possible to search and retrieve the information subjectively contained therein. Such a structure is materialized in the description, ordering, correlation, and presentation of metadata.

6.1.Metadata terminological considerations

Although metadata increasingly represent one of the objects of interest for Information Science, the term originated in Computer Science. In computational terminology, the prefix “meta” can be understood as “about”, so that metadata is data used to describe other data (SHELLEY; JOHNSON, 2015 apud CAPLAN, 2003). The term was designed by Jack E. Myers in the 1960s and in 1986 it was registered as a trademark of the Metadata Company, which operates in the development of software and services applied to medicine and healthcare. The association of the term metadata with the information needed to make computer files readable to humans occurred for the first time in the early 1990s (CAPLAN, 2003). Contradictorily, the W3C[9] refers to metadata as "machine-readable information for the Web" (W3C, 2001). For Cunha and Cavalcanti (2008), metadata corresponds to information that describes the structure of the data and its relationship with others, e.g., a label on a database record, indicating the field containing the author's name. Rowley (2008), goes further: Metadata means data related to data. The bibliographic records are a type of metadata. However, they are increasingly being used in the more specialized contexts of data referring to digital resources available in a network. Metadata also differs from bibliographic or cataloging data because the location information is contained in the record in such a way to allow the direct delivery of the document from appropriate application programs, or, in other words, the records may contain detailed information regarding the access and network addresses. In addition, bibliographic records are

[9] The World Wide Web Consortium (W3C) is an international community comprised of the W3C team, associated organizations, and the general public, under the leadership of Tim-Berners Lee and Jeffrey Laffe, whose goal is to develop standards for the Web.

designed for consultants to use them not only when assessing relevance and deciding whether they want to locate the original resource, but also as a unique identifier for the resource, so that they can request it – the resource or the document - in a way that makes sense to the person to whom the request is delivered. These functions remain important. Internet search engines use metadata in the adopted indexing processes to index Internet resources. The metadata should be able to describe locations and versions of documents that are in remote locations, as well as to adapt to the lack of stability of the Internet, the redundant data, the different perspectives concerning the granularity of the Internet (e.g., what is a document or a resource?), and different locations on several different networks.

Perhaps the most traditional (and also the most vague) definition for the term metadata (i.e., "data about data") becomes even more vague, considering that the term can be approached from different points of view, and by several areas of knowledge, as remember Bacca (2008, p. 7)

> When applied outside the original repository, the term metadata acquires an even broader scope. An Internet resource provider might use metadata to refer to information that is encoded in HTML meta tags for the purposes of making a Web site easier to find. Individuals who are digitizing images might think of metadata as the information they enter into a header field for the digital file to record information about the image file, the imaging process, and image rights. A social science data archivist might use the term to refer to the systems and research documentation necessary to run and interpret a magnetic tape containing raw research data. An electronic records archivist might use the term to refer to all the contextual, processing, preservation, and use information needed to identify and document the scope, authenticity, and integrity of an active or archival record in an electronic record-keeping or archival preservation system. Metadata is crucial in personal information management and for ensuring effective information retrieval and accountability in record keeping—something that is becoming increasingly important with the rise of electronic commerce and the use of digital content and tools by governments. In all these diverse interpretations, metadata not only identifies and describes an information object; it also documents how that object behaves, its function and use, its relationship to other information objects, and how it should be and has been managed over time.

Berestova (2017) talks about the terminological problem regarding the definition of the term metadata in Information Science and, considering the definition of metadata as "data about data" or "information about information", the author argues that such unfounded generalizations obscure the nature of the metadata phenomenon, violate the purity and logic of the reflection on such matter, and confuse the scientific knowledge process. Kogalovskii (2002, apud BERESTOVA, 2017, p. 28), refers to a developed study concerning the metadata term definitions. The author reports that twelve definitions were analyzed, however, none remained adequate, in his view: some due to its generic approach, others due to not adequately explaining the purposes of metadata or even, due to not adequately clarifying the idea (conceptual basis) of the metadata, although highlighting the link established between metadata and descriptive resources. However, the definitions were unanimous in associating the definition of metadata with "data about data" or "information about information". Kogalovskii concluded by emphasizing the possibility of using metadata to describe not only data, but also other resources, since metadata existed a long time before the advent of computer systems, e.g., the bibliographic descriptions, the various catalogs, classifications, and notes.

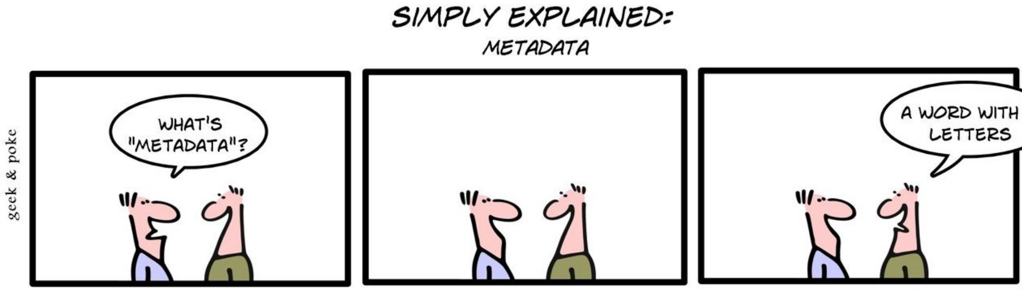
Therefore, it is important to note that metadata is not necessarily restricted to the digital format or refer exclusively to digital resources. From the perspective that metadata is a set of descriptive information that identifies and individualizes a particular item, any information of this nature, including the early analog catalogs, can be considered metadata.

From the electronic point of view, it can be said that the correspondence between the metadata and the terms used in a search (not disregarding the thematic descriptions, i.e., the thematic metadata) are the determining factors for the composition of the results presented, or not presented, for a particular search. It should be noted that the searches are not made from the documents themselves, nor from the resources representing them, but from the corresponding metadata containing information about such documents. Therefore, responsiveness rates achieved by the search engines are directly proportional to the quality and the consistency of the metadata composing the database to where the search is being submitted and to the correspondence between such metadata and the search terms (query texts).

The definitions for the term metadata which can be found in the literature are generally broad and insufficient and, in some cases, different and incompatible. The most common definition, “data about data”, cannot be considered incorrect, but in contrast, it disregards several conceptual, technological, contextual, and managerial aspects directly or indirectly involved with the term.

However, it cannot be disregarded that the variation in the application of metadata is wide and permeates among many areas of knowledge, with different approaches, contents, and purposes. This is also reflected in the metadata nature and definition and, in addition, constitutes an obstacle to the formulation of a universal definition, other than “a word with 8 letters”, as approached in Figure 5

Figure 5 -What is metadata?



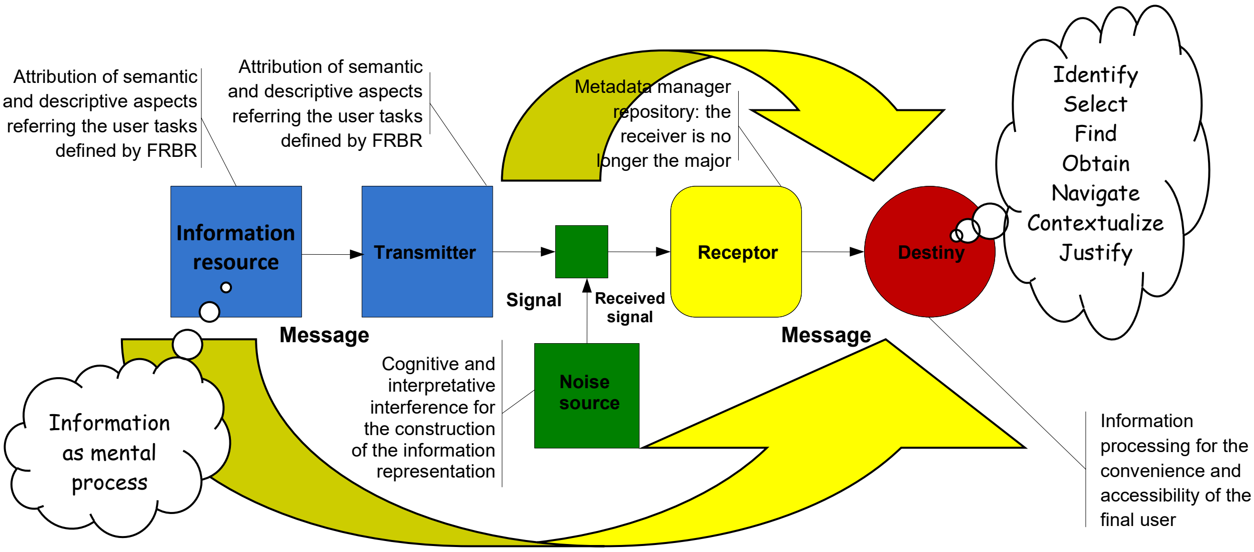
Source: Adapted from Geek and Poke (c2010?).

Considering the basal character of its conceptual understanding, the term metadata could be defined as the descriptive elements and structured qualifiers (electronic or not), considered indispensable for the identification, organization and findability of the information represented by the metadata. They consist of sets of values for the identification, description, contextualization, localization and administration of a resource of information, regardless the type of media in which it is stored, which allow the establishment of relationships with other data, entities and resources. Metadata corresponds to the most traditional way of representing information and its respective recording and storage supports.

The forms of registration, manipulation, storage and structure of the metadata are determinant for the information architecture, which will define the findability degree of the information which, by its turn, represents and integrates the set of attributes that justify the exhaustive discussions on the forms of metadata description and relations. In fact, such issues have been discussed by Information Science experts over the past decades. As a result of these discussions, there was adjusts on the understanding about the registration of information, focusing on its retrieval. One of these adaptations led to the understanding that the organization of information must prioritize, first, the researcher, as a member of the scientific community and, second, the information, as a mental and abstract process. In this scenario, the library community assumes a supporting role, as a manager and mediator of access to information, together with the information storage media, as sources of information.

Such understandings converge with the approach of the conceptual models of information representation (i.e., the FRBR Family[10]), whose recommendations represent trends in the ways of presenting and establishing relationships between metadata. From this perspective, the focus of the organization of information changes from the bibliographic catalog to its final user. In this context, it is not the registration rules, but the information recoverability degrees which enable the final user (i.e., the researcher), to perform the tasks of finding, identifying, selecting, obtaining, navigating and contextualizing the information in focus, as approached in Figure 6:

Figure 6 -The end-user in focus



Source: Adapted from Wurman (1995).

The application of metadata is decisive in the aspects of organization and information retrieval (be it digital or not) and, although they are attributed several uses and purposes, in Information Science they are responsible for enabling the organization, search, selection, identification, retrieval and, in some cases, the information access. In addition, the representation of information through metadata allows the exchange

[10] The Functional Requirements for Bibliographic Records (FRBR), initially published in 1998, is a recommendation of the International Federation of Library Associations and Institutions (IFLA), for the restructuring of catalog databases to reflect the conceptual structure of information resources (OCLC, c2018).

of information between two or more institutions. Such functionality favors the integration of diverse and heterogeneous information sources and allows a researcher to, in a single search, identify and retrieve documents that meet their information needs from different institutions (FUSCO, 2010). After all, as opportunely pointed out by Tennant (2001) “[...] isn't it true that only librarians like to search? Everyone else likes to find”.

6.2. The origins of the information registry

The registration of information is a human need, and its practice can be noticed since the beginning of human civilization. The human being has the intrinsic need and prerogative to think, to know, to register and to communicate, and the activity of registering serves not only to the concretization / fixation of knowledge in a particular support, but also to the need for communication and sharing of ideas, which is also inherent to human nature (BAPTISTA, 2007).

However, the definitions for the term information point to an abstract context, as the result of a mental process that indicates a set of knowledge. So, since information is an abstraction derived from human cognition, any form of registration or expression corresponds to the representation of a cognitive product. Inclusive, this fosters a confusing conceptual relationship between the terms document, information and knowledge. In this context, Alvarenga, (2003) argues that the representation made by the authors, at the time of expressing the results of their thoughts, is classified at primary level. Such thoughts, are derived from methodical observations of the nature and the social facts, using the available languages in the context of the knowledge production and communication. At this level of representation, the languages of different peoples and of specialties (i.e., the fields of knowledge) play a major role, including other iconic symbols and sound. After being produced, the records of knowledge contained in documents become part of the archives, libraries, documentation / information centers, and are then represented again (secondary representation), aiming at their inclusion in referential documentary systems.

In both cases (i.e., the primary and secondary representations), the representation comprises a cognitive process. When it culminates with the primary representation of knowledge, the representation stands out as an instance of the human cognitive process in accordance with the scope of the registration of thought in a documentary media, including the stages of perception, identification, interpretation, reflection and codification, which are all involved in the act of knowing a new being or thing, or delving into the knowledge of a being or something already known, using the senses, the emotion, the reason and the language. The beings exposed to the process of knowledge, that is, the beings about which one thinks, about which it is enunciated and about which a concept is built, integrate the essence of the field that philosophers call ontology: the universe of all beings existing (concrete and abstract) (ALVARENGA, 2003).

Alvarenga adds that the process of processing knowledge records for the purpose of storage in information systems, demands a new stage of representation, or a new approach to representation at the secondary level, this time not starting from the ontological being itself, but from the knowledge about the being, expressed in documents. In this sense, considers Alvarenga, secondary representation would have as a priority object not the collection of ontology of existing things and beings, but the collection of knowledge about these things and beings, which are objects of epistemology. Alvarenga still complements that in the primary representation, the final products are made up of concepts about beings (which forms the knowledge) and can be more or less intensely detailed and encoded through a symbolic language. As for in the secondary representation, which is an essential practice in documentary information systems, the same concepts contained in the primary records are succinctly identified among their fundamental constituent elements, from which is chosen the fundamental access points that guarantee the representation of such knowledge (document) for the purposes of future recovery. In this case, the concepts contained in the documents, as well as their emergency surfaces, constitute inputs for the secondary representation and should be identified, what requires that the information professional, in the process of knowledge organization, proceed to the identification of the elements of description and the thematic approaches that may be searched by potential users of the information system. Lastly, the author argues that in the professional performance of the treatment and organization of information, aiming at the intermediation between documents and users, archivists, librarians, museologists and other information professionals thus develop different types of representations, involving the replacement of primary information by specific records related to it, with a view to further recovery. In the catalogs and databases composing documentary information systems and services, the primary textual, sound, or iconic information is represented by new sets of information. In traditional technical processing, documents have been represented by a set of information related to its physical description and relevant access points (indexes), and such representation have been prepared and stored in a physical context regardless the primary document. The information in this type of representation comprises summarizations that attempt to describe the characteristics of the document, reflecting its origin and content and, facilitate its recovery (ALVARENGA, 2003).

Alvarenga's speech reinforces the understanding that the epistemological and ontological codification of knowledge in the form of metadata constitutes a new intellectual production that enables the mapping, management, organization, and retrieval of registered information. The author also points out that the human communication process is the major goal of knowledge representation, be it at the primary or secondary level, and points out that the representation at the secondary level can also involve the representation at the primary level, especially in those cases involving the production of a primary textual record about objects, sounds or image (be it real or virtual), that can integrate documentary collections.

Baptista (2007) highlights a third level of representation that can be considered as the retroconversion of the knowledge coded at the primary level again into a mental process, which is potentially responsible for the reproduction and multiplication of knowledge. The author proposes a model (Chart 2 - Knowledge representation model), based on the development and in the use of metalanguages, considering the need, not only for the representation of an object, but also for its conversion into informational resource, as a basic constitutive element of the knowledge.

Chart 2 - Knowledge representation model according to Baptista (2007)

Origin	**Primary representation (from the thought to the object)**	**Secondary representation (from the object to the registry)**	**Tertiary representation (from registry 1 from registry 2)**
Thinking as raw material	Concretization / fixation of thought in an object	Object representation	Record identification and location
Thinking as cognitive process	Passage from the abstract to the concrete	Passage from concrete to symbolic	Registration processing; from symbolic 1 to symbolic 2
Thinking as mental elaboration	Language as an expression of thought	Documentary language	Metalanguages
Thought as unrecorded information	Raw information	Registered information; transforming the object into an informational resource	Labeled information; virtual; controlled
Thought as individual knowledge	Externalized knowledge	Knowledge transfer	Multiplication of knowledge

Source: Adapted from Baptista (2007).

The development of knowledge and its ramification in numerous areas of specialization have exalted the importance of the representation of knowledge, aiming, above all, its communication and recovery. As a result, the preparation of records containing the descriptive representation of any supports is now made by different professionals, according to their interests, and from multiple locations: scientists; artists; companies; dealers; publishers; libraries; archives, museums, etc. (BAPTISTA, 2007).

Ironically, the perception concerning the importance of information representation took place in other disciplines, prior to in Information Science. According to Berestova, (2017), metadata emerged as a phenomenon following the advent of information and communication technologies (ICTs). However, the term became consecrated as a machine-readable resource only sometime later, with the rise of the digitization era. As a result, although possible, it is lost the meaning of relating the term information representation to the traditional formats in which the secondary sources of information were presented (i.e., bibliographic catalogs and bibliographies), since the use of the term metadata also emerged sometime later. Le Coadic (2004) points to the appearance of metadata as a new incarnation of language, affiliated with the management systems of sets of electrical signals (binary codes), which in fact, are products of Informatics. The author also points out that it is necessary to use the term information database: bibliographic, numerical, iconic (images) and not database or databank. The complementary relationship between Computing and Information Science is increasingly evident, to the point that Information Science becomes dependent on the former. This became more evident after the publication of the Functional Requirements of Bibliographic Records (FRBR), which suggest the information processing according to the establishment of relationships between entities and attributes. For instance, the entity-relationship concept was already consolidated and in full use in Computer Science at the occasion of the FRBR publishing. This evinces the relationship between Information Science and Computer Science and attests to the need for an effective approximation between both disciplines, aiming to simplify and expand access to scientific information. In fact, it is up to the Information Science to evince metadata not only as instruments of information representation, but as effective channels of organization and information access.

6.3.Metadata as secondary information sources

The typology applied to the term information is wide and varies: as to the publicity level (for instance, information can be public, private or confidential), as to the type of access, as to the format and storage support, as to the format of registration (digital or analog), as to the nature (scientific, technological, strategic), among other less usual characteristics. From a conceptual perspective, the typology of information sources is still further divided into primary, secondary and tertiary.

The primary information sources are understood to be all information presented in its original format, as conceived, and made public by the authors. Examples include journal articles, books, theses and dissertations, reports, standards, patents, laws, letters, speeches, photographs, works of art, among others.

The secondary sources present a systematization of the primary information, to facilitate its access, understanding and use. Include comments, analyzes and criticisms of primary information sources. Generally speaking, the secondary information sources are systematic information about primary information sources. Encyclopedias, dictionaries, bibliographies, indexes, files, databases, search engines, among others are examples of secondary information sources.

Tertiary sources of information are considered bibliographies of bibliographies. They are like guides for accessing primary and secondary information. The library itself as an institution is an example of a source of tertiary information.

Also, some publications that include both primary and secondary information, simultaneously. The boundaries between primary and secondary sources of information are tenuous and this subtlety has become even more pronounced as electronic publications have emerged. This is so true that some sources of information are sometimes considered primary and, other times as secondary, just as in the case of the catalogs.

6.4. Metadata as phenomenon and as a process

Metadata are phenomena, as intrinsic characteristics of resources. At the same time, they are processes, as tools for organizing information retrieval.

Metadata are inherent to every document[11], regardless the format, the type of support or any other attribute. Except for the metadata for the control and management of collections, the metadata concerning any publication are intrinsic to it and cannot be assigned to information resources at random, so that they are natural characteristics of the resources and of the information contained therein. For example, once published, the information resource is given immutable information about title, responsibilities, place (s) and date of publication. They are like natural phenomena, incorporated into the information resource.

As a process, the degree of importance and relationship established between the organization of information and metadata, is like that established between a building and its structural base (e.g., the foundation, the columns, and the beams). The

[11] In this paragraph, the term document is approached as any type of information support, including three-dimensional objects and people, considering that these are assigned qualifying elements, such as the identification information (e.g., the registration numbers in public agencies), the location (e.g., the address), the profession, and the place of birth.

metadata constitute the raw material for the organization of information and allows, the functionalities of the users' tasks as established by FRBR: Find, Identify, Select, Obtain and Explore, in the sense of structuring and making them feasible.

The way in which metadata are presented is determinant and directly proportional to the information findability indexes and is so important, that there are specific standards to guide its structure, value and content, such as the MARC21, the Dublin Core, the ISO 2709 and, the most recent, the UNIMARC and the Bibframe.

6.5.Metadata in the publishing market of bibliographic information

All bibliographic information contains metadata. The reverse is not necessarily true. This suggests that the systematic compilation of bibliographic metadata culminates in an arsenal of information that can contribute to the retrieval of information in scientific environments. As information technology matures, the focus shifts from machines to information itself. The value of technology lies in its ability to manage and exploit the product, that is, the information (WURMAN, 1995).

This justifies the efforts of scientific publishers to develop increasingly accurate, elaborate and semantic research tools. The perception of the commercial value of the information potentiated the focus change from the machines to the information itself and instigated scientific content publishers for the production of metadata compilations aiming to facilitate the identification, the search and, in some cases, the access to scientific contents. Such compilations correspond to the databases or as opportunely pointed out by Le Coadic (2004), the bibliographic information databases.

If, on the one hand, the disclosure of such a market niche for publishers added value to information, on the other hand, it contributed to increasing both the access to information and the quality of the bibliographic search instruments.

It is also worth mentioning that the bibliographic information databases, correspond to systematically structured compilations of bibliographic metadata which may or may not allow the access to the represented content in full. Regardless the type of access (e.g., referential, partial, or full text), it should be highlighted that the authorship of the compilation and the structuration of metadata within bibliographic databases marketed by publishing companies, are not even mentioned within such bibliographic databases nor in any other source as a means of strengthening their brands and achieving pharaonic financial goals.

And all in all, the role played by bibliographic information databases, commercialized by scientific content publishing companies and by cataloging databases, usually designed and maintained by libraries, are similar, in the sense of enabling the execution of the user's tasks suggested by the FRBR (i.e., Find, Identify, Select, Obtain and Explore). Perhaps the greatest distinction in this respect is the need for providing financial investments to publishers as payment for the temporary access to such commercialized bibliographic databases, whereas the cataloging information databases, are usually made available to the public for access and use, free of charge. (Un)Fortunately, librarians do not seem to have realized that.

The control, the description, the certification, the identification, the location and the access are the most common functions attributed to metadata, regardless their purposes, be it commercial, scientific, technological, administrative or data exchange, while the applicability aspects are directly related to the form they are presented and the type of information they represent.

6.6.Metadata typology

Metadata has always played an important role in organizing and retrieving information. Such importance was directly impacted by the improvement of the technological resources which enhanced the expansion of resources for the description, the management, the retrieval and, more recently, by the introduction of the concept and the practice of interoperability within bibliographic catalogs and the establishment of relationships between documents and the data representing them, even if allocated in different physical or digital environments. In this context, metadata assume and perform the functions of:

> • certifies the authenticity and degree of completeness of the content;
> • establishes and documents the context of the content;
> • identifies and exploits the structural relationships that exist within and between information objects;
> • provides a range of intellectual access points for an increasingly diverse range of users; and
> • provides some of the information that an information professional might have provided in a traditional, in-person reference

or research setting (BACCA, 2008, p. 6).

The production and use of metadata are becoming increasingly common in people's lives (in addition to researchers). If only a few years ago only academic and research institutions, were considered providers of information who had autonomy for the generation of metadata, nowadays, therefore, such practice is becoming more and more popular. For instance, even people with low educational levels usually assign tags to their publications (posts) in the social medias. Such tags, or hashtags, as they are popularly known, are words or phrases preceded by the "hash" symbol (#) which performs the function of identifying messages on a specific topic. Hashtags are used as a way of converting feelings, behaviors, geographic locations, and other attributes into searchable elements. Even if such resource is used by people who may not even know the function of such tags, the role played by them in the information organization cannot be denied. So, thanks to their representative character, hashtags can be considered as a variation of metadata.

The processing of information can be approached from numerous aspects, which can vary according to the type of information being considered, the interests of its generating and/or holding institution, its informational context, its targeted audience, etc. Such aspects can assume several purposes e.g., administrative, political, budgetary, statistical, and similar. So, it is important (and in some cases essential), the registration of metadata referring to the origin (acquisition mode), location, indexes of access and use. Chart 3 approaches existing types of metadata, their definitions and examples of where can be used.

Chart 3 - Types and functions of metadata

Metadata types	Definition	Examples
Administrative	Metadata used to facilitate the management and administration of collections and information resources (objects)	• Acquisition data • data on when and how the resource was created • Access and use restrictions • tracking of use and reproduction rights • Legal access documentation • Location information • Selection criteria for scanning
Descriptive	Metadata used to identify and describe collections and related information resources. They refer to the purpose of discovering, identifying, and selecting information.	• Catalog records • Search features • Differentiation between versions • Specialized indexes • Curatorial information • Hyperlinks of relationships between resources • Notes from authors and users
Preservation	Metadata related to the preservation management of collections and information resources (objects)	• Records concerning the physical condition of resources • Record of actions taken to preserve the physical version of the resources, e.g., data update and migration • Recording of any changes occurred during digitization or preservation processes
Technical	Metadata related to the functionality of systems or their behavior	• Hardware and software registration • Technical information on digitization, e.g., formats and compression rates • System response time record • security and authentication data e.g., passwords
Use	Metadata related to the rates and types of use of collections and information resources (objects)	• Circulation records • Physical and digital display records • Usage andu ser data tracking • Reuse of content and versions • Search logs • Metadata rights
Structural	Metadata that can be considered as connectors of digital objects. They relate to usage metadata, considering their purpose of allowing the use of a particular entity, however, the usage metadata are intended for human use while structural metadata are generally used in processing by machines and provide the smooth functioning of a system.	• Indication of the pages ordination within a book to form chapters • Visual structure of a digital environment
User-created	Metadata created by the author himself to identify, retrieve, categorize, and promote Web contents.	• Folksonomy

Source: Adapted from Bacca (2008); Caplan (2003); Lourenço (2007).

A precarious technological structure does not always result in the non-finding of information. Conversely, the opposite is not true. Poorly structured metadata usually impairs the retrieval of information. The metadata structure is important, even in environments where the information organization is manually controlled. There is no need to talk about "right or wrong metadata", as this approach varies according to the purpose of the record, its conceiving context, its storing, its retrieving and use (including the target audience). In addition, metadata can be presented in several formats, according to the guidelines of the standards adopted for their registration. For instance, the Chart 4 – Typology of metadata standards, show the most used types of metadata.

Chart 4 - Typology of metadata standards

Metadata formats	Examples
Data structure patterns (sets of metadata elements, schemas). are "categories" or "containers" of data that compose a record or other information object	The MARC fields set (Machine-Readable Cataloging Format), The Encoded Archival Description (EAD), The Dublin Core Metadata Element Set (DCMES), The Categories for the Description of Works of Art (CDWA), The VRA Core Categories
Data value patterns (controlled vocabularies, thesaurus, controlled lists). They correspond to the terms, names and other values used to fill data structure patterns or sets of metadata elements.	The Library of Congress Subject Headings (LCSH), The Library of Congress Name Authority File (LCNAF), The LC Thesaurus for Graphic Materials (TGM), The Medical Subject Headings (MESH), The Art & Architecture Thesaurus (AAT), The Union List of Artist Names (ULAN), The Getty Thesaurus of Geographic Names (TGN), The ICONCLASS
Data content standards (cataloging rules and codes). They correspond to the guidelines for the format and syntax of the data values used to fill in the metadata elements.	The Anglo-American Cataloging Rules (AACR), The Resource Description and Access (RDA), The International Standard Bibliographic Description (ISBD), The Cataloging Cultural Objects (CCO), The Describing Archives: a Content Standard (DACS)
Standards of data format / technical exchange (standards of metadata referring to the machine-readable format). This type of pattern is often a manifestation of the data structure pattern (described above), in a coded or marked format for processing by machines	The MARC21, The MARCXML, The EAD XML DTD, The METS, The MODS, The CDWA lite The Simple Dublin Core The XML Schema, The Qualified Dublin Core XML Schema, The VRA Core 4.0 XML Schema

Source: Adapted from Bacca (2008).

Such standards make possible the interoperability and findability of common elements between the registries (establishment of relationships). The more consistent and structured are the metadata, the greater the information representativeness, and the interoperability and the possibilities for research, manipulation, and interrelationships with other information objects[12]. In contrast, some types of tasks may require one or more specific types of metadata, which in some cases includes aspects about the content and the completeness in the description of the data.

The care in structuring and describing metadata results in better information management practices and better access experiences for the researcher. However, it is up to the documentalist to be sensitive to perceiving the limit between the enough, the necessary and the excess. This requires some knowledge concerning the characteristics and information desires of the public to whom the information resources being represented are destined (BACCA, 2008).

6.7.Untold truths about metadata

a) Metadata has authorship

“Cataloging is an art, not a science. No rule can replace the experience and the common sense, but some of the results of experience can be guided by standards” (CUTTER, 1904, p. 6). Although Centenary, Cutter's observation could not be more up to date. However, the scope of his discourse could be expanded, as the experience and the common sense are not only applicable to cataloging, but to any human activity area, especially those concerning information management and organization. But the purpose of citing Cutter on this occasion was to highlight the personal and interpretive character intrinsic to the information representation, more specifically to the descriptive representation, both in the form of catalogs and in the form of bibliographic references. The representation of information implies the interpretation and the ability to reproduce the represented content or image, as a decoding-encoding process. Since information representation is the fruit of a cognitive and interpretative process, it can be considered a new product of the intellect. Despite of this, the arrangement of metadata is not usually attributed to any authorship, which can be an inconvenience, considering,

[12] An information object is a digital item or a group of them, regardless of type or format, which can be considered or manipulated as a single object by a computer. However, such a concept can become confusing in the case of use to refer to both digital substitutes of the original object or items (for example, scanned images of works of art or cultural materials, a PDF of an entire book) and, for descriptive records relating to objects and / or collections (for example, catalog records or help resources).

above all, that the retrieval (or not) of the represented information is directly related to the veracity and consistency of the recorded metadata. Talking of the usage and/or the administrative metadata, they can be automatically conceived, as is the case of circulation indexes concerning a particular resource. However, even though the counting of statistical indexes is conducted automatically, their compilation and interpretation depend on the human intellect and, before that, the programming of the equipment to compile such data is also the result of human intellectual effort. Thus, whether by merit or demerit, the indication of the responsibility (authorship) for the creation of metadata would be a form of recognition for the work performed and, at the same time, an incentive to the public commitment to the reliability in the representation of information.

b) Metadata are not necessarily digital

And more than just that. In mid-1995, following the promotion of the Dublin Core Metadata Element Set, two phenomena occurred: first, the strengthening of relations between the library and the Web communities (Information Science and Informatics), including the migration of concepts and terminology from information technology to Information Science and second, the librarian’s perception that their production of data about data (metadata), in the form of bibliographic catalogs, was centuries old. (CAPLAN, 2003). Therefore, those who attribute the concept of metadata exclusively to the digital environment are mistaken. And, amazingly, the International Federation Library Association (IFLA), seems to be included in the list of “mistaken” entities, considering its statement that “metadata is data about data. The term refers to any data used to assist in the identification, description and location of electronic resources in a network” (IFLA, 2005). Considering, metadata as representation of information, it is certainly possible to state that the practice of descriptive representation is ancient, and arose even before the writing itself, and the rock inscriptions can support such understanding. Later, the catalogs appearance in ancient libraries, indicated not only the need for representing, but also the concern about the location of information. And even today, metadata is present from the most everyday environments, such as traffic signs, to environments controlled by information technology resources, such as initiatives based on the semantic web and linked data. Thus, considering even the guidelines suggested by the conceptual model Functional Requirements for

Bibliographic Records (FRBR), metadata can be considered as the description of the entities related to an information resource, regardless of the type of support they are on: printed, digital, three-dimensional, audiovisual or any other, without even addressing the issues of forms of access: whether analogue or electronic, via the network or via the Web, online or offline.

c) Metadata represents more than "just" the description of an object

As discussed in section 5.6 - Metadata typology, the description is one of the several functions performed by metadata. Despite this, there is a tendency, mainly within the confines of common sense, concerning the association of the concept and the application of metadata, to the documentary description, exclusively when, in fact, metadata can make possible the contextualization and the identification of the content of a document, inform about usage rates, acquisition modalities, conservation status and possible conservation and restoration processes that the object may have passed through, its location and additional and/or related content.

d) Metadata may stem from several sources

Metadata can be manually written, manually or automatically collected through databases, or even automatically generated (BREITMAN, 2007). Metadata can be provided by humans, extracted from large databases, or obtained automatically. A major project on the Internet today is the self-generating information portals, capable of updating their information automatically (BREITMAN, 2007).

e)Metadata are dynamic and modifiable across its useful life

Yes, metadata have a useful life period and besides, they evolve during the useful life of the information system or the object to which they refer. As stated by Breitman (2007), metadata are created, modified, and even discarded during the useful life of the resource to which they refer.

f) Metadata can be simultaneously assigned to more than one information object with different meanings

> "One information object's metadata can simultaneously be another information object's data, depending on the kinds of aggregations and dependencies between information objects and systems. The distinctions between what constitutes data and what constitutes metadata can often be very fluid and may depend on how one wishes to use a certain information object" (BACCA, 2008, p. 14).

g)Metadata as database registers gatherers

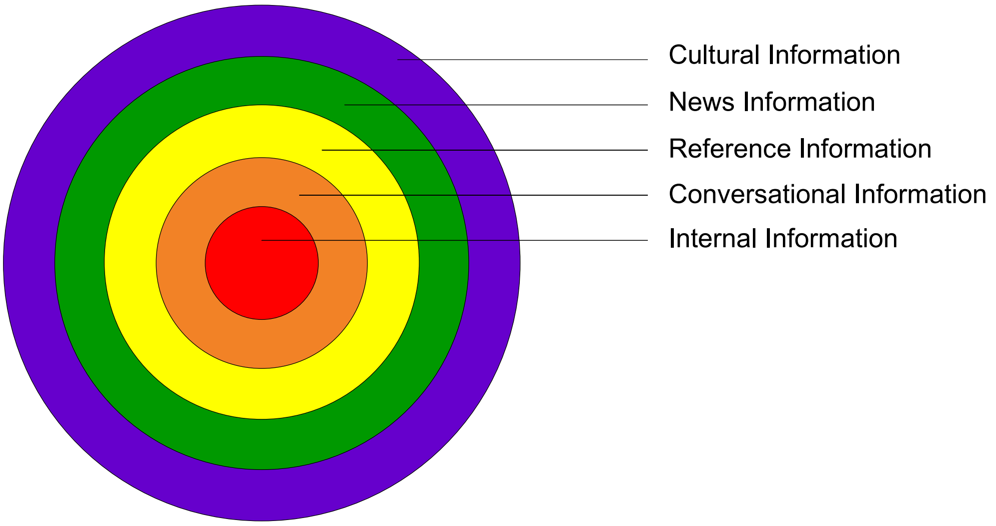
Metadata can also make it possible to search across multiple collections or to create virtual collections from materials that are distributed across several repositories – but only if the descriptive metadata records are the same or can be mapped across all the collections (BACCA, 2008, p. 14).

6.8.Metadata as a mitigating element of information anxiety

Ambiguity is one of the characteristics of the term information, to which can be assigned different meanings, in the most varied contexts. Despite this ambiguous scenario, "information has become the most important word of our decade, the livelihood of our life and our work" (WURMAN, 1995). This fostered the explosion of non-information (data explosion), which cannot always be considered information (considering it as what it informs, which has intrinsic value). This reaffirmed the commitment of Information Science in the treatment, aiming at the management, organization, contextualization (semantics) and promotion of access to data.

Wurman (1995) proposes the division of all information acting on human life into five levels, as visually represented in Figure 7:

Figure 7 - The five information rings according to Wurman (1995)



Source: Adapted from Wurman (1995).

The first ring represents the internal information. It corresponds to the set of brain information that governs the human vital system. This is the least controlled level of information. The second ring represents the conversational information. It is the form of oral communication that allow the exchange of information between people. This is perhaps the most controlled level of information. The third ring, the reference information, can be anything (a manual or a dictionary, for example). It consists of all the information that operates the systems of Science and Technology and human life. The fourth ring, news information, consists of current events, which do not necessarily affect an individual's life, but which have the potential to alter his worldview. The fifth and final ring, cultural information, represents the least quantifiable and most abstract form of information. It corresponds to the set of information that comes from the other rings which determines beliefs and behaviors and the society nature as a whole.

The representation of the levels of information in the form of rings, as proposed by Wurman, indicates the urgency levels about human life, but they are not intended to suggest the existence of hierarchical levels between them. On the other hand, such information is related to each other, so that the interchange among the five rings can generate new information for themselves or for the other rings. For instance, a conversation (conversational information), can encourage or support the writing of a document, which in its turn can be considered as reference information or news information, depending on its content and context in which it is used.

Still according to Wurman (1995), both the excess and the lack of information can culminate in information anxiety, which can act at any of the levels represented in Figure 7. The author states that there are several situations that usually cause information anxiety: not understanding the information; feeling overwhelmed by its volume; not knowing if a particular information exists; not knowing where to find it; and, perhaps the most frustrating, knowing exactly where to find it, but not having the passkey. You are sitting at your computer, which contains all the listings to justify the money you are using to develop a new product, but you cannot remember the name of the file. Information remains on the threshold and out of reach (WURMAN, 1995).

Wurman (1995) continues: you are trying to describe yourself as a wine lover, but you have no idea how to spell “oenophile”. In fact, dictionaries are very useful if you know how to spell, but if you cannot remember how the word starts, nothing doing. This is the nightmare of inaccessibility – trying to find something without knowing the topic to which it relates. How do you ask about something if you do not know what it is called? This is information anxiety. We are surrounded by reference materials. But, without the ability to use them, they are only sources of anxiety. Wurman finally, states that he thinks of them as buddhas sitting on his shelf, with all the information and a smile of wisdom. For him, the challenge is to gain access to them and make them more accessible to others. (WURMAN, 1995).

Wurman lists causal situations for information anxiety, through a discourse marked by expressions that make evident his concern with finding and accessing information, like: feeling overwhelmed by its volume; not knowing if certain information exists; not knowing where to find it; knowing exactly where to find it, but not having the access key; information remains on the threshold and out of reach; and, this is the nightmare of inaccessibility.

The first version of Wurman's text was published in English, in 1989. Perhaps at that time, the organization of information was not as challenging as it is today, since the proportions of the volume of information to be processed were smaller at that time, in comparison with the current scenario. Perhaps the inefficiency of information systems would further fuel the author's concern. However, the improvement of technological resources, at the same time that it favored and facilitated the processes of organization and information retrieval, it also potentiated the exponential generation of new data, which represented the information, reaching pharaonic volumes. And as Wurman rightly pointed out, not only the scarcity, but the excess of information generates information anxiety.

Once, in a placard, the following words were seen: “data is the new tapioca”[13],[14]. The metaphor illustrates, satirically, the scenario in which data are everywhere, contained in any context, and applied in the most different purposes. Tapioca is mentioned here as a connotative term for the popularity and proliferation of data, which are neither self-organizing nor self-retrieval.

Among data, tapiocas, buddhas and the like, Information Science found in technology the antidote to “remedy” or at least minimize the effects of information anxiety: the metadata.

[13] Tapioca is the starch extracted from cassava, usually prepared in granulated form. It is the main ingredient of some typical Brazilian delicacies.

[14] The expression “data is the new tapioca” is also being used as a banner in a website of a Brazilian big data startup, available from https://www.99jobs.com/semantix.

Méndez Rodríguez (2002) points out the traditional roles of metadata: identification and description of information, search and retrieval, and location of documents. The author also adds some emerging roles to the traditional ones like, the forms of access, the valorization of the content, the visibility of information and the accessibility of documents, among other functions. He also indicates that the functions of metadata can be focused on two aspects: the technological approach (system) and the researcher approach (user). From the system approach, metadata, enable interoperability and data sharing between different systems, particularly when applied together with the formats of information representation and conceptual models. From the researchers approach, metadata shows the location and the forms of access, in addition to its usefulness, which corresponds to the contextualization.

Such a scenario favors the understanding that the metadata contributes to minimizing the effects of information anxiety, considering above all, the correspondence with the researchers' expectations, especially regarding electronic information:

a) The user needs to recognize the existence of the resources, that is, recover and retrieve the sources of information.
b) A user who uses the network for research needs broader information concerning a particular resource, that is, to know if it will be useful in his Professional and/or research context. He needs to know the source, the history or the intellectual responsibility of the source, its integrity and authenticity, the relationships with other resources, the specific characteristics of the domain to which such information applies, etc. he also should know whether a resource fits different levels of use; whether he can extract the content using any specific tool or application; and know what kind of publication it is (if the document is a manual, a scientific monograph, a dissemination work, etc.).
c) As a further information provider in a constant feedback process in the production of knowledge, the user should also know the intellectual property rights assigned to the requested information and what level of use or reproduction he can make of it.
d) In addition, the user, as an information consumer, needs to know under what conditions he can access a particular information object, like the need for pay for its access and under what conditions and terms he should do so. This client-

user must also know the technical interfaces, the access protocols, the search types or formats, etc., allowed by the information system that he uses.

e) Lastly, the user also needs a value judgment on the contents of a particular source of information, for example, if you include materials that you are likely to consider unsuitable for minors. (MÉNDEZ RODRÍGUEZ, 2002).

Except for the concerns related to the understanding of the information content itself, Méndez Rodrígues suggests that metadata meets the researchers' information needs in both the descriptive and the contextual contexts by stating that the metadata increases the likelihood that the users will be able to retrieve the information best fitting to their search queries and that they can assess its relevance, usefulness, and accessibility. The use of metadata for providing a well-structured and rigorous description of the web resources can increase the accuracy of the search and its order of relevance, in such a way that allows greater automation of the queries and dispenses the manual validation of the results presented by the information retrieval systems (MÉNDEZ RODRÍGUEZ, 2002).

Metadata plays the role of interpreters between human cognition and machine language. Well-structured metadata contributes to the satisfaction of the researchers' interests concerning the identification, description, contextualization, and location of information and contributes to the reduction of uncertainty, which is practically the same as saying that they contribute to the reduction of anxiety about information.

6.9. The semantic character of metadata

There are differences between what is meant and what is actually said; between what is said and what others hear; between what people hear and what they understand; between what they understand and what they remember; between what they remember and what they can relay. People only listen to what they want and how they want it, according to their own experiences, paradigms and pre-judgments. There is information that individuals do not perceive and do not see; information that they see and don't care about; information that they see and don't understand or don't decode; information that they see and don't use; information that they seek; information that they guess information. The state of mind and the mood can affect how people handle information (ANGELONI, 2003; DAVENPORT, 1998).

In the context of representing information in the form of metadata, there should be no differences between the meaning registered in the primary source of information and the content described by the metadata, so that they (the metadata) must correspond to the pure and faithful representation of the represented object and of its content.

The representation of information does not admit differences between what is being represented and what is represented, between what is represented and what is perceived, between what is transmitted and what is actually understood. It is like in the popular joke “broken telephone”, in which the objective is that the message generated by a sender is understood by the receiver, after passing through other agents, as it was generated by the sender. It is evident that in the scope of information representation, this requires a cognitive construction, through which should be determined the points of access to a work, the identification of varying forms of authorities (people, entities and events), the selection of primary terms (authorized access points) and secondary terms (remissive access points), the thematic delimitation (indexing) and the preparation of abstracts, in addition to the description itself. Such arguments support the understanding that the set of metadata composing the descriptive records, the bibliographic references and others, also constitute an intellectual production, corresponding to the result of the interpretation and the appropriation of the context and meaning of the information being represented, although is not common the assignment of authorship for metadata.

Despite the importance of the attribution of meaning in the data value assignment and the technology contribution in this process, Wurman (1995) criticizes the human tendency to demonstrate so much consideration for computers, to the detriment of humans and adds that the better information processing can result in increased data flow but give very little assistance on examining the results for a search, decide what to do with it or find a broader meaning. Meaning requires meditation, which takes time, and the pace of modern life works against the idea of giving us time to think. (WURMAN, 1995).

In the face of such notes, the development and the enhancement of technologies for metadata and document management, the semantic treatment of data, and the promotion of the access to them, especially considering the explosion of non-information (explosion of data devoid of the meaning attributed by cognition), are challenges for Information Science. In such a scenario, the fidelity in the description, in the contextualization and in the attribution of semantic aspects to the metadata becomes a commitment and a way in which the metadata-producing agent demonstrates respect for the information systems and, consequently, for the users of such metadata and for the authors of the information represented.

Considering that “data is the new tapioca”, the semantic character of metadata is established prior to the use of ontologies or any other vocabulary control language. An example that highlights the responsibility for assigning value to information is the slogan widely used by a food company in the 1990s in their advertising campaigns: “Image is nothing. Thirst is everything. Obey your thirst”, as reproduced in Figure 8.

Figure 8 - Advertising campaign ran on communication channels in the 1990s

Sources: Singhi (2017) and Quem disse (c2021).

The advertising campaign was widely broadcast on television and in the print media, and the mention of the slogan on the agenda was presented with a dark background, which denoted the mood of drowning, followed by another image, with the advertised drink, which referred to the idea of freshness and satiation of the thirst feeling.

The Figure 9 refers to a video published in 2013, presenting a satirical version of the advertising campaign above, is an example of the responsibility for assigning value to information through metadata. The video reproduces the idea of the original advertising campaign; however, the advertised drink is replaced by cod liver oil. Until the scene showing the label of the bottle containing the liquid used in the filming (cod liver oil), the assistant is led to the clear impression that the video reproduces the image of the drink originally announced being served in a glass with refreshing ice cubes, while a female voice pronounces the following words: "drink what you think is good, because television can make anything look delicious".

Figure 9 -Video scenes aired on YouTube

Source: Fernandes (2013).

Similarly, in the universe of Information Science, the format and the structure used for presenting metadata can induce the researcher to misperceptions about the document and/or the information represented or referred by metadata. In Librarianship, there is a classic example of the work authored by Sérgio Buarque de Holanda, entitled "*Raízes do Brasil*" (Roots of Brazil), about Brazilian historical aspects. In this case, the vagueness of the title promotes ambiguities in the understanding of the subject addressed by the work, whether History or Botany[15]. As a consequence of such misinterpretation, the metadata attributed to such work may not correspond to its content and, therefore, its retrieval possibilities may be reduced, at least in the context it was expected to be included. It is worth reminding that the establishment of (proper) relationships across records representing documents, as suggested by Functional Requirements for Bibliogaphic Records (FRBR), is directly dependent on the consistency of the metadata assigned to them. This example reinforces the need for attention to the semantic characteristic of the information representation through the metadata use, which has not replaced but, in contrast, complemented its bibliographic functions.

Assigning meainings to data, to the information and to the knowledge (no, they are not synonymous!) is not as simple a process as it seems. Individual characteristics, which form the mental model of each person, interfere in the encoding / decoding of these elements, often resulting in individual distortions that may cause problems in the communication process (ANGELONI, 2003).

---

[15] In Portuguese, the word "roots" may assume several meanings in various disciplines. One of them refers to the plant organ that perform the main functions of serving as a means of fixation to the soil and to absorb and conduct water and minerals (referring to Botany). The other refers to an event or the existence of something, referring to the source or the origin (referring to History). (HOUAISS; SALES, 1999).

Angeloni (2003) also states that the biggest challenge is not to obtain the data, the information and the knowledge, but the acceptance that distortions occur in the encoding-decoding process, and that there are ways to mitigate it. The author exemplifies people's interference in the encoding-decoding process and the distortion in the transformation of data into information and of information into knowledge by the following fact. Different people facing the same fact tend to interpret it according to their mental models, which lead them to perceive it differently. For example: a BMW car, the latest type, convertible, zero kilometer, totally destroyed in an accident in which the driver hit a centennial tree knocking it over, can be encoded-decoded and distorted in the following ways: some people will be led to decode the information based on their material values: “What a pity, this is such an expensive car! Is he insured?” While other people, more susceptible to human values, will focus on the human being: "Did the accident result in injuries?" Other people with ecological interests will still have their attention focused on the fate of the centennial tree: “Why this tree? Couldn't it have been in another BMW?”. Angeloni concludes that being aware of these and many other interferences in dealing with data, information and knowledge in the decision-making process is the first step to mitigate them. (ANGELONI, 2003).

Data are raw and meaningless elements, which are disconnected from reality (ANGELONI, 2003). However, the researcher's interest is not the metadata itself, but the resource it represents. In this context, the documentalist's responsibility is divided into two aspects: the first, refers to the reliability in the representation of information, so that the perception of the researcher when consulting an information record should correspond to the closest possible to the reality. The second aspect, refers to the identification of what type of information should be represented, considering the different types of metadata which can be linked to a single information resource and that, not all of them are of interest to the researcher and/or will be effective in the representativeness in the retrieval of the resource.

Another aspect not to be disregarded: metadata, in addition to representing a document or a resource, also enable the functionalities of the Semantic Web. In the digital environment, the use of metadata, combined with the functionalities of ontologies, allow the connection of different resources for the simple fact that they are “the same thing” or represent “the same meaning”. This provides a wide and linear navigation experience (GILLILLAND, 2008). In the end, the essence of this discussion lies in the understanding that metadata is not the universal solution for discovering resources in the digital environment (GILL, c2016), but they are the main inputs for descriptive representation and, if it were not for them, the organization and information retrieval, would be absolutely intangible.

7. THE METHOD

Data supporting this research were extracted from The SCImago Journal & Country Rank which is “an open access scientometric directory” (GUERRERO-BOTE; MOYA-ANEGÓN, 2012, p. 675) which was launched in 2004 (SHOTTON, 2018) and “that includes the journals and country scientific indicators developed from the information contained in the Scopus® database” (SCIMAGO, c2020). SCImago covers the 27 major thematic areas divided into 313 specific subject categories (also referred to as disciplines in this study), comprising over 34.000 journal titles from more than 5.000 international publishers and country performance metrics from 239 countries worldwide (SCIMAGO, c2020). ScImago is one of the most authoritative sources of citation data (SHOTTON, 2018). The selection of this database for this study is supported (justified) by the aforementioned statements, added to the expressive numbers of 31971 journals titles ranked and also by the detailed analysis of the metrics data for each journal title, what includes the index of citations received by each journal in the previous 4, 3 and 2 years.

SCImago’s thematic scheme is based on The All Science Journal Classification (ASJC), developed by Elsevier and adopted by the Scopus database. Both the SCImago and the ASJC were considered by this analysis for grouping journals according to thematic categories and subcategories to support approaching related disciplines journals data. We considered at least two journal titles by each subject area.

The total number of journals composing SCImago database (24702 journals) at the date of the data collection were converted in percentages. This made possible to identify the representativeness of the total number of journals grouped in each SCImago subject area in relation to the total number of journals integrating SCImago database.

What follows is the methodology adopted to select the journals and articles to sample composition and, the criteria adopted to the qualitative and quantitative analysis of journals, articles, mentions, quotations and bibliographic references.

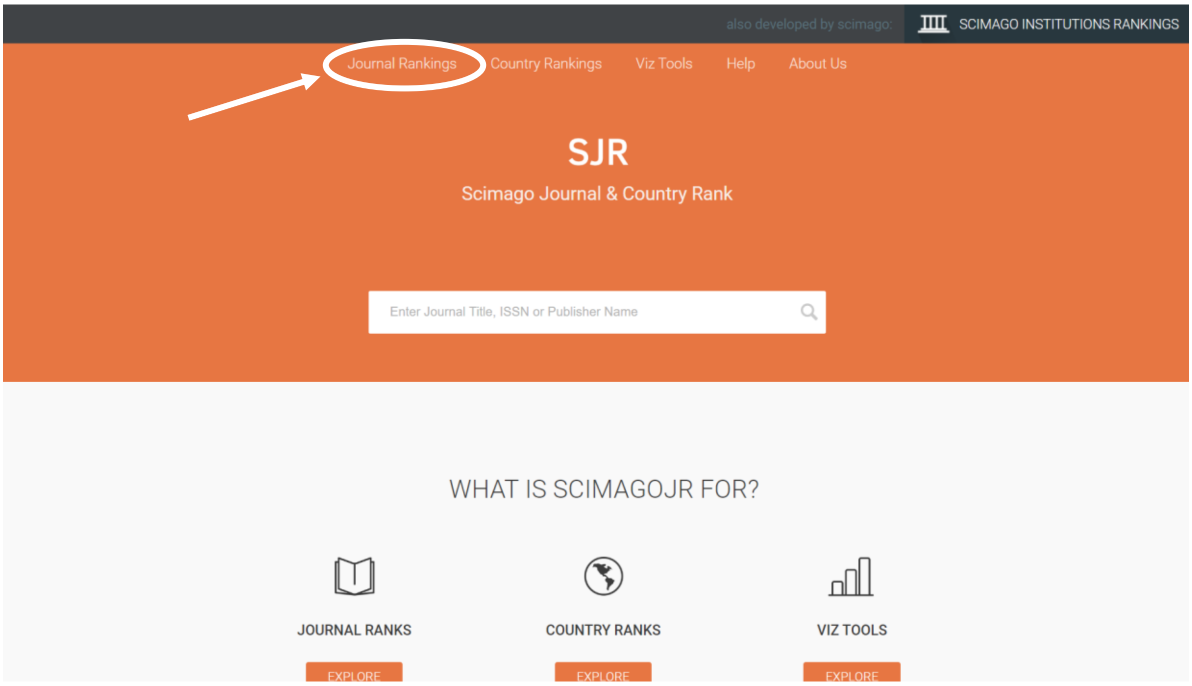
7.1.The journal selection procedures

The procedures considered in this section are part of a protocol for selecting journal samples, developed to the purposes of attending the needs of this research. The full protocol is available from Santos, Peroni and Mucheroni. (2020c) and the data collected from the recommended parameters were gathered in an Excel file, which is available from Santos, Peroni and Mucheroni. (2020b). Such raw data also is available in Santos (2021).

The first step comprised an overview of the SCImago database arrangement. This analysis showed that Journal titles are classified and subdivided under 27 subject areas which will be integrally considered in this methodology: Medicine, Social Sciences, Arts and Humanities, Engineering, Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Computer Science, Mathematics, Environmental Science, Business, Management and Accounting, Psychology, Materials Science, Earth and Planetary Sciences, Physics and Astronomy, Economics, Econometrics and Finance, Chemistry, Pharmacology Toxicology and Pharmaceutics, Nursing, Chemical Engineering, Neuroscience, Immunology and Microbiology, Health Professions, Energy, Decision Sciences, Veterinary, Dentistry, and Multidisciplinary.

Once the 31971 publications titles considered by SCImago database compose a too wide universe to be fully considered by any research, it was necessary to determine a representative sample for the total amount of indexed journals. For determining this sample, we carried out searches at SCImago database, available at SCImago (c2021), according to the steps showed in Figures 10 and 11:

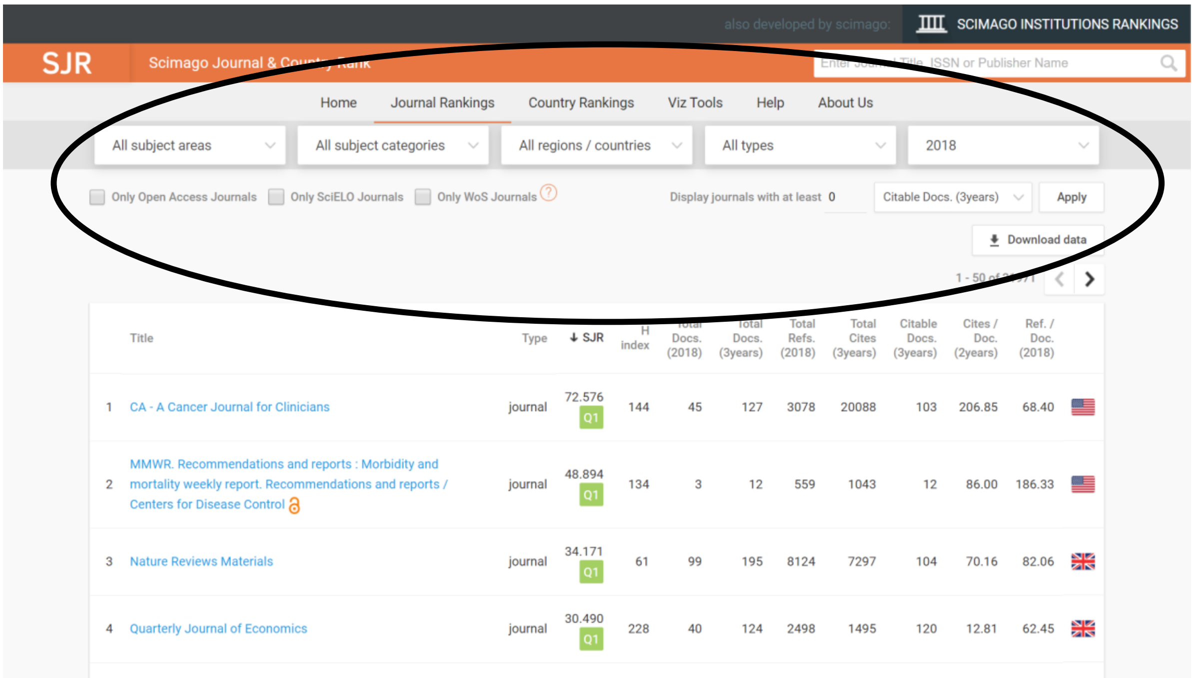
Figure 10 -SCImago database homepage



Source: SCImago (c2021).

By clicking at the “Journal Rankings” at SCImago homepage, the user will be led to the rankings search page (Figure 11):

Figure 11 -SCImago database searching page



Source: SCImago (c2021).

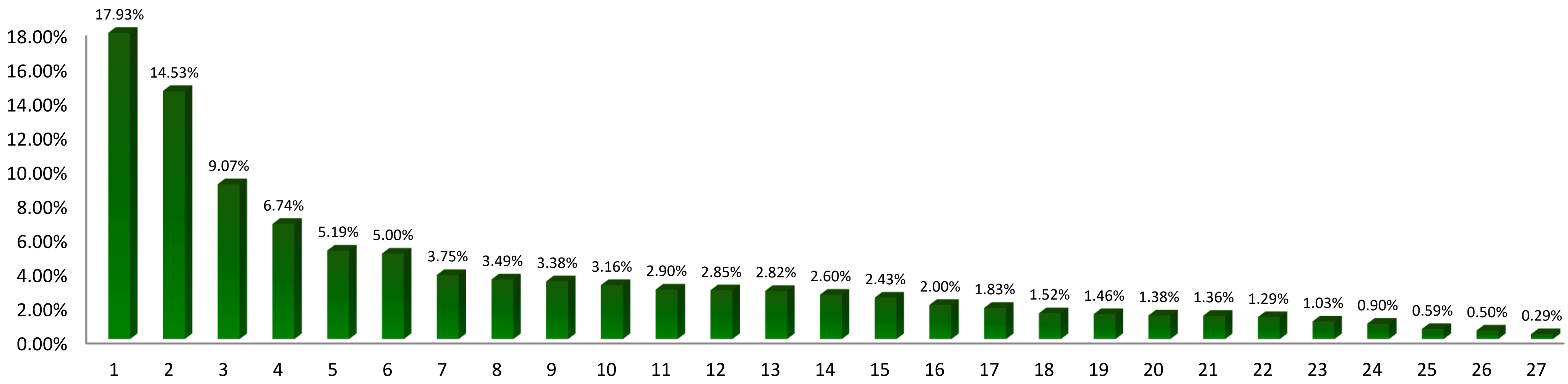
The searches considered journals from all subject areas, including all subject categories and from all regions ranked in the year 2018. This assured a representative range of data that made possible to have a broad view from citing and referencing perspectives. In the sequence, the following search criteria was be applied to the searches:

a) Subject area: once all the subject areas should be represented in the sample, we carried out 27 searches, one for each of the subject area composing SCImago database.
b) Subject category: the subject category is a criterion which was considered in a whole at this moment. So, the “all subject categories”, selected by default by the database, was maintained.
c) Region/country: The method’s purpose was to provide a broad view of the research universe, represented by the sample. Facing this, the “all regions/countries” option was selected.
d) Type of publication: The SCImago database ranks several types of publications: journals, book series, conferences and proceedings and trade journals. For the purposes of this study, only the “journals” option was considered.
e) Year: we considered the most recent released journal ranking, which, in the occasion of the data collection, was the year 2018, considering the data collection period in November 2019.
f) The selection boxes “Only Open Access Journals”, “Only SciELO Journals” and “Only WoS Journals” were left blank in all searches, because for the approach of this method there was no interest on evaluating the form of access or the indexing source of the journals.
g) Finally, for the last “selection box” we chose the option “Citable Docs. (3years)”. This limited and ranked the search results among the citable documents published by the journals in the previous three years.

Since these search parameters were delimited, the search results were displayed after clicking the “apply” button. The results provided by SCImago database were exported into a Microsoft Excel spreadsheet, which allowed us to take advantage of the tools used for separating, filtering, and selecting data.

This search resulted in 24702 journal titles, distributed into the 27 SCImago subject areas. Graphic 1 represents the percentual distribution of journals composing SCImago database, according to the adopted subdivision by subject areas.

Graphic 1. Percentual distribution of journals across the 27 subject areas subdivision of the SCImago Journal and Country Rank



Total amount of journals ranked in SCImago database, by subject knowledge area

1	Medicine	7224
2	Social Sciences	5855
3	Arts and Humanities	3654
4	Engineering	2717
5	Agricultural and Biological Sciences	2089
6	Biochemistry, Genetics and Molecular Biology	2016
7	Computer Science	1511
8	Mathematics	1405
9	Environmental Science	1362
10	Business, Management and Accounting	1273
11	Psychology	1168
12	Materials Science	1150
13	Earth and Planetary Sciences	1136
14	Physics and Astronomy	1047
15	Economics, Econometrics and Finance	977
16	Chemistry	807
17	Pharmacology, Toxicology and Pharmaceutics	737
18	Nursing	611
19	Chemical Engineering	589
20	Neuroscience	557
21	Immunology and Microbiology	549
22	Health Professions	518
23	Energy	416
24	Decision Sciences	361
25	Veterinary	237
26	Dentistry	201
27	Multidisciplinary	116

Total: 24702 Journals / 31971 SCImago entries

Considering the discrepant range of distribution of journals among the subject areas, from 116 titles in multidisciplinary subject areas up to 7224 journal titles in Medicine subject area, it was defined a method to determine the quantity of journals that should be selected from each subject area aiming to get a representative sample of the total amount of journals.

This method considered the representativeness of the total amount of journal titles classified in each subject area in relation to the total number of journals classified by SCImago (24702 journal titles). These indices are represented by the percentage indicators at the top of each column in Graphic 1.

From this perspective, to ensure a sample proportionally similar to SCImago database, considering the number of journal titles indexed per discipline, the number of selected journal titles per discipline in our sample was based on the representativeness of each subject area in the SCImago database as a whole. So, the percentual indexes corresponding to the representativeness of each subject area in SCImago database was considered as the percentual of journals to be selected to our sample in the respective subject area. This procedure was applied for the 27 SCImago subject areas. For instance, from the total of 24702 of ranked journal titles at SCImago database at the occasion of data collection (100%), 7224 of those titles corresponded to the Medicine subject area (17.93% of the whole database). So, our sample of journal titles representing the Medicine subject area should be composed by 17.93% of the ranked journal titles under that subject area. That is to say that our sample should consider 17.93% of the 7224 Medicine journal titles, which corresponded to 27 journal titles. For the cases in which such method resulted in decimal numbers, i.e., when the number of journal titles to be selected corresponded to decimal numbers, we considered rounding them up to the next whole number, according to mathematical principles. The method did not admit less than 2 journals per subject areas thus, for the cases in which the aforementioned method resulted in any number lower than 1.99, we considered selecting 2 journal titles for the respective subject area.

According to the method described above, the number of journals to be selected per subject area was as described in Chart 5:

Chart 5 - Number of journal titles selected per subject area for sample composition

The 27 Subject Areas subdivision by SCImago			
SCImago Journal Subject Areas	Total number of journals at SCImago database[16]	Representativeness of the SCImago journals titles by subject area	Expected number of selected journals by subject area
Medicine	7224	17,93%	27
Social Sciences	5855	14,53%	22
Arts and Humanities	3654	9,07%	14
Engineering	2717	6,74%	10
Agricultural and Biological Sciences	2089	5,19%	8
Biochemistry, Genetics and Molecular Biology	2016	5,00%	8
Computer Science	1511	3,75%	6
Mathematics	1405	3,49%	5
Environmental Science	1362	3,38%	5
Business, Management and Accounting	1273	3,16%	5
Psychology	1168	2,90%	4
Materials Science	1150	2,85%	4
Earth and Planetary Sciences	1136	2,82%	4
Physics and Astronomy	1047	2,60%	4
Economics, Econometrics and Finance	977	2,43%	4
Chemistry	807	2,00%	3
Pharmacology, Toxicology and Pharmaceutics	737	1,83%	3
Nursing	611	1,52%	2
Chemical Engineering	589	1,46%	2
Neuroscience	557	1,38%	2
Immunology and Microbiology	549	1,36%	2
Health Professions	518	1,29%	2
Energy	416	1,03%	2
Decision Sciences	361	0,90%	2
Veterinary	237	0,59%	2
Dentistry	201	0,50%	2
Mutidisciplinary	116	0,29%	4
Total	**24702 Journal occurrences**	**100%**	**158 journal titles**

Once the journal titles sample was sized, we applied the following criteria to define the journal titles to compose our sample:

a) Only journals, and nor any other kind of periodical publications, like book series and proceedings were considered eligible for sample composition;
b) journals were selected according to 2018 SCImago average citation per document ranking in the three previous year's period. The selected journals

[16] The numbers showed in this table refers to those verified at the occasion of the data collection.

were the ones which have reached the highest citations scores, that is, the most cited ones in the three previous years (2015 up to 2017);

c) the sample did not admit journal titles from the same subject area within the same subject category. In those cases, the procedure adopted was replace the journal title with the smaller citation score by the next journal title in the citation ranking score that fits the selection requirements. This procedure was repeated until this and all previous requirements were met and until the predefined number of journal titles indicated in Chart 5 were reached in each thematic area;
d) the sample did not admit more than one journal from the same subject area within the same subject category. For those cases, the journal with the smaller citation score was replaced by the immediately next journal title in the citation ranking score that fitting the selection requirements;
e) publishers integrating editorial groups were considered as a unique publisher. For example, three different journals published respectively by Wiley, Blackwell and Wiley-Blackwell were not considered eligible to integrate the sample under the same knowledge area, considering that all these publishers belong to Wiley Publishing Group. The same criteria was applied to all similar cases, like Springer, Springer Verlag and Springer Netherlands, in which case all publishers were considered as Springer Publishing Group;
f) all selected journals were supposed to have one (only one and any one) quartile index attributed. Journals with no SCImago quartile index (Q1, Q2, Q3, Q4) were not considered eligible to sample composition;
g) due to the multidisciplinarity character of journals, SCImago may categorize journal titles in more than one subject area simultaneously. In these cases, the sample considered journals classified in only one subject area. In case of a journal title was evaluated and classified in more than one subject area (have more than one quartile index), the journal title was considered not eligible for the sample and then, was replaced by the following journal title in the citation ranking score that fits all the selection requirements;
h) the sample did not admit journals classified as “miscellaneous” in the subject category field at SCImago database. Exceptionally this criterion was not applied to the Multidisciplinary subject area, which followed an exclusive methodology for journal titles sample selection. Regarding to the Multidisciplinary subject

area, the number of journals expected to be selected, according to data showed in Chart 5, did not correspond to the representativeness of the total amount of the journal titles ranked by SCImago database. However, the first four positions of this subject area were taken by three of the most traditional journals of scientific and academic universe: Nature, Science and Proceedings of the National Academy of Sciences of the United States of America. Considering the multidisciplinary characteristics of such journal titles, all of them were considered eligible for representing Multidisciplinary subject area in our research sample which were represented by 3, instead of 1 journal title as expected, according to the previous method, based on the representativeness of the Multidisciplinary subject area in SCImago database.

In addition to the attendance of the previous requirements, the following three criteria were considered, in this order, to select the journal titles effectively composing the sample:

a) The first one was the values showed in the total cites ranking column at SCImago database. The sample considered the journals corresponding to the highest values, which corresponded to the most cited journals of each subject area, considering that the most cited journals might be the most important or the most popular in their respective subject areas.
b) The second criterion was the publisher. Our sample did not admit more than one journal from the same publisher under the same subject area. This assured the coverage of the different biases adopted by the editors regarding the presentation and formatting of citations and references in scientific articles.
c) The third and last criteria for selecting journal titles was the subject category in which the journal title was classified in the SCImago database. Each journal title is labeled in SCImago database by subject area and by subject category. Some journals may be classified simultaneously in more than one subject area and/or category. As we intended to evaluate the editorial habits in each particular discipline, the sample did not consider journal titles classified in more than one subject area and/or category.

So, the sample considered the journal titles best ranked in the total cites rates and, whose publisher and subject area or category were not the same of the previously selected journal titles under the same subject area. For the cases in which any of the selected journal title did not accomplish to all the journal titles selection criteria, it was replaced by the immediately following journal title in the “total cites” ranking, attending to the whole criteria previously introduced. Such process was carried in all subject categories and repeated until the number of journal titles determined for the respective subject area was reached.

The previously approached method suggested a sample of 158 journal titles. However, some disciplines did not have a number of journals greater than or equal to that suggested in the Chart 5. So, 149 journal titles accomplished to the journal sample selection requirements for composing our research sample.

7.2.The journal issues selection procedures

The sample took the date of 31 October 2019 as an upper bound reference from which the most recent journal issues fully attending the selection criteria were selected. Articles published between 1 October 2019 and 31 October 2019 (or the most recent equivalent period for journals not publishing articles in this interval) were considered for selecting the journals in case they did not organize their articles to any issue. Special issues and supplements were not considered eligible for composing the sample, since the editorial policies with which they are submitted may differ from those applied to regular issues. For the same reason, articles that were not original research communications, like special articles, letters to the editor, and book reviews, were not considered eligible for sample composition purposes.

7.3.The articles selection procedures

Once the journal issues that should compose our research sample were defined, we established a random method for selecting the articles of each journal issue for sample composition. However, we noticed that some journals did not have a regular periodicity, e.g., some electronic journals publish articles insofar as they are being approved and accepted for publication by the editorial board. On the other hand, some of them do not even mention a specific issue number for each journal issue.

Each journal is represented by 5 articles published in the most recent issue published between October 1st and October 31st, 2019. For journals not releasing any issue in this period, the sample considered the immediately previous issue published before October 1st. For issues containing more than 5 articles, the selection considered a probabilistic systematic random sampling technique, based on the average number of articles published by the journal in the aforementioned period. As for the journals containing less than 5 articles, the sample considered all those attending the selection criteria.

We considered that since all articles published by the same journal are subject to the same editorial policies, the variance of data collected from articles published by the same journal, should not add significant data to the analysis discussion, so that the coverage of the analysis should prioritize as many journals as possible per knowledge area, rather than increasing the number of articles analyzed per issue, in order to guarantee a more comprehensive coverage of how reference styles are applied to scientific articles.

Since not the standardization issues themselves (i.e., the conformity of the metadata presentation in relation to the guidelines of the reference styles they were supposed to be formatted), but the analysis of the typology of presentation of metadata and their relationships with the information they stand for were the main focus of the analysis, there were no justification for conducting the analysis of a larger volume of articles submitted to the same editorial policies, instead of prioritizing the coverage on the different ways of presenting mentions, quotations and their respective bibliographic references by the various editors active in the editorial market.

Therefore, we considered that a sampling size of five articles per issue would be an adequate sample which were not too large to demand too many efforts in the evaluation of redundant (and, therefore, irrelevant) data and approaches but, at the same time, not so small as to limit the analysis to a superficial view on the dynamics of bibliographic metadata within scientific articles.

The average number of articles published by all selected journals in the previously introduced chronological period delimited by the research (between October 1st and October 31st, 2019), was not admitted for dimensioning the articles sample. We observed that some journals publish hundreds of articles per month, and because of that, the monthly average of published articles would increase considerably. This might mask the real monthly average of articles by journal titles whose publishing rates are not as expressive. On the other hand, using the monthly average of published articles per journal title for supporting the determination of the number of articles to be selected per issue for sample composition, might have put some journal titles in a disadvantage situation from our sample perspective (specially the less representative ones in SCImago database), since some of them publish fewer articles than the monthly average observed in the period covered by the research.

In addition, we did not find any stablished standards to determine the order of appearance of articles within journal issues. Some journal titles structure the order of appearance of articles within their issues by subjects, others by date of acceptance. Others do not have a stablished criterion for that. So, by choosing the first 5 appearing articles of each selected issue, for example, could mean admitting the risk of analyzing 5 articles under the same subject, in some cases. Since this might also have represented redundancy on the evaluation of the editor's view, a randomly and exponential method of selection was adopted for choosing the articles for sample composition.

Facing the previously introduced issues, we developed a probabilistic systematic random sampling technique, considering a formula, showed below, where each "X" represents the average number of articles published by each selected journal title in the in the period covered by the research, as previously mentioned:

X; 2X, 3X, 4X, 5X

So, for each journal issue, the first selected article was the one occupying the "X^{th} position" in the issue summary. For example, considering that 30 is the average of articles published by a particular selected journal in the chronological period considered for the research, the 30^{th}, 60^{th}, 90^{th}, 120^{th}, and the 150^{th} articles of each issue should have been the selected ones.

For issues whose summary had fewer eligible articles than the result of the average number of articles published by the journal title in the in the period considered by the research multiplied by 5, a cyclical and systematic counting was conducted, starting with the first article of the issue, and continuing until the last, returning to the first, successively, until the average number of articles published in the month of coverage of the method was reached. When the first article was chosen, the counting should restart from the article immediately following to the chosen one, which did not integrate the counting afterwards. This process was repeated 5 times for each journal issue, until the 5 articles per issue were selected.

Book reviews, letters to the editor and other formats than original communications (articles), were not considered as eligible for this method. Since the terminology applied by different publishers to name its respective journal sessions is not uniform and also considering that not every journal designates original communications as so, the sample also considered the following terms as synonyms of original communications: articles, papers, researches, original papers, original articles, original researches, research papers, research articles, regular papers, regular articles, regular researches and regular articles. Journal issues which did not publish any original communication in the period covered by this method, were discarded from the sample and replaced by the immediately previous eligible issue under the same journal title. It is important to point out that a journal issue publishing at least one up to 5 original communications within the selected journal issue remained considered eligible for sample composition, even in those cases in which we observed less than 5 published works attending the articles selection criteria.

Some journals divide their summaries in subject categories and then, articles which thematic is multidisciplinary might be simultaneously considered in more than one subject category within the journal issue summary. That is to say that it is true that one article could appear twice in the same summary, in different subject categories. For those cases, the sample considered the “duplicated article” only once for doing the previously introduced counting for selecting articles per issue.

Since the research is supported basically by the analysis of bibliographic references metadata, the sample considered articles whose bibliographic references were presented in an exclusive list, designated for the sole purpose of indicating the bibliographic references of the cited works within the article’s text body, regardless of how this session was named by the publisher. Exceptionally, we admitted articles whose bibliographic references were distributed in footnotes throughout the text, if they were presented in a format corresponding to one of the widely accepted reference styles such as Vancouver, Chicago or APA, or similar. Journals that did not meet these requirements, even if they met all the others, were not admitted in the sample.

For the analysis, we prioritized accessing the .pdf version of the articles. Exceptionally, in cases that this format is not available, the HTML format, printed (hardcopy) or any other available versions were accepted. For those cases in which was not possible to have access to any of the selected articles of a selected issue, the journal was replaced by the immediately following eligible journal according to the methodological parameters.

For those cases in which it was not possible to get access to the articles of a selected issue in any way, the journal title was replaced by the previously immediately eligible one, according to the aforementioned methodological parameters.

7.4.The journal analysis procedures

The journal analysis considered, initially, the following aspects:

a) Modality of access: we verified whether the modality of access to the journal was free access or restricted access. Within the scope of this method, free access, which is considered a synonym of "freely available", is understood by the unrestricted online access to a determined content, free of costs or other access requisites. At first, there are three accessing modalities for journals: the freely available, when the contents published within a particular journal can be accessed with no need of any counterpart, as a payment for example, the restrict access, which means the need to pay for accessing a particular content and, the mixed access, which applies to journals which makes available both freely access content and restrict access content. For proceeding the modality of access evaluation, we used the version 79 of Google Chrome browser, and an Italian IP with an individual Internet connection (i.e., not institutional).
b) Reference style adopted: the identification of the reference style adopted by each selected journal which allowed to point the most used reference styles in each subject area and how publishers are susceptible to develop their own reference styles.

c) Reference styles adaptations: for those publishers adopting a widely adopted reference style (e.g., Vancouver, APA, Chicago among others), we analyzed whether they recommended any adaptation on the guidelines of the adopted reference style. That specific part of the analysis evinced the level of explicit influence of publishers on the way bibliographic references, mentions and quotations are presented in their respective journals.
d) Maximum number of bibliographic references: Some discussions demand a more in-depth scientific basis than others. In these cases, it is common that articles are plenty of mentions and quotations. However, considering the purposes and the general length of an article, it is not recommended to have too extended bibliographic references lists. First, because having plenty of quotations and mentions in a text might makes its reading tiring and boring, in addition to obfuscating the author's own arguments. Second, because the discussion presented in an article should be not too short as not to include the exposition of purposes, methods, results, and the author's arguments, nor too long in order to be out of character and confused with other types of publications. Since there is no convention regarding to the range of the length of bibliographic references lists (and in fact, there are no means for it, without impairing the specificity of each work that may require longer and in-depth discussions), authors are expected to use common sense when writing their scientific productions. On the other hand, some publishers use templates for diagraming and in these cases, there might be some specific area for bibliographic references lists, which no possibilities of length adjustments and so, authors have to write their texts in order to fit in that specific space delimited by the publisher. This specific part of the analysis aimed to evince such limitations imposed to authors by publishers which may partially limit author’s needs of discussing or arguing on a particular issue.
e) The way instructions for publishing are transmitted to authors: this topic is not about the instructions for publishing properly. This is regarding the way instructions for presenting mentions, quotations, and bibliographic references (the reference style itself) are presented to the authors. Some editors indicate a specific reference style to be used in their publications, including those developed by themselves. Others give few and superficial examples of

reference styles guidelines and that is all. There is not any clear and explicit instruction concerning the way of presenting mentions and quotations: whether in text bodies, between single or double quotation marks, when indents should be applied, etc. Still, some publishers accept submissions in any bibliographic format and make available a team which is responsible for the formatting of the accepted works (not always free of charge for the author). This specific part of the analysis identified the cases in which the publishers did not indicate a specific bibliographic format in which the author would be able to find all the instructions on how to proceed for presenting mentions, quotations, and bibliographic references, should be considered as unclear instructions.

f) Management reference style softwares: There are plenty of management reference style softwares available to the scientific community, some free of charge, some not. Some publishers recommend the use of these tools, some do not. To identify the intensity of the editorial appeal for the use of these instruments, we considered whether the publisher recommended their use or not.

g) Bibliographic reference list name: the way bibliographic references lists are named within journals might also be variable. To check the veracity of this premise and the various ways in which the session of bibliographic references was called in journals, we carried a verification on this matter.

h) Bibliographic references assortment: there are commonly two ways of bibliographic references assortment: the citation-sequence, which is most used by the medical journals and related areas, and the author-data assortment, which is more commonly used by the humanities researchers. We carried an analysis on which of the citation systems are being adopted by publishers and how in-text reference pointers referring to each of them are being presented in text bodies.

7.5.The article analysis procedures

Articles were individually analyzed. The data collected from such observations comprised the most substantial part of data supporting the discussions. What follows are the aspects considered in the analysis and some related considerations, which were compiled in the raw data of the research available from Santos (2021), where all the following approached aspects are showed from a quantitative perspective:

a) Modality of access to the article: articles were classified as freely available online or restrict access. For proceeding the modality of access evaluation, we used the version 79 of Google Chrome browser, and an Italian IP with a particular Internet connection (i.e., not Institutional connection), from which we actually tested the access of all selected articles of the sample. At this point of the analysis, it is important to clarify that we verified both the modality of access for the journal and for the article. For instance, specifically considering journals of the Medicine subject area, most articles are indexed by the US National Center for Biotechnological Information (NCBI)[17]. However, even with restricted access on the publisher's website, the full texts from those journals can be found freely available from the NCBI database. In such cases we considered the editor’s choice regarding to the way of access to the article. That is, the parameter to define whether the article was freely available or not, was always the publisher’s option, according to data observed from their websites, no matter if the article was freely available in other sources than the publisher’s own. Sample issues were not considered free access as well
b) Format of the article file: some publishers make their articles available in PDF. files, others in HTML, and others in both formats. We detected the range of availability of each format among the selected journals.
c) Information for citation: some publishers make available a tool for automatically write the bibliographic reference of the articles published in their journals. These tools may either export files data to the read by the bibliographic reference manage softwares, or give the bibliographic reference for the article, ready for use, according to the reference style chosen by the user from the options given by the publisher in his webpage. We proceeded an observation concerning the modality of the availability of such tool within publisher’s webpages and such data supported a discussion based on quantitative data.
d) Availability for bibliographic reference metadata: Some publishers include a small header or a footnote in their articles, which may appear in the first page

[17] Available from https://pubmed.ncbi.nlm.nih.gov/

or in all pages of an article, in which can be found basic metadata for writing the bibliographic reference of the article itself. These metadata generally comprise (at least) the journal title, the year of publication, the number of the volume and issue and initial and final article page numbers. This feature was quantified into a yes or no evaluation regarding to its presence within the articles.

e) The way of presenting journal titles within bibliographic references: According to the reference style adopted, the titles of the cited articles within the bibliographic references may be given in full or abridged. The point is that the abridged version of the journal titles may led the reader to an ambiguous comprehension of the cited work's title and, consequently, constitute a barrier to the access to the cited work. Considering this, we quantified and identify the subject areas adopting each of the forms of presenting journal titles within bibliographic references.
f) The total number of authors: this is a basic data, collected to possibly support further discussion on the relationship between the number of authors of an article and the accuracy of information given in mentions, quotations and bibliographic references.
g) Non-textual cited content's sources provision: Graphics, photos, figures and other types of non-textual content also may be included in a text body as a mention or as a quotation. The premise here was that not always the source from where those excerpts were extracted or adapted are properly indicated. The expected procedure is to always inform the source from where the external data included in the text body were taken, regardless their format. However, this habit was suggested not to be a common practice among authors and so, this evaluation indicated the portion of the articles properly indicating the source of non-textual cited content and the portion which do not.
h) In text-reference pointers format: we also verified the correspondence between in-text reference pointers referring to mentions and quotations and their respective bibliographic references, i.e., the bibliographic references referring to the works mentioned or quoted in the text body, referred by the in-text reference pointers. So, we considered the way in-text reference pointers were presented within the article text bodies and this particular observation allow to

identify: first, whether they were clear and understandable; second, whether they evinced the link with the bibliographic reference corresponding to the mentioned or quoted excerpt regarded to the in-text reference pointer; third: whether the metadata set provided in the in-text reference pointers better fit to the concept of FRBR Work, FRBR Expression, FRBR Manifestation or FRBR Item. Whenever any inconsistence between in-text reference pointers and respective bibliographic references entries were detected, we counted an unconformity, since such events might configure a barrier on the identification of the cited work within the bibliographic references list. That is to say that in-text reference pointers referring to mentions and quotations included in the text bodies, not matching any bibliographic reference entries in the bibliographic reference lists, were considered as unconformities, even in those cases in which such linking was implicit but had to be inferred, as well as mentions or quotations whose work containing the cited content was not represented within the bibliographic reference list.

i) Links between in-text reference pointers and bibliographic references: we analyzed whether the in-text reference pointers referring to mentions and quotations were hyperlinked to the bibliographic reference containing the description of the works they referred in the text body. Besides being a convenient tool for the reader, such links explicit the correspondences between mentions, quotations, and their respective bibliographic references. We also proceed an evaluation on the cases in which in-text reference pointers were hyperlinked with the respective bibliographic references in the bibliographic references list and considered whether the linkage between them were round, which meant that by clicking on the in-text reference pointer the reader was expected to be submitted to the correspondent bibliographic reference in the bibliographic reference list and, by clicking on the bibliographic reference the reader should be sent back to the excerpt of the text where the work, which was represented by the bibliographic reference, was mentioned or quoted.

j) Total and average number of mentions and quotations per article: we counted the total mentions and quotations per article and such data supported a discussion on how the subject areas tend to cite external publications and, which was the most common way of doing it: in a literal form (using quotations) or in an interpretative form (using mentions). Quotations were considered as so, regardless the length of the quoted excerpt, be it a phrase, a paragraph, or an expression, as long as it was properly identified as such, according to the reference style adopted, which clearly expressed the intention of the author in quoting a particular content. Although non-textual contents, e.g., figures, also may be mentioned and quoted within text bodies, they were not considered by this analysis.

k) In-text reference pointers referring to mentions providing pagination data: a mention is characterized by the reproduction of an excerpt of a cited work under the words and the interpretation of a citing author. By using this way of citing, the indication of the pagination data concerning the precisely point of an external publication, i.e., the cited work, where the mentioned content can be found, is not mandatory, according to most of the reference styles and bibliographic standards. Therefore, we assume that the provision of such pagination data in the in-text reference pointers referring to mentions is an author's choice. However, it was also assumed that the indication of the page numbers in one in the reference pointer referring to a mention requires the indication of such metadata in all the remaining in-text reference pointers referring to other mentions within the same text body, to keep the uniformity. Facing this, we quantified the percentual rates of in-text reference pointers referring to mentions providing and not providing pagination metadata.

l) In-text reference pointers referring to quotations not providing pagination data: quotations are characterized by being a transcription of an excerpt of an external (cited) work. And as being a literal reproduction of an external content, the quote should be evidenced in the text. Most (if not all) of the reference styles recommend that quoted excerpts should be somehow detached from the self-authored content by using quotation marks, followed by an in-text reference pointer. The provision of page numbers within in-text reference pointers is one of the factors usually differing in-text reference pointers referring to mentions from those referring to quotations. Starting from those statements, the in-text reference pointers referring to quotations were analyzed within the selected journals in order to quantitatively identify: a) whether the initial and final page numbers where the quoted excerpt may be found in the cited work were given within the in-text reference pointers referring to quotations; b) whether it was possible to identify the exact starting and finishing point of the quoted excerpt within the citing work, that is, whether the quotation was presented between quotation marks, and which type of them. In case of external contents were not presented between quotation marks, even whether the in-text reference pointers gave the page numbers, we considered them as mentions, since there was no way to assure that those particular excerpts were effectively mentions or quotations.

m) The way quotations are presented in the article texts: The form of presentation of quotations was also a point of observation, starting from the premise that reference styles do not provide clear and enough instructions concerning the way of including quotations within the articles' text bodies. Considering that we did not found a standard representing an agreement on the length that defines long quotations within scientific publications, which are generally presented with an indent, and short quotations, which are generally presented between quotation marks, we developed a brief discussion on the uniformity on the way quotations are presented within journals, regarding the use of quotation marks and indentation formatting.

n) FRBR relation between in-text reference pointers referring to quotations and bibliographic references: In a comparison with mentions, quotations tend to be more evident and precise within a text body. Since the metadata referring to the pages numbers where a quoted excerpt may be found in the cited work are given within the in-text reference pointer referring to a quotation, a FRBRized feature can attributed to such in-text reference pointer. That means that by giving the page numbers, the in-text reference pointer is referring to an FRBR Manifestation, while the omission of this metadata in this context usually links the in-text reference pointer to a FRBR Expression. Nevertheless, the quotation itself refer to a FRBR Work, regardless the embodiment its Expressions may take, and, from this perspective, it does not matter whether the citing author consulted the electronic or the printed version of a work (considering that their contents are the same). Starting from this approach, this point of the analysis analyzed the relationship established between quotations, their respective in-text reference pointers and their respective bibliographic references, from a

FRBRized point of view. The initial premises indicated that bibliographic references generally refer to FRBR Manifestations, while in-text reference pointers, not necessarily. From this perspective, we quantified the percentual of bibliographic references and in-text reference pointers referring to quotations, according to the FRBR entity they corresponded: FRBR Work, FRBR Expression, FRBR Manifestation and FRBR Item.

o) Total number of bibliographic references: Starting from the premise that the total number of bibliographic references included in the bibliographic references list should be equal to the number of works cited within the text we proceed an evaluation on the total and average number of bibliographic references per article and per subject area and, on their assortment within the bibliographic references lists. This allowed us to identify characteristics regarding to citing needs and habits within the subject areas.

p) Mentions and quotations included in the text body not providing a respective bibliographic reference: all mentions, and quotations should have a correspondent bibliographic reference in the list of bibliographic references, no exceptions admitted. The reverse situation, likewise, is also valid: there should not be a bibliographic reference in the bibliographic reference list not corresponding to a mention or quotation within the article text body. We carried out a quantitative verification on this issue.

q) Type of publications cited: The number of bibliographic references referring to the types of publications (e.g., articles, books, proceedings, etc.) cited within the articles composing our sample were verified. All bibliographic references composing our sample were individually analyzed and the quantitative results supported the verification on the most relevant types of publications (i.e., the most cited ones) on each subject area. Books published simultaneously in press format and in electronic support, were counted as printed copies, unless there was an explicit note or metadata in the bibliographic reference suggesting that the author was referring to the electronic version of the publication. The bibliographic references not providing enough elements for the identification of the type of publication to which they referred, by themselves, were classified as "undefined".

r) DOIs and hypertext links provision within bibliographic references: Since a DOI Hyperlink is a unique identifier for an electronic publication available online, it is convenient to indicate it in the bibliographic references whenever it is possible. The same concept applies to publications which are available online, under a hypertext link, although they are subject to change, as opposed to DOI hyperlinks and DOI numbers. However, there is a premise that DOIs and hypertext links are rarely included in bibliographic references, even considering that they develop such a crucial role in information identification. To verify the validity of this premise, a quantitative analysis on the indication of hypertext hyperlinks, DOI hyperlinks and DOI numbers within the bibliographic references were carried out. We quantified the following aspects within bibliographic references composing our sample: a) the number of freely available online bibliographic references not including hypertext links or DOIs, b) the total number of bibliographic references including hypertext links and, c) the total number of bibliographic references including DOI hyperlinks or DOI numbers. We noticed that some publishers presented the bibliographic references within their articles by using clickable hyperlinked bibliographic references, which might point to an embedded hypertext or a DOI hyperlink. In these cases, the DOI could not be clearly identified at first sight, especially considering the printed version of the articles, since it is embedded in the hyperlinked bibliographic references. Because of this, all hyperlinked bibliographic references were considered as a hypertext hyperlink provider, even in cases in which such hyperlink pointed to a DOI hyperlink.

s) Link rot provision within bibliographic references: Electronic resources available on the Internet are subject to be reallocate or made unavailable permanently. Consequently, the links originally pointing to these resources, lost their functionality and do not permit the access to the file, web page, server or other source to where these links were originally pointed at. This phenomenon, called link rot, is frequent among bibliographic references and this analysis intended to quantify the incidence of such link rots within the bibliographic references lists of the articles composing the sample. So, for each hypertext hyperlink or DOI hyperlink included in the bibliographic references list we carried out an analysis to: a) quantify the bibliographic references including a hypertext link rot, b) quantify the bibliographic references including DOI hyperlink rots and, c) quantify the number of hypertext links or DOI hyperlinks not pointing to the cited work (e.g., some hyperlinks point to bibliographic catalogs and not to the cited work itself).

t) Format of bibliographic references numbers at the bibliographic references list: for the articles adopting the citation-sequence system, the format of the number at the beginning of each bibliographic reference were observed. These data provided an overview of the range of formats adopted by publishers to express similar information within the bibliographic references included in the articles composing the sample.

u) Bibliographic references metadata compilation: Basically, for each set of quotation or mention and their respective bibliographic references, we considered whether the bibliographic reference permitted the reader to precisely identify the content being mentioned or quoted considering, included, the need for indicating the consulted pages in the cited source. Based on such analysis, we established a common metadata set considered across disciplines to properly represent citing and cited works in papers through bibliographic references considering the identification of represented information in a clear and unmistakable way as its main function. Each bibliographic reference was individualized under the aspect of the metadata set composing them. We identified the type of metadata composing bibliographic references (i.e., work author, work title, year of publication, etc.). Such analysis resulted in the dataset available in found Santos (2021) and evinces the most considered metadata for bibliographic references composition, per type of cited work and per discipline. This list supported the delimitation of the most considered metadata across disciplines for describing cited works through bibliographic references, which we called “starred metadata set”.

The analysis addressed mentions (i.e. rephrasing a passage or idea introduced in a cited work without quoting it explicitly), quotations (i.e. a reference to an explicit textual passage of a cited work reported in the citing work), their respective in-text reference pointers (i.e. the textual devices, such as “[3]”, that denote the bibliographic references related to mentions and quotations), and bibliographic references both from qualitative and quantitative aspects to demonstrate the dynamic relationship between them, the interdependence in the fulfillment of their main purposes (and their relation with bibliographic catalogs), and the importance on assuring "clear, precisely, stated, and commonly shared understanding" metadata within all formats of representing information comprised by descriptive representation (IFLA Study Group…, 2009, p. 2).

The in-text reference pointers referring to quotations and respective bibliographic references were analyzed from the standpoint of the Functional Requirements for Bibliographic Records (FRBR), designed by the International Federation of Library Associations, which is an entity-relationship based conceptual model for describing bibliographic records for all types of materials (IFLA, 2009). This analysis considered the correspondence between the concepts of FRBR Expression and FRBR Manifestation entities and the elements provided by in-text reference pointers and respective bibliographic references.

In the last part of the analysis, we considered the total of 34140 bibliographic references[18] composing the bibliographic references lists of the 729 selected articles which were analyzed both in quantitative and qualitative aspects. For the quantitative perspective, we detected the most cited types of works in each discipline and also, the overall most cited ones. As for the qualitative analysis we detected the structure of bibliographic references for each type of cited work, considering different reference styles' formatting guidelines, i.e., the identification of the descriptive elements adopted for the bibliographic references for each type of cited work. Such descriptive elements were classified and assorted according to the Resource Description & Access (RDA) core elements (session 1.4), which provided the identification of the overall rate of use of each descriptive element within the bibliographic references composing the sample by subject categories.

Based on such analysis, we defined the "standard metadata set" considered in bibliographic references across disciplines in our sample, which we called "starred metadata set". In the ambit of this study, the "starred metadata" should be understood as the set of the most used descriptive elements observed within bibliographic references representing each type of cited work detected in our sample.

[18] Which in fact, (should) represent cited works

Finally, such “starred metadata set” was analyzed from The Semantic Publishing and Referencing Ontologies (a.k.a. SPAR Ontologies) perspective, particularly the FRBR-aligned Bibliographic Ontology (a.k.a. FaBiO), aiming to verify its applicability on the description of published or publishable entities within bibliographic references, focusing on those identified as cited works in our sample. The discussions were carried out in the subject category level, to make data more digestible and the discussions more fluid. Further studies considering a micro-view form the issues approached in this research are encouraged.

The previous analysis, particularly those supporting the “starred metadata set” definition, supported a discussion on SPAR Ontologies. We proceeded an analysis on such metadata set and FaBiO Ontology, which integrates the set of ontologies composing the SPAR Ontologies, to verify its suitability on describing bibliographic metadata considered in bibliographic references, particularly considering the “starred metadata set”. These considerations supported the verification whether FaBiO Ontology is prepared in structure, terminology (vocabulary) and range to attend particularly the “starred metadata set” for writing bibliographic references. The analysis also identified suggestions of improvement for FaBiO Ontology to better attend bibliographic needs and expectations concerning the proper description bibliographic metadata.

7.6. An overview on the sample

The sample is composed of 729 articles (we retrieved in PDF format) published in 147 journals from 27 thematic areas, arranged under four main top thematic categories (i.e., subject categories), as shown in Chart 6:

Chart 6 - Journals sample quantitative overview ("Avg" means average)

Subject categories	Subject Areas	Total number of journals	Total number of articles	Freely available journals	Freely available articles	Avg number of authors per article	Total biblio-graphic references	Total number of mentions	Avg mentions per article	Total number of quotations	Avg quotations per article
Health Sciences	**Medicine**	27	132	33%	42%	9	5340	8400	64	5	0.04
	Nursing	2	10	0%	50%	2	440	768	77	17	1.70
	Veterinary	2	10	0%	50%	5	336	561	56	0	0.00
	Dentistry	2	10	0%	10%	6	381	533	53	0	0.00
	Health Professions	2	10	50%	50%	4	271	446	45	1	0.10
	Health Sciences overall rates	**Total 35**	**Total 172**	**Avg 16.6%**	**Avg 40.4%**	**Avg 5.2**	**Total 6768**	**Total 10708**	**Avg 59**	**Total 23**	**Avg 0.37**
Social Sciences	**Arts and Humanities**	10	45	20%	40%	3	1752	2556	57	614	13.64
	Business, Management and Accounting	5	25	40%	40%	2	1642	2719	109	73	2.92
	Decision Sciences	2	10	0%	50%	2	428	749	75	0	0.00
	Economics, Econometrics and Finance	2	10	0%	75%	2	457	610	61	1	0.10
	Psychology	4	20	0%	5%	4	1536	2959	148	32	1.60
	Social Sciences	17	81	35%	52%	2	4571	6953	86	835	10.31
	Social Sciences overall rates	**Total 40**	**Total 191**	**Avg 15.8%**	**Avg 43.6%**	**Avg 2.5**	**Total 10386**	**Total 16546**	**Avg 89.33**	**Total 1555**	**Avg 4.76**
Life Sciences	**Agricultural and Biological Sciences**	8	40	25%	40%	5	2281	3445	86	1	0.03
	Biochemistry, Genetics and Molecular Biology	8	39	25%	41%	6	2229	3524	90	0	0.00
	Immunology and Microbiology	2	10	0%	50%	6	521	786	79	0	0.00
	Neuroscience	2	10	50%	80%	6	606	848	85	0	0.00
	Pharmacology, Toxicology and Pharmaceutics	3	15	33.3%	53,3%	9	821	1213	81	0	0.00
	Life Sciences overall rates	**Total 23**	**Total 114**	**Avg 26.6%**	**Avg 52.8%**	**Avg 6.4**	**Total 6458**	**Total 9816**	**Avg 84.2**	**Total 1**	**Avg 0.01**

Continues

Continuation

Subject categories	Subject Areas	Total number of journals	Total number of articles	Freely available journals	Freely available articles	Avg number of authors per article	Total biblio-graphic references	Total number of mentions	Avg mentions per article	Total number of quotations	Avg quotations per article
Physical Sciences	**Chemical Engineering**	2	10	50%	50%	5	452	706	71	0	0.00
	Chemistry	3	15	33.3%	40%	6	682	846	56	0	0.00
	Computer Science	8	39	0%	20%	3	1125	2167	56	0	0.00
	Earth and Planetary Sciences	4	20	25%	50%	7	1157	1942	97	0	0.00
	Energy	2	10	0%	10%	3	583	924	92	0	0.00
	Engineering	10	48	30%	35.4%	4	1470	2148	45	13	0.27
	Environmental Science	5	25	20%	32%	4	1547	2147	86	43	1.72
	Materials Science	4	20	0%	5%	5	845	1156	58	1	0.05
	Mathematics	5	25	20%	60%	3	569	900	36	0	0.00
	Physics and Astronomy	4	20	75%	50%	4	1101	1758	88	2	0.10
	Physical Sciences overall rates	**Total 47**	**Total 232**	**Avg 25.3%**	**Avg 35.24%**	**Avg 4.4**	**Total 9531**	**Total 14694**	**Avg 68.5**	**Total 59**	**Avg 0.21**
Multidisciplinary	**Multidisciplinary**	4	20	25%	35%	11	997	1697	85	1	0.05
	Multidisciplinary overall rates	**Total 4**	**Total 20**	**Avg 25%**	**Avg 35%**	**Avg 11**	**Total 997**	**Total 1697**	**Avg 85**	**Total 1**	**Avg 0.05**
Total 5 subject categories	**Total sample overall rates**	**Total 149**	**Total 729**	**Avg 21.8%**	**Avg 41.4%**	**Avg 4.74**	**Total 34140**	**Total 53461**	**Avg 75**	**Total 1639**	**Avg 2**

Regarding the number of authors per article, on average, co-authorship was more frequent among Multidisciplinary articles compared with articles of the other subject categories. Data collected within the scope of this investigation did not demonstrate reasons justifying such behavior and, therefore, this subject may configure an object of further studies.

The sample size reproduces the representativeness of each thematic area according to SCImago Database. All the results considered in this study were based on the data extracted from articles composing the sample and/or from those available from publishers' webpages, as they have been provided in the occasion of the data collecting, i.e., from November 2019 to May 2020.

Regarding publishers' nationality, 18 countries are represented within the journals sample. On average, the main 3 publisher's nationalities per subject category are as follows: Health Sciences: United Kingdom (46%), United States (38%) and Japan (10%); Social Sciences: United States (40%), United Kingdom and Netherlands (both with 26%); Life Sciences: United States (32%), Switzerland (24%) and United Kingdom (20%); Physical Sciences: United States (33%), United Kingdom (23%) and Netherlands (20%) and lastly, Multidisciplinary with the United Kingdom and United States (both with 50%). This allows us to consider United Kingdom, United States, and the Netherlands as the most contributing countries in the journals sample, with a respective average representativity of 36.01%, 27.74%, and 16.62% of the whole sample.

100% of analyzed journals provide their articles in downloadable PDF files. From this, 72.8% of journals also provide articles in HTML format. (i.e., Nursing, Veterinary, Dentistry, Business, Management and Accounting, Decision Sciences, Agricultural and Biological Sciences, Immunology and Microbiology, neuroscience, Pharmacology, Toxicology and Pharmaceutics, Chemistry, Earth and Planetary Sciences, Energy, Environmental Science, Material Science, and Multidisciplinary).

Regarding the modality of access, 40.4% of the articles composing the sample is free access, against 59.6% of restricted access articles. All gathered data are presented in the following sections by means of appropriate plots. The raw data on which all findings are based can be freely accessed in Santos, (2021).

8. THE RESULTS

The results are divided into three sessions: The first one, referring to citing and referencing systems in Social Sciences and Medicine articles from different theoretical and practical perspectives, considering bibliographic references as a facet of descriptive representation. These particular disciplines were considered separately from the remaining 25 subject areas composing the study (as shown in Graphic 1) first, because they were the most representative disciplines in our sample. Second, both disciplines showed particular (and sometimes opposite) characteristics and habits which were approached by Santos, Peroni, Mucheroni (2021).

The second part expands the results showed in the first part, by including the whole 27 disciplines approached in Graphic 1 in a deeper and broader view of the citing and referencing habits across the subject areas.

The third part considers the findings showed in Part II as the starting point to a data crossing between the descriptive elements (metadata) usually composing bibliographic references observed in the bibliographic references of the articles analyzed, with the SPAR Ontologies, particularly considering FaBiO Ontology. The main purpose of this specific observation was to verify the applicability of FaBiO Ontology to the description of the bibliographic elements most used within bibliographic references, per type of publications being cited.

The raw data for the results presented next, can be found at Santos (2021).

8.1. Part I – Citing and referencing habits among Medicine and Social Sciences journals

This session contains the discussions concerning the analysis considering particularly the Social Sciences and Medicine subject areas, whose results were published as a scientific article in the Journal of Documentation (SANTOS, PERONI, MUCHERONI, 2021).

8.1.1.Part I – Citing and referencing habits among Medicine and Social Sciences journals – Data

Considering the data available on SCImago in November 2019, we selected 46 journals and 213 articles from both the subject areas in consideration, i.e. Medicine and Social Sciences. The articles contained a total amount of 9,911 bibliographic references and 16,193 mentions and quotations overall. We obtained 27 journals, 132 articles, 5,340 bibliographic references, 8,400 mentions and 5 quotations in Medicine, and 19 journals, 81 articles, 4,571 bibliographic references, 6,953 mentions, and 835 quotations in Social Sciences.

The sample size proportionally corresponds to the representativeness of both disciplines, i.e. Medicine and Social Sciences, within SCImago database. The method adopted for the selection of the journals allowed us to come up with a balanced and representative population of journals which supported us in doing reasonable initial refletions on citing and referencing issues, that may be extended in forthcoming studies.

The sample comprises journals owned by publishers from 13 different countries. Publishers of Medical journals are from Canada, China, Egypt, Germany, Hong Kong, Netherlands, New Zealand, Switzerland (each country with 1 journal representing 3.7% of the Medicine sample), United Kingdom (8 journals representing 29.6% of the Medicine sample) and United States (11 journals representing 40.7% of the Medicine sample). Publishers of Social Sciences journals are from Brazil, Germany, Portugal, South Korea (each country with 1 journal representing 5.9% of the Social Sciences sample), United Kingdom (7 journals representing 41.1% of the Social Sciences sample), and United States (6 journals representing 35.2% of the Social Sciences sample).

Gathered data included information to support a discussion from multiple points of view, i.e., the publisher, journal, and article perspectives. Also, we collected more granular data about in-text reference pointers and bibliographic references, which provided a view on the citation apparatus. These different viewpoints supported an overview of the variations with which bibliographic data appears and how they relate to each other in articles considering the subject areas we analyzed.

8.1.2.Part I – Citing and referencing habits among Medicine and Social Sciences journals – Findings

33% of the Medicine journals and 35% of Social Sciences journals composing the sample are Gold/Diamond Open Access journals[19]. From an article perspective, 42% of the Medicine articles and 52% of Social Sciences articles can be freely accessed from the journal website without paying any fee – because either the journal is a Gold/Diamond Open Access one or it is a Hybrid journal.

Analyzing the works cited by the articles included in our sample, we observed that 55% of the works cited by the Medicine articles are freely available online (either as Green, Gold or Diamond Open Access items) and 50.1% of the bibliographic references referring to them (1462) do not provide a DOI URL (e.g. an URL starting with either “http(s)://doi.org/” or “http(s)://dx.doi.org/” followed by a DOI) to access them directly from the Web. Along the same line, only 35% of the works cited by Social Sciences articles are freely available online and 61% of the bibliographic references referring to them do not provide a DOI URL.

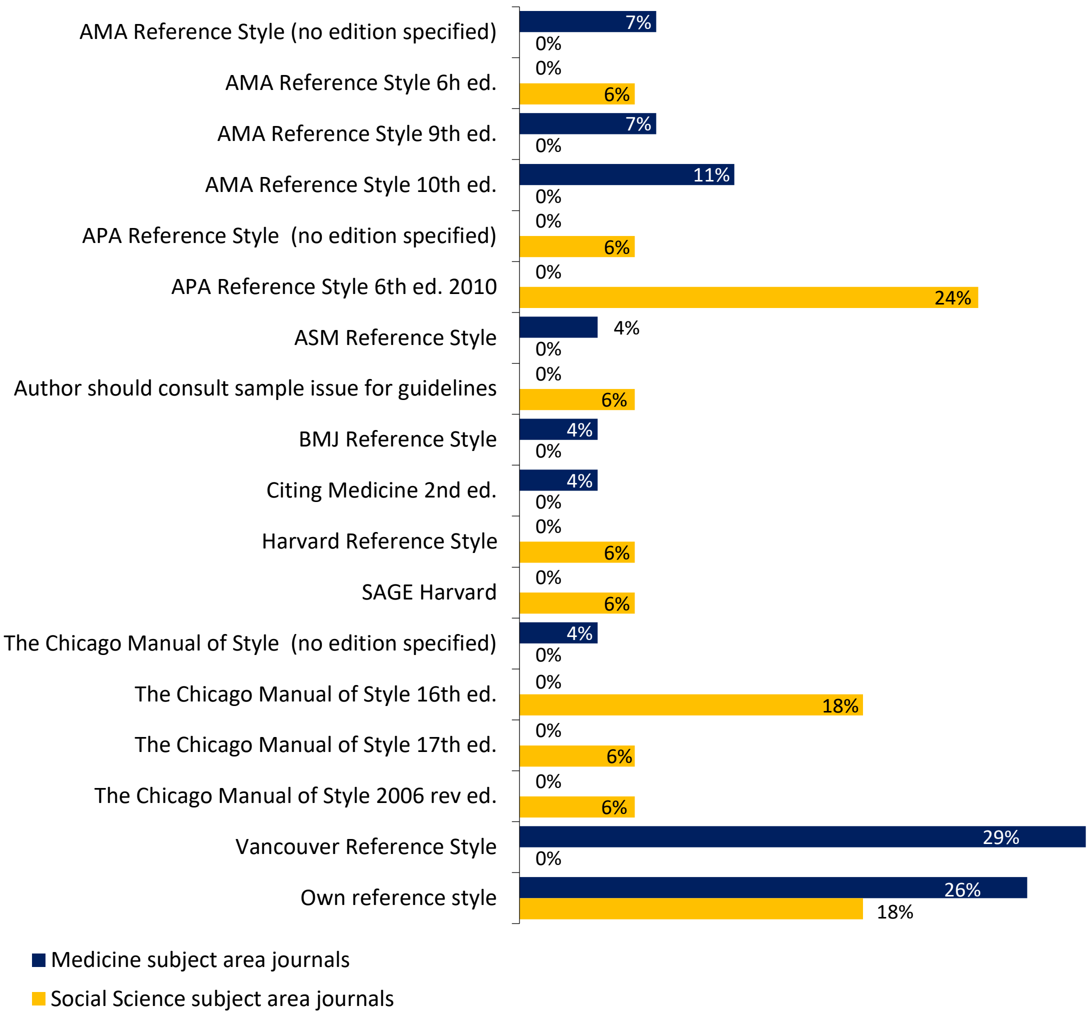
Considering all the bibliographic references included in Medicine articles providing a URL that did not include any DOI (even when hidden behind a hyperlink), 20% of them specified a URL, while only 12% of bibliographic references in Social Sciences articles contained an URL. However, 42% of the URLs indicated in bibliographic references of Medicine articles often referred to records within a bibliographic database (e.g., a library’s bibliographic catalog, like Pubmed[20]). We observed similar behavior in 2% of the bibliographic references in Social Sciences articles.

The data we gathered included the reference styles adopted by the journals, according to the recommendations in the instructions provided for authors. We noticed a huge variety of reference styles among the journals in the sample, as shown in Graphic 2.

[19] See (Piwowar *et al.*, 2018) for a definition of all the Open Access levels.

[20] PubMed is a free resource supporting the search and retrieval of biomedical and life sciences literature with the aim of improving health–both globally and personally (available in https://pubmed.ncbi.nlm.nih.gov/about/).

Graphic 2. Percentage of the adoption of reference styles in the journals in our sample



70% of Medicine journals adopted widely accepted reference styles like Vancouver, Chicago, AMA, and APA. Among these, 26.3% customized the adopted reference style guidelines according to their specific needs. Such customization practice was not observed in any Social Sciences journal in our sample, which usually directly reused the original reference style chosen as it is.

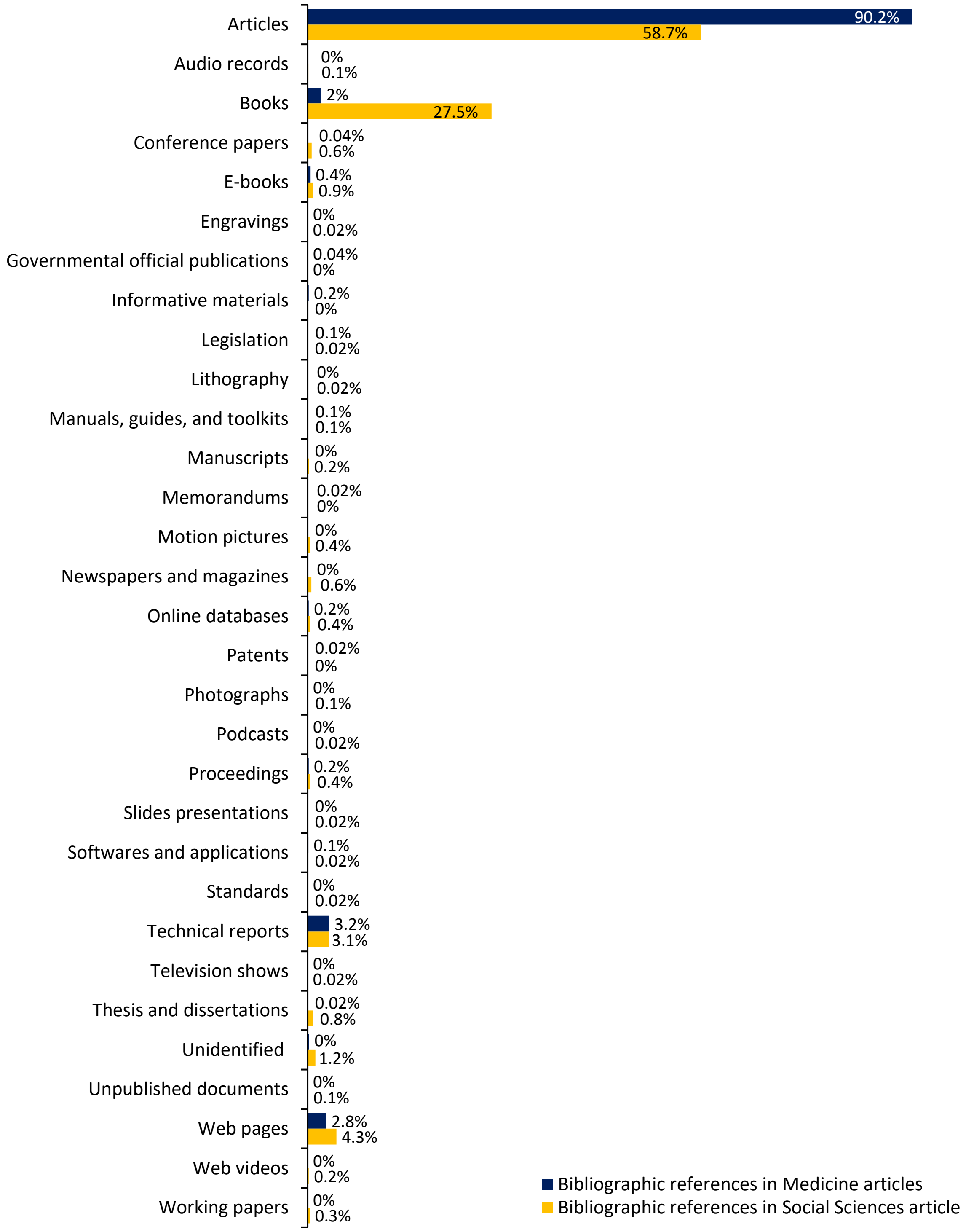
Considering only journals providing own reference styles (Medicine: 26%; Social Sciences: 17.6%), the reference styles of 33% of Medicine journals and 23.5% of Social Sciences journals did not provide clear, comprehensive, and exhaustive guidelines for accurately describing and arranging bibliographic metadata in bibliographic references. The reference styles can be classified into three non-disjoint categories:

1. those which did not provide guidelines for describing some types of publications like grey literature, e-prints, technical reports, speeches, etc.;
2. those which did not provide instructions on how to properly establish correspondences between bibliographic references and the in-text reference pointers used in mentions and quotations and;
3. those which did not provide instructions on how to proceed with particular bibliographic issues, like citing secondary or indirect sources (e.g., quoting quotations), structuring and formatting DOI metadata, etc.

In Social Sciences articles, we noticed variations in the title attributed to the section containing bibliographic references. 87.6% of articles name it *References*, 6.2% name it *Referências* (References, in Portuguese, following the article’s language), and 6.2% name it *Notes*. 100% of Medicine articles name the bibliographic references section as *References*.

The analysis of the type of publications cited by the articles in our sample showed 33 different types of works, as shown in Graphic 3.

Graphic 3. Percentage of the types of the cited works derived from analyzing the related bibliographic references (BRs) of all the articles in our sample

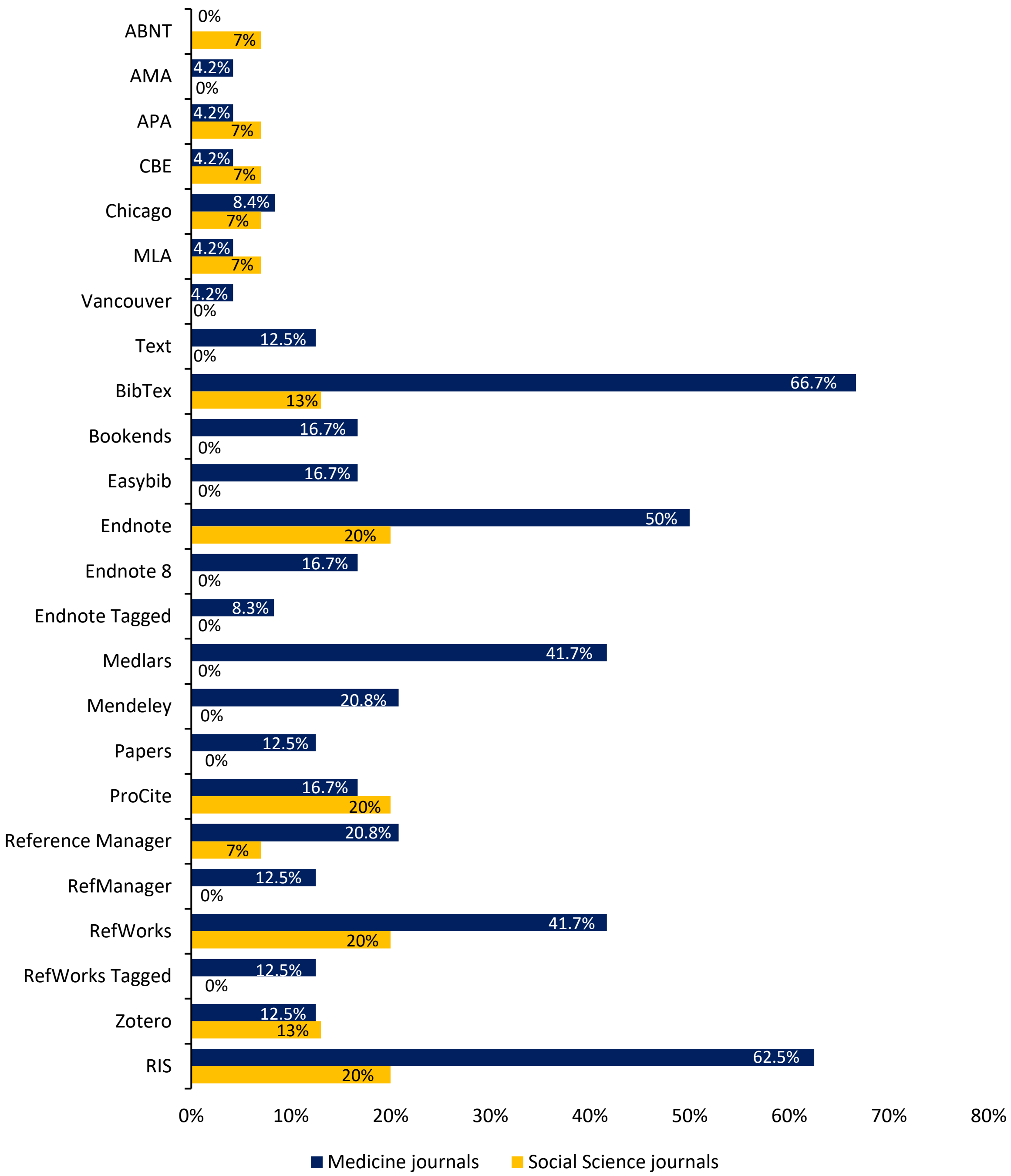


41% of Medicine journals and 29.4% of Social Science journals recommended the use of one or more reference management softwares. From this, 90% of journals in Medicine recommended the use of Endnote (https://endnote.com/), 27% recommended Mendeley (https://www.mendeley.com/), and 18% recommended

Zotero (https://www.zotero.org/). Considering Social Sciences journals, 100% suggested Endnote, and 20% suggested Zotero.

Some journals (Medicine: 89%; Social Sciences: 88.2%) enabled one to download bibliographic metadata of their articles in textual or machine-readable formats, as shown in Graphic 4. From this, a portion of publishers (represented by 12,5% of Medicine articles) provided bibliographic metadata for their articles in text format, with no information regarding the reference style in which it is formatted.

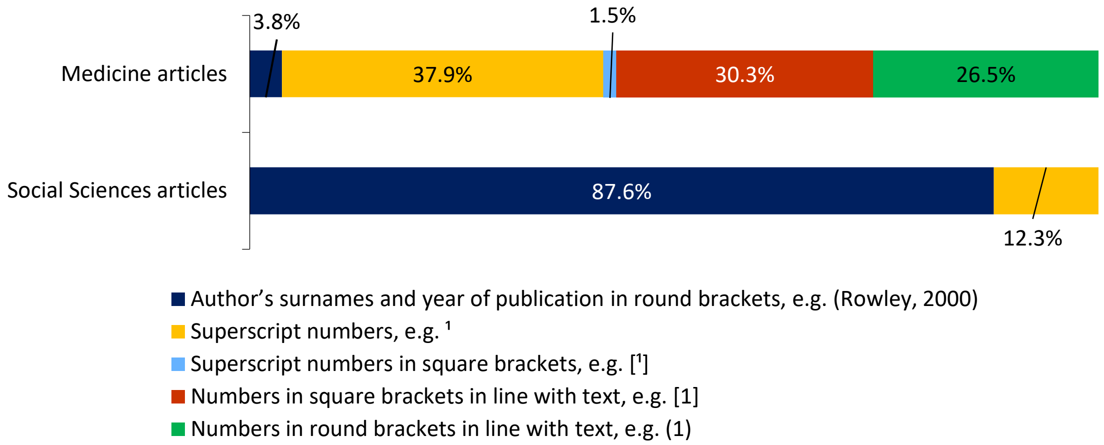
Graphic 4. Percentage of the export formats of bibliographic records of the articles published in the journals in our sample. The "Text" category refers to journals providing bibliographic references in plain text with no specification of the reference style in which it is formatted).



96% of Medicine articles and 97.5% of Social Science articles provided their own bibliographic metadata in headers, footnotes, or the first page of PDF files they provide.

Usually, articles adopt one out of two systems for identifying in-text reference pointers. The first one is the *author-date* system – e.g. "(Doe, 2020)" – adopted by 3.7% of Medicine articles and 88.2% of Social Sciences articles. The second, the *citation-sequence system*, based on numbers (e.g. "[3]"), was adopted by 96% Medicine articles and 5.9% of Social Sciences articles. The remaining 5.9% of Social Science articles adopted both styles within the same articles. All these data are introduced in Graphic 5.

Graphic 5. The format used for representing in-text reference pointers referring to mentions and quotations within articles



Considering articles adopting the citation-sequence system, we observed variations on the format of the numbers associated with each bibliographic reference, as shown in Graphic 6. Within this system, the numerical arrangement of bibliographic references did not correspond to the order in which the respective in-text reference pointers appear in the text in 8.66% of Medicine articles. For instance, let us consider the first paragraph of a Medicine article, shown in Figure 12. The first in-text reference pointer (e.g., "[41]") refers to the bibliographic reference in position 41 and the second one (e.g., "[40]") to the bibliographic reference in position 40. Similarly, the three in-text reference pointers that follow do not follow an ascendant numerical sequence: [1, 21, 31], [13, 27], [31, 34]. The journal in which this article was published adopts Vancouver reference style, according to which "references should be numbered consecutively in the order in which they are first mentioned in the text" (ICMJE, c2020), something that is not happening in this case. No additional instructions regarding the numerical arrangement, neither for in-text reference pointers nor for the bibliographic references, were found within the reference style.

Graphic 6. Different uses of the numbering systems for ordering bibliographic references in the bibliographic references list, considering articles adopting citation-sequence system.

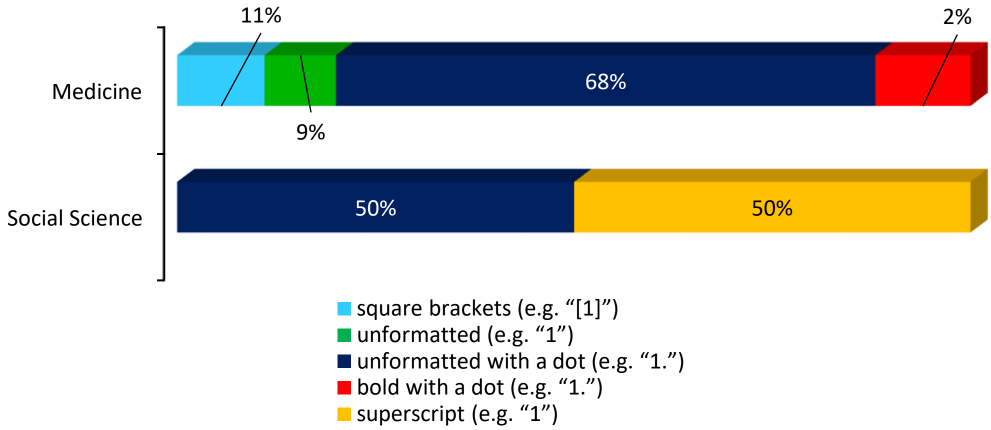


Figure 12 -An excerpt from a paragraph in a Medicine article in our sample[21]

The World Health Organization (WHO) defines mental health as "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community." [41] One in four people worldwide are affected by mental health concerns [40].

We observed misuses regarding the alphabetical arrangement of bibliographic references in 40% of Medicine articles adopting such system and 3% of Social Sciences articles adopting the author-date citation system. Figure 13 includes an excerpt of the bibliographic reference list of one of the articles in our sample, in which its bibliographic references were expected to be sorted in ascending order by authors' surnames.

[21]Passages referred in Figures 3 to 8 were extracted from articles in our sample.

Figure 13 -An excerpt of the bibliographic reference list of an article in our sample

Maparyan, L. (2012). The womanist idea. New York, NY: Routledge.

Morris, E. W. (2007). ""Ladies" or "loudies"? Perceptions and experiences of Black girls in classrooms." Youth & Society, 55(4), 490-515.

Ladson-Billings, G. & Tate, W.F. (Eds.). (2016) Covenant keeper: Derrick Bell's enduring education legacy. New York: Peter Lang.

Johnson, L. (2017). The Racial Hauntings of One Black Male Professor and the Disturbance of the Self(ves): Self-Actualization and Racial Storytelling as Pedagogical Practices. Journal of Literacy Research, 49(4), 476-502.

Lorde, A. (1984). Sister Outsider: Essays and Speeches. Trumansburg, NY: Crossing Press.

Matthew, P.A. (2016). Written/Unwritten: Diversity and the hidden truths of tenure. Chapel Hill, NC: The University of North Carolina Press.

Some publishers also provided hyperlinks to connect in-text reference pointers to the corresponding bibliographic references. This behavior was observed in 60% of Medicine articles and 25% of Social Science articles. However, the reciprocal hyperlink, i.e., between a bibliographic reference to the in-text reference pointers denoting it, was not a usual feature provided in the Medicine articles. Instead, considering only the articles with hyperlinked in-text reference pointers, 100% of Social Sciences articles provide backlinks to the in-text reference pointers, while only 10% of Medicine articles implement such functionality.

Regarding the uniformity of descriptive metadata, journal titles of cited articles may appear in distinct forms within bibliographic references. These differences may exist either when comparing bibliographic references lists of articles from different issues of the same journal or even between bibliographic references lists included in the same journal issue. Differences regarded to the abbreviation of journal titles provided in bibliographic references were observed only in Medical journals since 100% of Social Sciences articles gave full titles of cited journals in their bibliographic references. Considering the 93% of Medicine articles providing titles of cited journals in the abridged format in their bibliographic references, 4% adopts the ISO 4 rules[22], 8%

[22] Refers to a Standard published by the International Standardization Organization (ISO), entitled "ISO 4:1997 Information and documentation — Rules for the abbreviation of title words and titles of publications" (https://www.iso.org/obp/ui/#iso:std:iso:4:ed-3:v1:en).

adopts the ISSN List of Title Word Abbreviations (LTWA) guidelines[23], 1.6% adopts the recommendations of the National Center for Biotechnology Information Database, 44.3% adopts the recommendations of US National Library of Medicine (NLM, also referred by publishers as Pubmed Database and Index Medicus)[24]. For the remaining 42% of Medicine journals, the source in which abbreviations should be based on was not identified.

Our data revealed that articles did not usually provide sources of cited non-textual content – i.e., all that information presented by using visual signs, tables, graphs, photographs and images, illustrations, schemes, verbal communications, audio, and video recordings or any other type of manifestation made without using argumentative text as the main language. Non-textual content was observed in 92.4% of Medicine articles and 71.6% of Social Sciences articles. Of these, 84.8% of Medicine articles did not provide the source of non-textual cited content and 4% of articles provided the source for only part of non-textual cited content. In the Social Sciences articles in our sample, 31% of them provided the source of non-textual cited content and 12% of articles provided the source for part of it.

Quoting and mentioning proved to be more frequent in Social Sciences articles than in Medicine articles. In Medicine, we observed an average of 64 mentions and 0.04 quotations per article, while in Social Sciences we had 86 mentions and 10.31 quotations per article. 71.6% of Social Science articles included at least one quotation, which seemed to confirm the impression that Social Sciences usually quote more compared with Medicine articles, in which only 2.27% of articles used quotations.

About the markup used to show and identify quotations within a text, 100% of quotations in Medicine articles used double quotation marks (i.e., “”) to markup the text in run-in quotations. No long quotations were detected in Medicine articles in our sample. In Social Sciences articles, quotations were marked up in different ways. For run-in quotations, 70.6% of the articles adopt double quotation marks (i.e., “”), 24.1% of the articles adopt single quotation marks (i.e., ‘’), and 1.7% of the articles adopt both double and single quotation marks. Considering long quotations, 31% of Social

---

[23] Refers to a list containing standardized abbreviations used for words in scientific citations. (Available at https://www.issn.org/services/online-services/access-to-the-ltwa/?lang=en).

[24] Available at https://www.ncbi.nlm.nih.gov/nlmcatalog/journals.

Science articles used simple indentation, while 3.4% of the articles presented quotations in indented passages and italicized characters between double quotation marks.

Usually, providing the page numbers in the in-text reference pointer associated with a mention is optional. However, it was used in 41.9% of the articles in Social Science, while Medicine did not show such behavior. Instead, the page numbers in in-text reference pointers referring to quotations were found in 68.9% of Medicine articles and 68.9% of Social Science articles.

From a FRBR perspective, in-text reference pointers with page numbers were used to refer to a particular FRBR Manifestation of a cited document (i.e., a particular edition). Instead, the remaining in-text reference pointers, i.e., those without the specification of page numbers, pointed to the FRBR Expression level of the cited documents (i.e., their content), despite the particular format specified in the metadata of the related bibliographic references, which usually described the cited document at the FRBR Manifestation level.

The article analysis revealed that considering articles with quotations, 33.3% of Medicine articles and 48.2% of Social Sciences articles specified either in-text reference pointers or bibliographic references (or both) related to quotations that did not provide easy access to the text quoted. With *easy access,* we mean to locate a quoted passage within the cited document without having to perform complementary searches, like queries on indexes and summaries, or to read long excerpts to identify the quotation in its original source. One example of these cases is illustrated in Figure 14, which reproduces a passage from a Social Science article containing a quotation of a Collin’s work, published in 2000. We noticed that the in-text reference pointer matches two different bibliographic references in the bibliographic references list, both published in 2000. Since the in-text reference pointer do not provide any specification of which bibliographic references it referred to, e.g., by adding an alphabetical character to the date of publication, like 2000a, and assuming that the provided pagination is correct, the only way of retrieving the quoted passage on the cited work is by consulting both works referenced by the two bibliographic references.

Figure 14 -An excerpt from a Social Sciences article

In the text-body:

It also develops space for Black women researchers to do work that Collins (2000) describes as activating "epistemologies that criticize prevailing knowledge and that enable us to define our own realities on our own terms..." (p. 292).

In the bibliographic references list:

Collins, P.H. (2000). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. New York, NY: Routledge.

Collins, P. H. (2000) "What's going on? Black feminist thought and the politics of postmodernism." In E. A. St. Pierre & W.S. Pillow (Eds.) Working the ruins: Feminist post-structural theory and methods in education (41-73). New York: Routledge.

The excerpt in Figure 15, extracted from an article from the Social Science subject area, illustrates some of the issues mentioned above. In the original source, the year of publication of each cited work within the in-text pointer is connected to the respective bibliographic reference in the bibliographic reference list through a hypertextual link. By clicking on the link provided in the in-text reference pointer for "Okoh and Hilson, 2011", we were sent to the bibliographic reference shown in Figure 16.

Figure 15 -An excerpt of a Social Sciences article

"This is the case for many countries such as Ghana, Tanzania, Senegal, and Mozambique (Fisher, Mwaipopo, Mutagwaba, Nyange, & Yaron, 2009; Aizawa, 2016; Bryceson & Geenen, 2016; Hilson & Garforth, 2012; Okoh and Hilson, 2011; Persaud, Telmer, Costa, & Moore, 2017)."

Figure 16 -The bibliographic reference of one of the in-text reference pointers introduced in Figure 15

Hilson, G. (2011). Artisanal mining, smallholder farming and livelihood diversification in rural Sub-Saharan Africa: An introduction. Journal of International Development, 23(8), 1031–1041.

None of the works authored by Hilson described in the bibliographic reference list of such article was co-authored with Okoh. The only work dated from 2011 considered in the bibliographic references list was the one represented in Figure 16. By checking the work represented in the bibliographic reference in the webpage of its publisher, we confirmed that Okoh was not an author of the work referred by means of its own bibliographic reference. Thus, there is no evidence that the in-text reference pointer and bibliographic reference referred to the same work, neither that the quoted passage is actually a quotation, or a mention incorrectly marked up as a quotation.

Figure 17 illustrates a situation in which the author does not use the same approach to specify in-text reference pointers within the text. Indeed, since the in-text reference pointer referring to the bibliographic reference number 41 appears after the period ending a quotation and no page number was provided in it, readers have to infer autonomously to which passage in-text reference pointers 40 and 41 refer to since it is ambiguous whether 41 refers to the previous sentence or not. These are only a few examples demonstrating how ambiguous bibliographic metadata was within the articles in our sample.

Figure 17 -Another excerpt from a Social Science article.

…community". [41] One in four people worldwide are affected by mental health concerns [40].

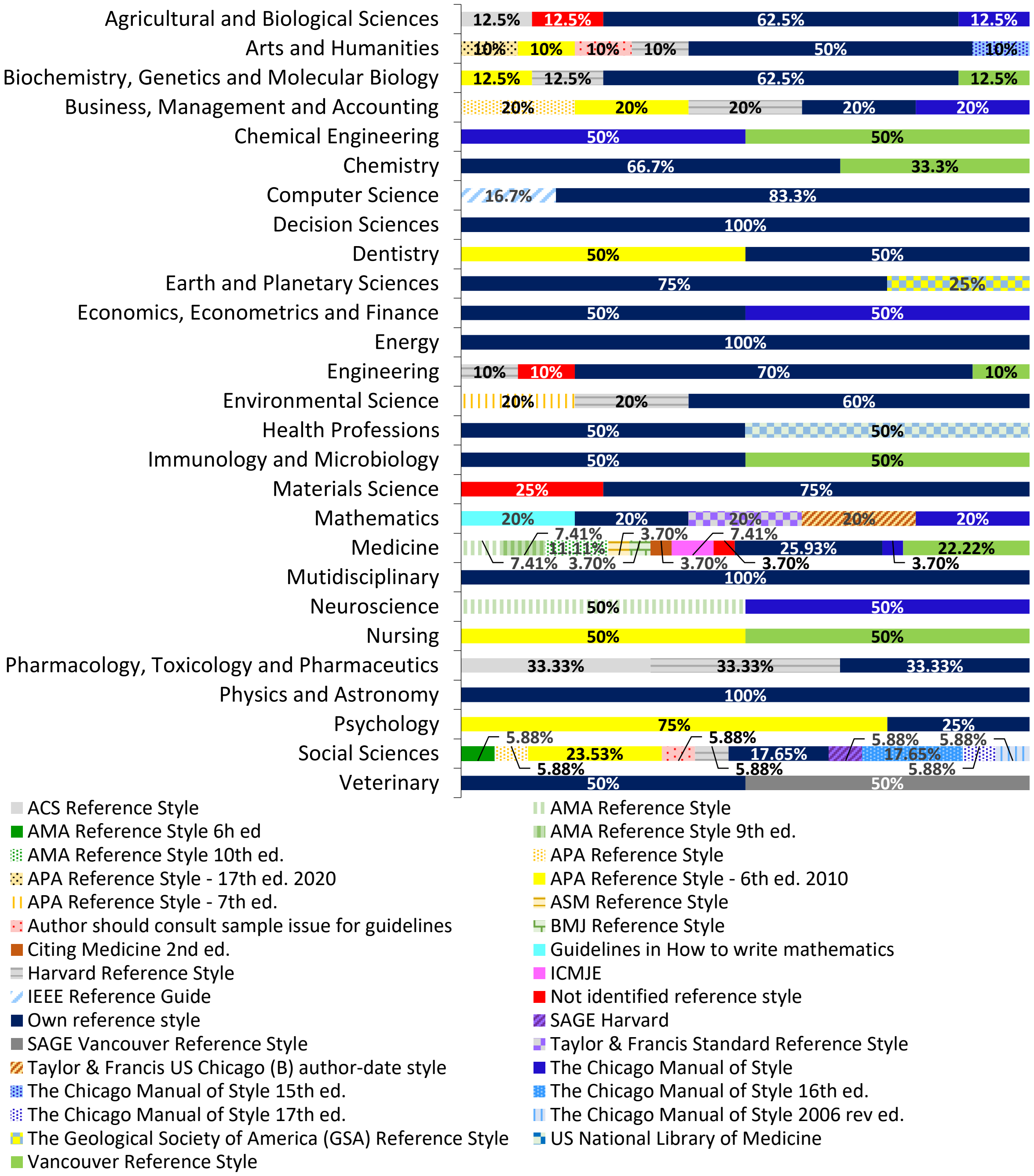
8.2.Part II – Citing and referencing habits across all scholarly disciplines

Based on the total of bibliographic references gathered which, theoretically, reflects the volume of works being cited within a citing work, Life Sciences showed the highest citing habits rates with an average of 56 bibliographic references per article, closely followed by Social Sciences subject category with an average of 54 bibliographic references per article.

100% of articles of the sample contain mentions, while only 19.2% contain quotations. The average perspective over subject categories showed the highest indexes concerning mentioning and quoting habits in Social Science. Nevertheless, the dispersion among data addressed in quoting and citing rates denotes a heterogenicity in the degrees of intensity of citation habits both within and across the disciplines. For instance, considering articles from Psychology, the percentual increment rates between articles with the lowest and the highest indexes of mentions per article is 1.048% (i.e., articles in Psychology tended to contain a similar number of mentions), as for Dentistry this variation is of the order of 100%.

Heterogenicity was a characteristic also observed among the reference styles adopted by journals, even considering disciplines under the same subject category, as shown in Graphic 7.

Graphic 7. Percentual representation of reference styles adoption, by subject area



According to Graphic 7 and Chart 7 (below), the four most adopted reference styles are APA 6th ed. 2010 (8.05% of total journals sample), Vancouver (10.73% of total journals sample), Chicago (4.69% of total journals sample), and those referred as “own reference styles" (46.97% of total articles sample), corresponding to customized versions of widely accepted reference styles, i.e., Vancouver, Chicago, and APA or, reference styles authored by the publishers themselves.

Data pointed to such reference styles as the most adopted ones across all disciplines, with particular reference to the Multidisciplinary subject category. While considering the widely accepted reference styles, it can be said that Health Sciences and Physical Sciences appear technically tied as the disciplines that are most adept to Vancouver Reference Style with 14% and 14.89% of journals, respectively.

Disregarded the reference styles authored by publishers, Vancouver proved to be the most adopted one across disciplines (10.73% of journals sample). In Second place, The APA Reference Style (6th ed. 2010), which achieved the rate of 8.05% of adoption across disciplines also showed to be the most adopted within Social Sciences' journals (52.94%). As for Life Sciences, the most adopted reference style was Chicago (no edition specified), with the rate of 8.69% of adopting journals in such subject category and 4.69% of adopting journals across disciplines. Chart 7 considers data showed in Graphic 7 in a less granular way, since it focuses on the reference styles showing the highest rates of adoption per subject category. Subject categories in bold represent the best adopters of each reference style.

Chart 7 - Percentage of journals per subject category adopting the four most adopted reference styles

Reference Styles	Reference styles' adoption rate within the sample	Subject categories	Percentage of journals per subject category adopting the reference style within the sample
Vancouver Reference Style	10.73%	**Health Sciences**	**14%**
		Social Sciences	0%
		Life Sciences	8.69%
		Physical Sciences	**14.89%**
		Multidisciplinary	0%
APA 6th ed. 2010	8.05%	Health Sciences	5.71%
		Social Sciences	**52.94%**
		Life Sciences	4.34%
		Physical Sciences	0%
		Multidisciplinary	0%
Chicago Reference Style (no edition specified)	4.69%	Health Sciences	2.85%
		Social Sciences	5%
		Life Sciences	**8.69%**
		Physical Sciences	4.25%
		Multidisciplinary	0%
Publisher's own reference style	46.97%	Health Sciences	28.57%
		Social Sciences	35%
		Life Sciences	52.17%
		Physical Sciences	59.57%
		Multidisciplinary	**100%**

In total, we found 31 different guidelines on formatting citing and referencing data and metadata provided or indicated by publishers in their journal's webpages. 15 different reference styles were detected within journals composing the Social Sciences subject category, disregarding the multiple editions of a single reference style. Indeed, Social Sciences was the subject category with the widest range of reference styles, followed by Health Sciences (14 reference styles), Physical Sciences (11 reference styles), Life Sciences (8 reference styles), and lastly, the Multidisciplinary subject category in which only one reference style was detected (i.e., "own reference style").

On average, our sample showed the adoption of around 3 reference styles per subject area. Medicine and Social Sciences rates overcame the average with respectively 11 and 10 different adopted reference styles. Arts and Humanities; Business, Management and Accounting and, Mathematics adopted 5 different reference styles each – all of them comprising journals adopting "own reference styles".

The most appropriate (i.e., the less inappropriate) behavior concerning this matter was observed in 48.14% of the subject categories adopting two different reference styles, namely: Chemical Engineering; Chemistry; Computer Sciences; Dentistry; Earth and Planetary Sciences; Economics, Econometrics and Finance; Health Professions; Immunology and Microbiology; Materials Science; Neuroscience; Nursing; Psychology, and Veterinary. But, in the meantime, it should be noted that except for three disciplines – i.e., Chemical Engineering, Neuroscience and Nursing – part of journals of the remaining disciplines adopts a widely accepted reference style – e.g., APA, Chicago and Vancouver reference styles – while the other part of the journals adopts “own reference styles”.

Graphic 8 shows an analysis of journals adopting “own reference styles”, as introduced in Graphic 7 and Chart 7. This analysis considered whether the reference styles provided enough instructions for metadata description considering mainly the following aspects:

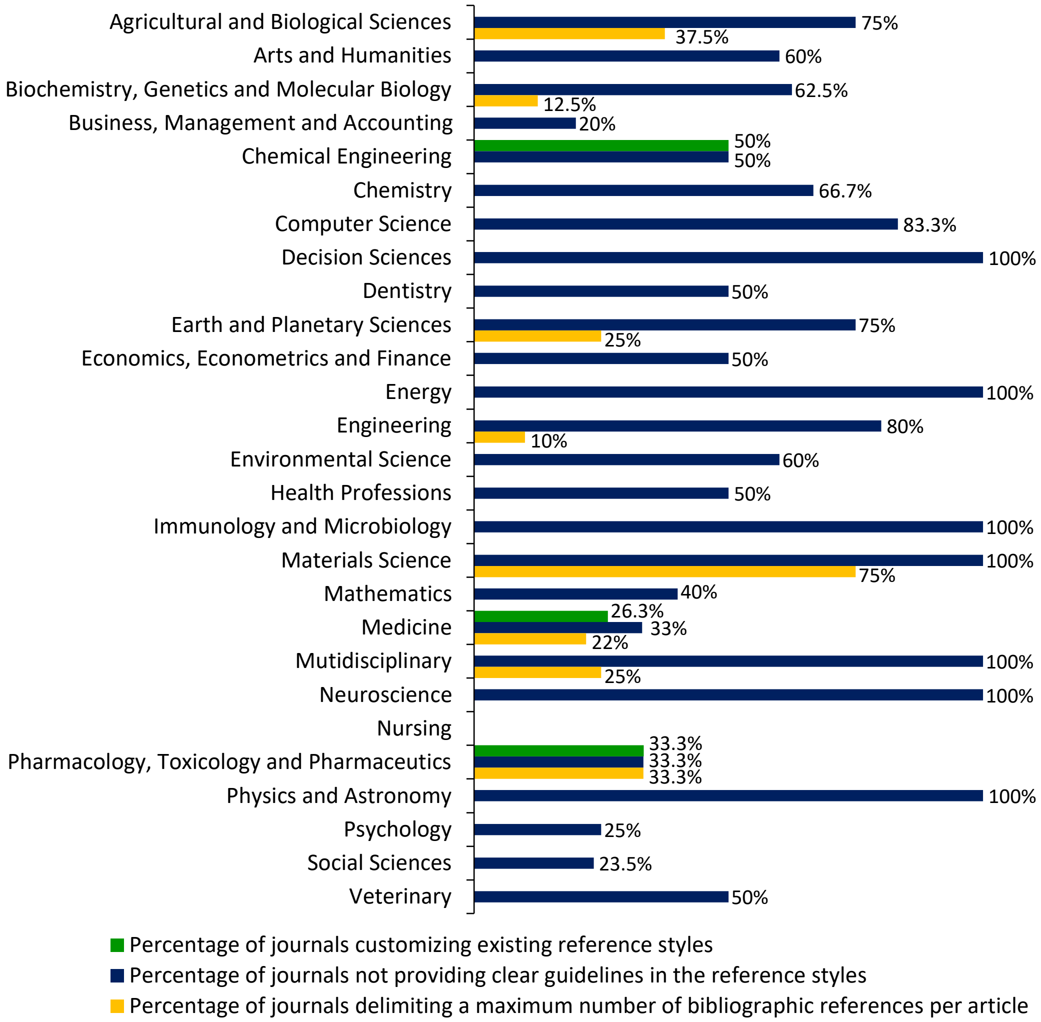
a) The guidelines coverage: we evaluated whether the reference styles provide clear guidelines[25] for (at least) the most frequently cited types of publications, i.e., books, articles, proceedings, and correlated papers and events, electronic content available online, and grey literature (as further shown in Graphic 10). The results of such evaluation showed that reference styles usually do not address instructions for describing all the types of publications cited by the analyzed articles. For instance, we noticed that reference styles do not usually provide clear instructions regarding the differences between the description of a work presented in a conference and the description of the same work in the version published in the conference proceedings. Some reference styles do not

[25] By “clear guidelines” we mean an extensive and comprehensive set of guidelines for describing and formatting citing and referencing metadata, i.e., in-text reference pointers concerning mentions and quotations, and bibliographic references. Journals clearly stating the adopted reference styles were considered as “clear guidelines providers”, based on the understanding that by clearly stating the adopted reference styles, the responsibility of providing clear instructions on this matter automatically falls upon the reference style itself, instead of the publisher. Since the purposes of this work do not include the analysis of the widely adopted reference styles’ contents, i.e. Vancouver and APA, the verification of the level of clarity of those guidelines remains an open question to be approached in further studies.

provide instructions on how to describe some cited types of publications, like thesis, dissertations, software, legislation and others;

b) The guidelines specificity: we considered the level of detail of the guidelines in consideration. We noticed that such instructions usually do not comprise various common aspects of the bibliographic universe, i.e., the description of DOI or hypertext hyperlinks metadata or shared authorships and the proper bibliographic references assortment to be adopted in the bibliographic references lists.

Graphic 8. Percentual distribution of journals considering aspects from the respective reference styles



Considering the gathered data, we noticed that the average number of journals adopting reference styles authored by publishers per subject category is close to that of “non-providers of clear citing and referencing guidelines”, as shown in Chart 8. This scenario, if confirmed in further studies, may suggest that the adoption of non-standard guidelines may result is higher possibility of not having clear guidelines.

Chart 8 - The relation between the average rates of journals per subject category providing reference styles authored by publishers and the average rate of journals per subject category adopting reference styles not providing clear citing and referencing guidelines

Subject category	The average rate of journals providing reference styles authored by publishers	The average rate of journals adopting reference styles not providing clear guidelines
Health Science	35.19%	35.93%
Life Sciences	41.67%	44.17%
Social Science	43.78%	46.42%
Physical Sciences	65%	70.50%
Multidisciplinary	100%	100%

62.49% of reference styles were classified by the analysis as “not providers of clear guidelines”. In 70% of the subject areas, the average rates of journals adopting reference styles classified as “not clear” are equal or higher than 50%. In 18.51% of the sample (i.e., Decision Sciences, Energy, Materials Science, Physics and Astronomy and Multidisciplinary), the rate of journals adopting “own reference styles” classified as “not clear” is 100%. From the subject area perspective, some disciplines showed a critical scenario, in which 100% of the reference styles were addressed as “not clear”, namely Decision Sciences (corresponding to 16.66% of Social Sciences subject category); Immunology and Microbiology and Neuroscience (corresponding to 40% of Life Sciences subject category); Energy, Material Science, Physics and Astronomy (corresponding to 30% of Physical Sciences subject category) and Multidisciplinary subject area (corresponding to 100% of the Multidisciplinary subject category). The best behavior concerning the clarity of the reference style adopted by journals was observed within the Nursing subject area in which no journal was classified as a provider of unclear guidelines.

On average, 54.96% of journals do not provide clear guidelines concerning citing and referencing, and nevertheless, 6.12% of publishers, on average, limit the maximum number of bibliographic references allowed per article. These limits usually corresponded to 30 bibliographic references with particular cases considering 35 and others considering up to 50 bibliographic references in specific situations. The overall indexes place Multidisciplinary subject category as the one in which publishers more exert influence on author's writings, by limiting the total number of bibliographic references per article (25%), followed by Life Sciences (17%) and, Health Sciences and Physical Sciences (4% each). However, we detect articles whose bibliographic references list extent is longer than the limit established by the publishers of the journals they are published in. Social Sciences showed as the more flexible subject category concerning the extent of bibliographic references lists, with no publisher establishing maximum limits.

The average of bibliographic references per article, considering the whole sample, is 47.75. From the subject categories perspective, the rates are as follows: 56.32 for Life Sciences, 54.39 for Social Sciences, 49.85 for Multidisciplinary, 44.82 for Physical Sciences, and 36.65 for Health Sciences. From the subject area perspective, it can be said that 25.92% of disciplines clearly define an upper limit for the number of bibliographic references per article. However, in 70.37% of the subject areas, the average number of bibliographic references per article is at least 40% higher than the overall limit of 30 bibliographic references per article. A shown in Chart 9 (below), the only 3 disciplines in which the average number of bibliographic references per article was under the limit of 30 bibliographic references per article were Health Professions (27.10), Computer Science (28.85), and Mathematics (22.76). The averages of bibliographic references per article are addressed within Graphic 9. Values for the sequences between 131-140, 171-220 and 231-270 bibliographic references per article resulted null and, therefore, are not shown in Graphic 9.

Graphic 9. Distribution of articles per subject area, according to the number of bibliographic references included in bibliographic references lists

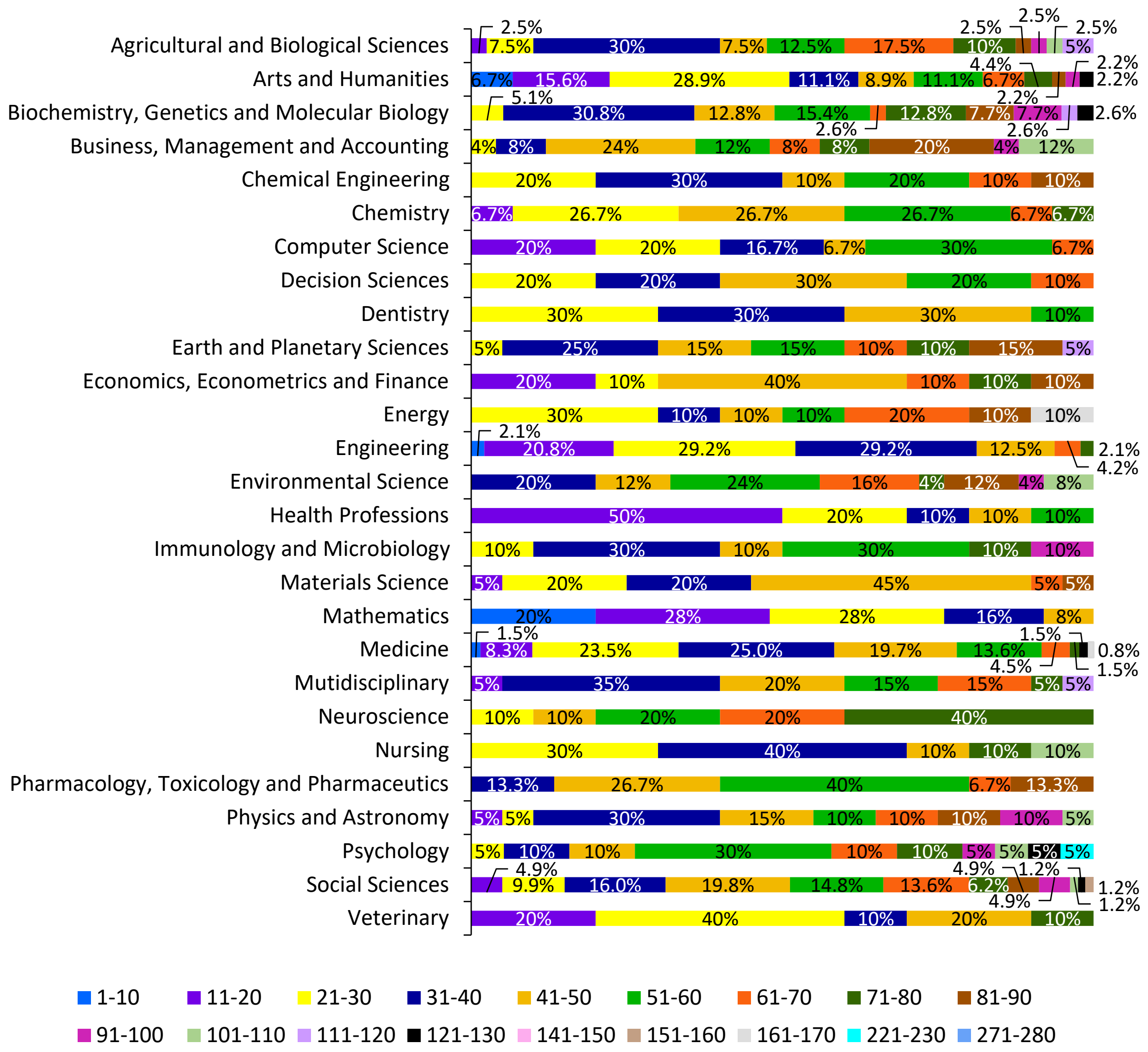


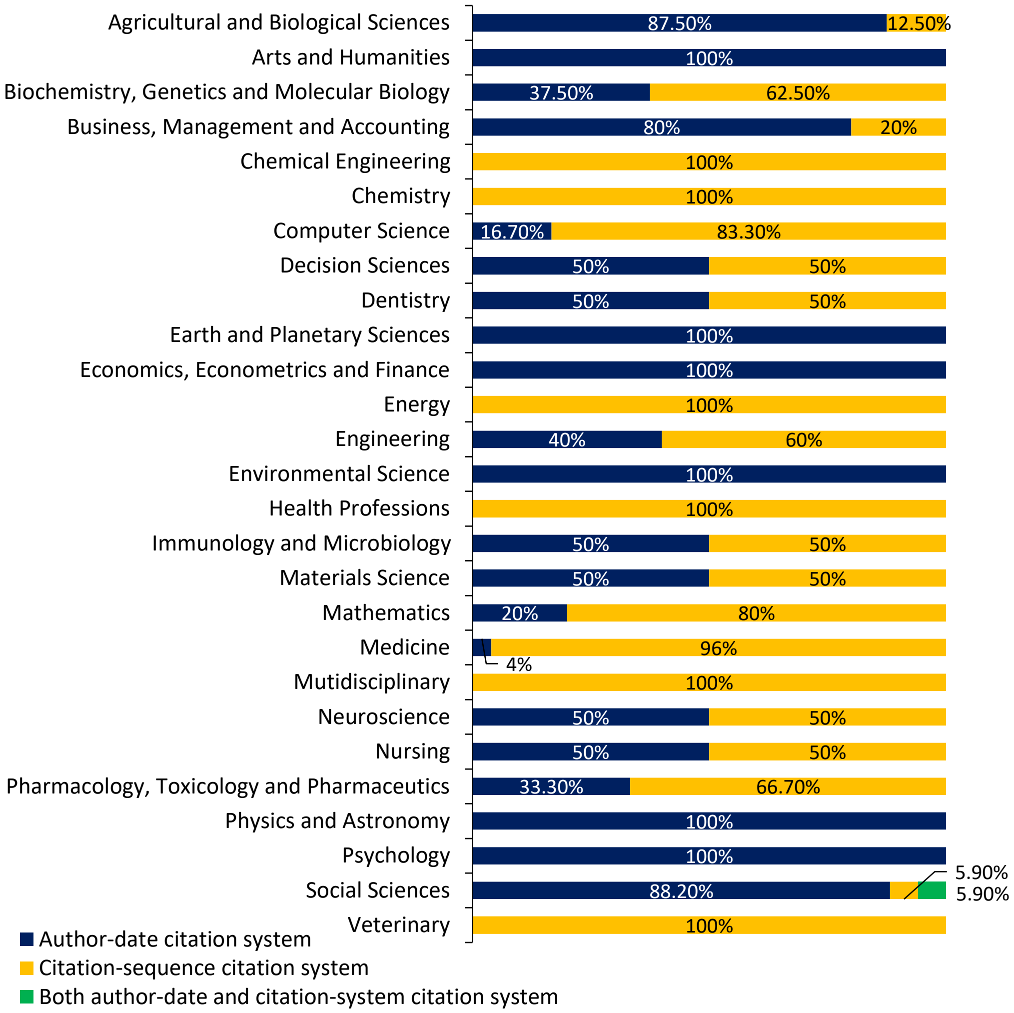
Chart 9 - Classification of subject areas, according to the percentage variation of the number of bibliographic references included in articles' bibliographic references lists

Subject areas	Average bibliographic references per article	Average mentions and quotations per article
Agricultural and Biological Sciences	57.03	86.15
Arts and Humanities	38.93	70.44
Biochemistry, Genetics and Molecular Biology	57.15	90.36
Business, Management and Accounting	65.68	111.68
Chemical Engineering	45.20	70.60
Chemistry	45.47	56.40
Computer Science	28.85	55.56
Decision Sciences	42.80	74.90
Dentistry	38.10	53.30
Earth and Planetary Sciences	57.85	97.10
Economics, Econometrics and Finance	45.70	61.10
Energy	58.30	92.40
Engineering	30.63	45.02
Environmental Science	61.88	87.60
Health Professions	27.10	44.70
Immunology and Microbiology	52.10	78.60
Materials Science	42.25	57.85
Mathematics	22.76	36.00
Medicine	40.45	63.67
Multidisciplinary	49.85	84.90
Neuroscience	60.60	84.80
Nursing	44.00	78.50
Pharmacology, Toxicology and Pharmaceutics	54.73	80.87
Physics and Astronomy	55.05	88.00
Psychology	76.80	149.55
Social Sciences	56.43	96.15
Veterinary	33.60	56.10

From a broader view, considering subject category perspective, the incrementing percentage considering the gap between the minimum and the maximum number of bibliographic references comprised within articles' bibliographic references lists were: 237% for Life Sciences, 352% for Physical Sciences, 484% for Multidisciplinary, 534%

for Health Sciences and 621% for Social Sciences. From a more granular perspective, considering the subject are perspective, the lowest variation range registered was 100% for Dentistry subject area and impressive 2925% for Arts and Humanities subject area.

Graphic 10. Percentual distribution of articles per subject area, according to the citation system adopted



The analysis identified two citation systems within articles composing the sample, according to data represented within Graphic 10, i.e., the citation-sequence system (which uses numbers to refer to bibliographic references within the text, e.g., “[3]”) and the author-date system (which uses the surnames of the authors of the article plus a date, e.g., “(Doe et al., 2020)”). On average, the citation-sequence system is adopted by 51.37% of the article's sample, against 48.41% adopting the author-date system. This scenario prevents assuming one of the citation systems as the most adopted overall. However, by considering the subject categories perspective, we can say that

Health Sciences, Physical Sciences, and Multidisciplinary articles tended to adopt citation-sequence system (respectively 79%, 57% and 51% of journals, on average), Social Sciences and Life Sciences tend to adopt author-date system (respectively 86% and 52% of journals, on average).

Basically, in-text reference pointers referring to mentions and quotations considering articles adopting the author-date system, provided (at least) the cited author's surname and the year of publication of the cited work. For works authored by more than one author, the in-text reference pointers usually provided only the first surname of the first author of the cited work, followed by the Latin expression et. al. and its year of publication (e.g., Schmidt et al., 2021). The exceptions were observed for the cases in which there were mentions or quotations of two or more publications published in the same year, whose first author has the same surname. In these cases, in-text reference pointers referring to mentions usually provided the cited work's author' surname and the year of publication. As for in-text reference pointers referring to quotations, they (should) provide the exact location (e.g., the page number) where the quoted excerpt could be found in the cited work, in addition to the cited work's author' surname and the year of publication.

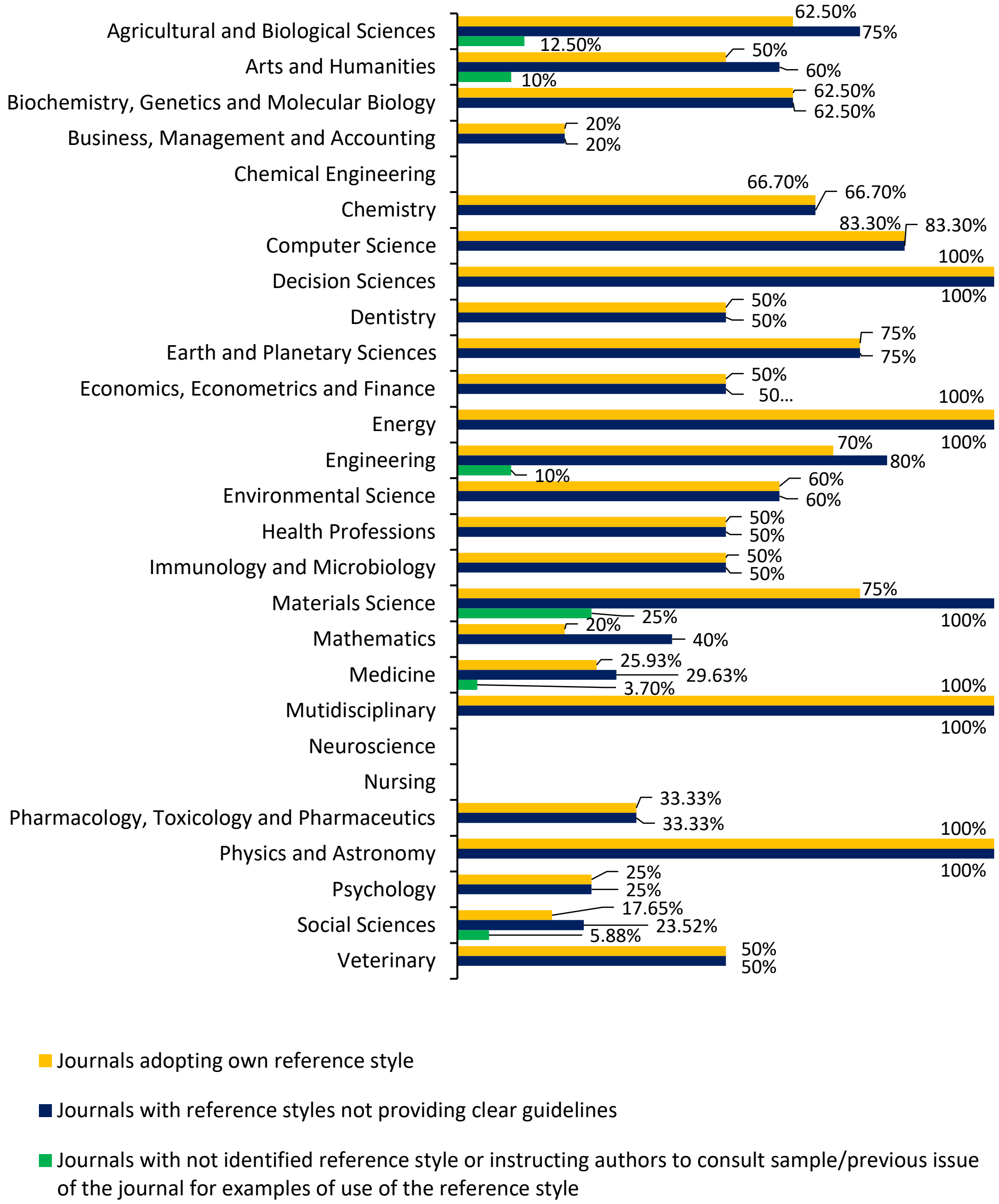
It is worth mentioning the case of articles from a journal representing 5.9% of total Social Science articles in which we observed the adoption of both author-data and citation-sequence systems, simultaneously within the same article. In such articles, superscript numbers were used as in-text reference pointers for marking up mentions and quotations. At the end of the article, there was a session called "Notes" containing a list of bibliographic references sorted under an ascending numerical order whose numbers corresponded to the superscript numbers introduced in the text body, like "1 Key 1949, Munro 1920." (i.e., the expected data to be provided by in-text reference pointers referring to mentions and quotations in an article adopting the author-date system). At the end of the "Notes" session, there was the "References" session, containing the bibliographic references of the works cited in the text body, alphabetically ordered. So, to identify the works cited by the articles published in this Social Science journal, readers are expected to use the superscript numbers used as in-text reference pointers in the text body to identify the surname of the cited author and the year of the publication of the cited work according to the data contained in the

“Notes” session. Only then, readers should be able to consult the list of references to identify the work cited in the text body. This configures a clear example in which, although the necessary metadata for identifying the cited works are all provided, the editorial politics adopted by such journal do not favor its access and interpretation by the reader.

A similar case was observed in articles from an Arts and Humanities journal, with the difference that instead of providing complete in-text references pointers in the “Notes” session, in this case, the in-text reference pointers considered in the text bodies referred to the bibliographic references appearing in the footnotes. However, the articles from this journal also included an alphabetically sorted list of bibliographic references referring to the works cited in the text body and, therefore, those bibliographic references included in footnotes were absolutely dispensable.

Graphic 11considers a data-cross checking and evinces the correspondence between the rates of publishers adopting “own reference styles” and the rates of reference styles not providing clear instructions on referencing and/or citing.

Graphic 11. Approaches concerning the relation between reference styles authored by publishers and the level of clarity of the guidelines for formatting citing and referencing data

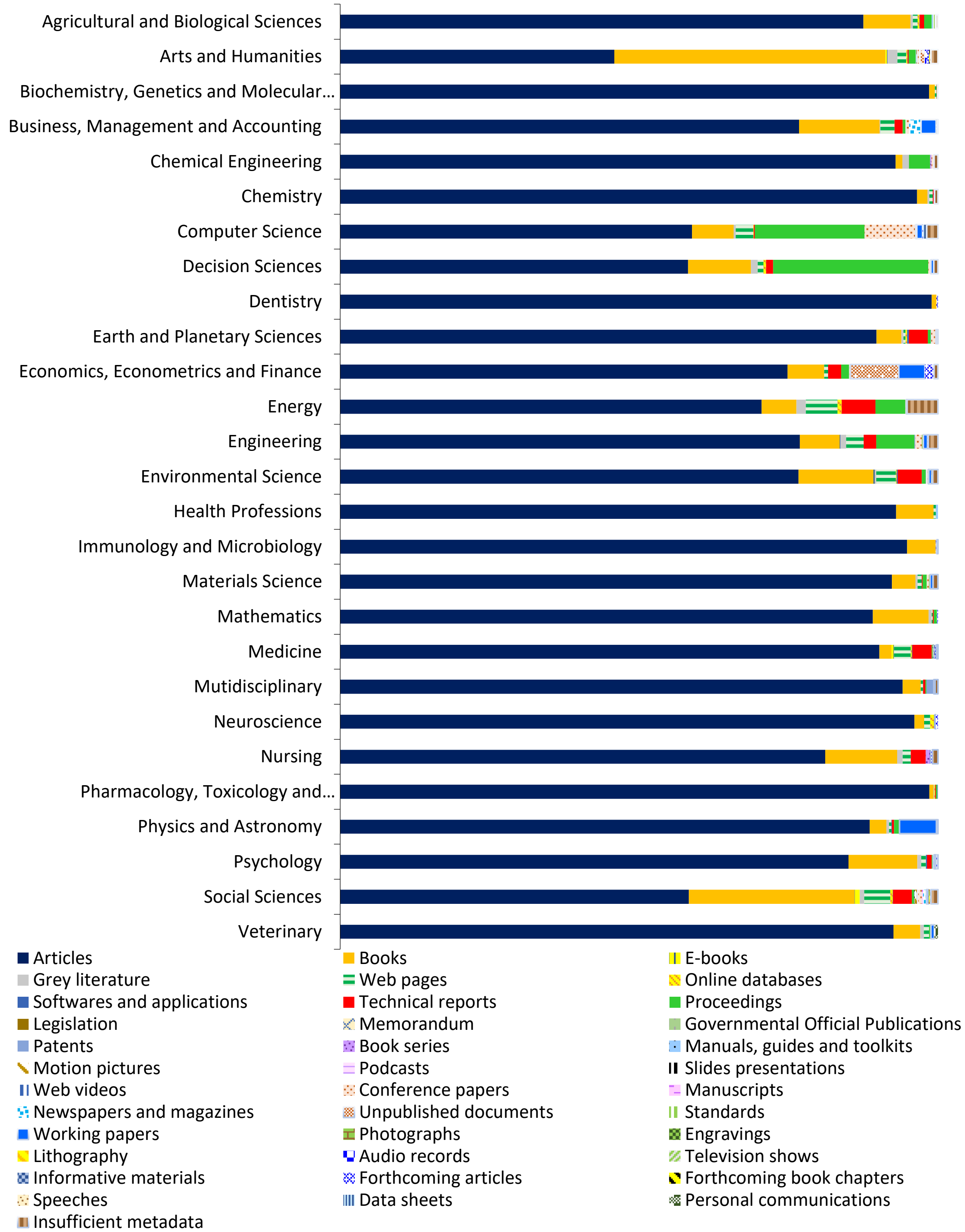


It should be mentioned the portion of journals for which it was not possible to identify the adopted reference style (1.90% of journals on average) and, the portion of 10% of Arts and Humanities journals and 5.88% of Social Sciences journals which, besides not clearly stating the adopted reference style, still instruct authors to consult any previous issue of the journal or the sample article provided by the publisher which should be considered as guidelines for formatting, citing and referencing metadata.

Although the elements provided by publishers are insufficient even to classify the instructions concerning their level of clarity, they were put together in Graphic 11 to support the statements on how publishers may be negligent concerning citing and referencing matters.

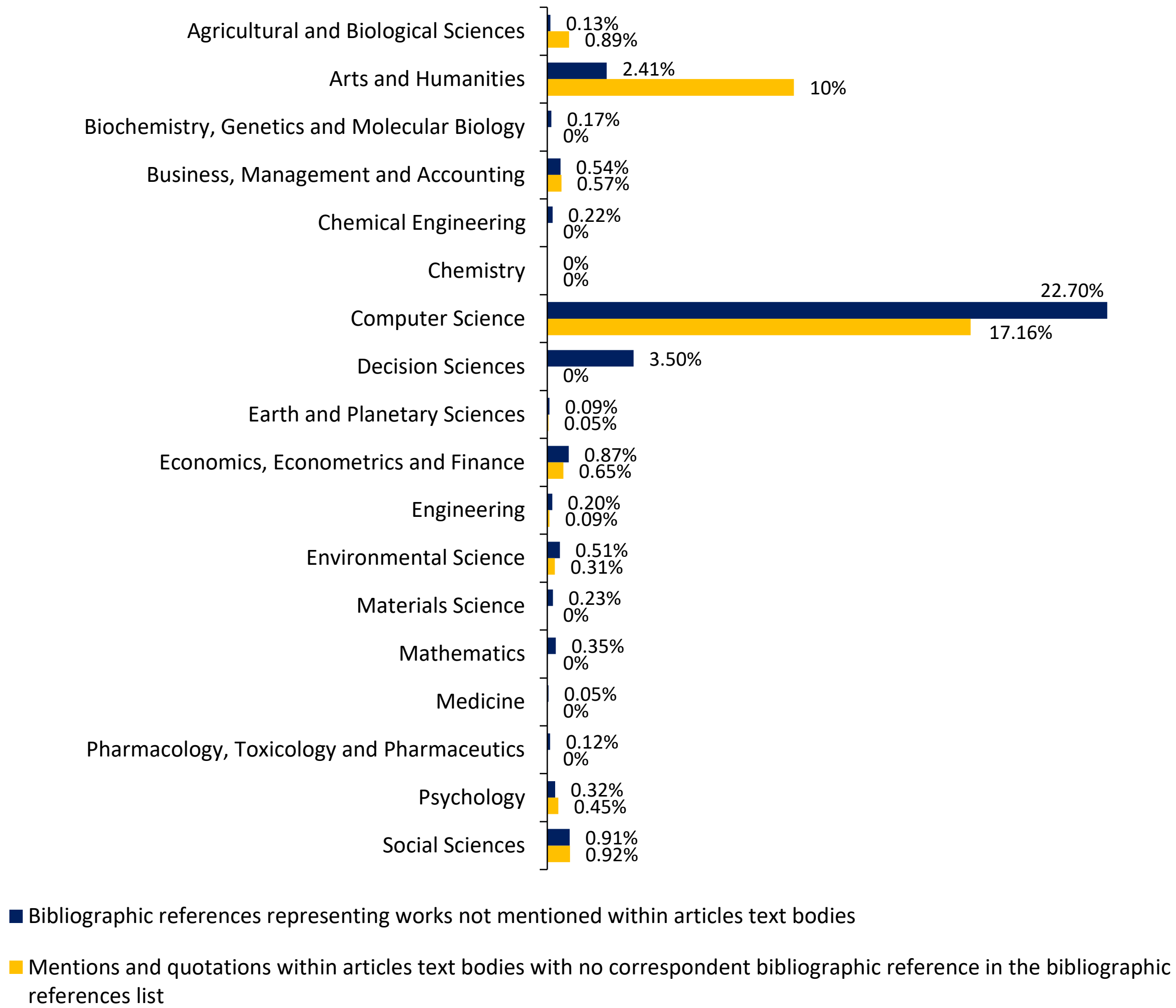
36 different types of publications were detected among the works cited by the articles of the sample which, on average, represent the following percentages of the whole sample: articles (83.55%), books and book chapters (7.93%), proceedings (2.53%), web pages (1.30%), technical reports (1.17%), working papers (0.67%), conference papers (0.51%), thesis and dissertations (0.47%), unpublished documents (0.35%), newspapers and magazines (0.10%), online databases (0.13%), e-books (0.07%), patents, software and applications (0.05% each), book series, manuals, guides and toolkits (0.04% each), datasheets, personal communications, standards and, web videos (0.03% each), motion pictures (0.02%), audio records, informative materials, forthcoming articles, forthcoming book chapters and, manuscripts (0.01% each). Legislations, memorandums, governmental official publications, podcasts, slides presentations, and photographs showed average rates equal to or lower than 0.004% each. Regarding works presented in scientific events, we noticed that bibliographic references may refer to the conference paper itself or its version published in the proceedings, corresponding to similar types of publications. In addition, sometimes (e.g., in the Computer Science subject area) books and, more rarely, journals issues may contain the proceedings of scholarly events as well. A visual representation of those approaches is provided in Graphic 12, considering the subject area's perspective. The “insufficient” data category represents the average rate of bibliographic references per discipline whose provided metadata do not allow the precise identification of the type of publication they refer to.

Graphic 12. Distribution of bibliographic references according to the type of publication they refer to (considering the whole sample)



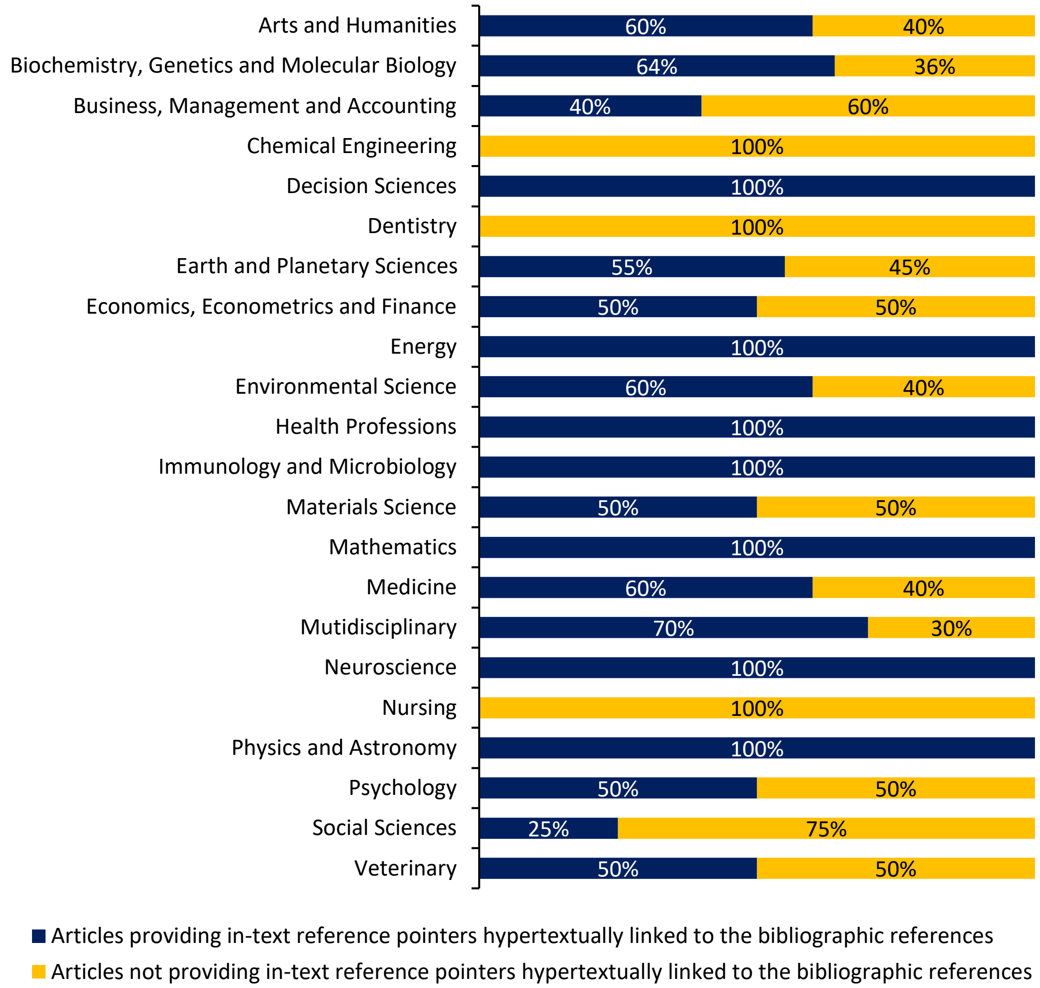
It is expected that the number of bibliographic references found in a bibliographic references list is equivalent to the number of works cited by the corresponding citing work. However, this situation was not true among the articles of our sample. Graphic 13 shows the rates of bibliographic references that were not denoted in the text body, and vice-versa. The in-text reference pointers associated to mentions and quotations that did not match any of the bibliographic references were counted as non-cited referenced mentions or quotations. The subject areas for which we did not observed any of these behaviors – namely Dentistry, Energy, Health Professions, Immunology and Microbiology, Multidisciplinary, Neuroscience, Nursing, Physics and Astronomy and Veterinary – are not shown in Graphic 13. Also, it is worth mentioning that data extracted from articles of a single Computer Science journal are entirely responsible for the rates retrieved for this discipline shown in Graphic 13. The indexes also comprise cases of in-text reference pointers denoting shared authored cited works, in which authors are given in full in the first mention/quotation of the cited work in the text and in abridged form (i.e., using the expression “et al.”) within the other in-text reference pointers referring the second and subsequent mentions/quotations of the cited work within the text.

Graphic 13. Percentual distribution of bibliographic references not denoted in text bodies and mentions and quotations not corresponding to a bibliographic reference included in the article's bibliographic reference list, per subject area



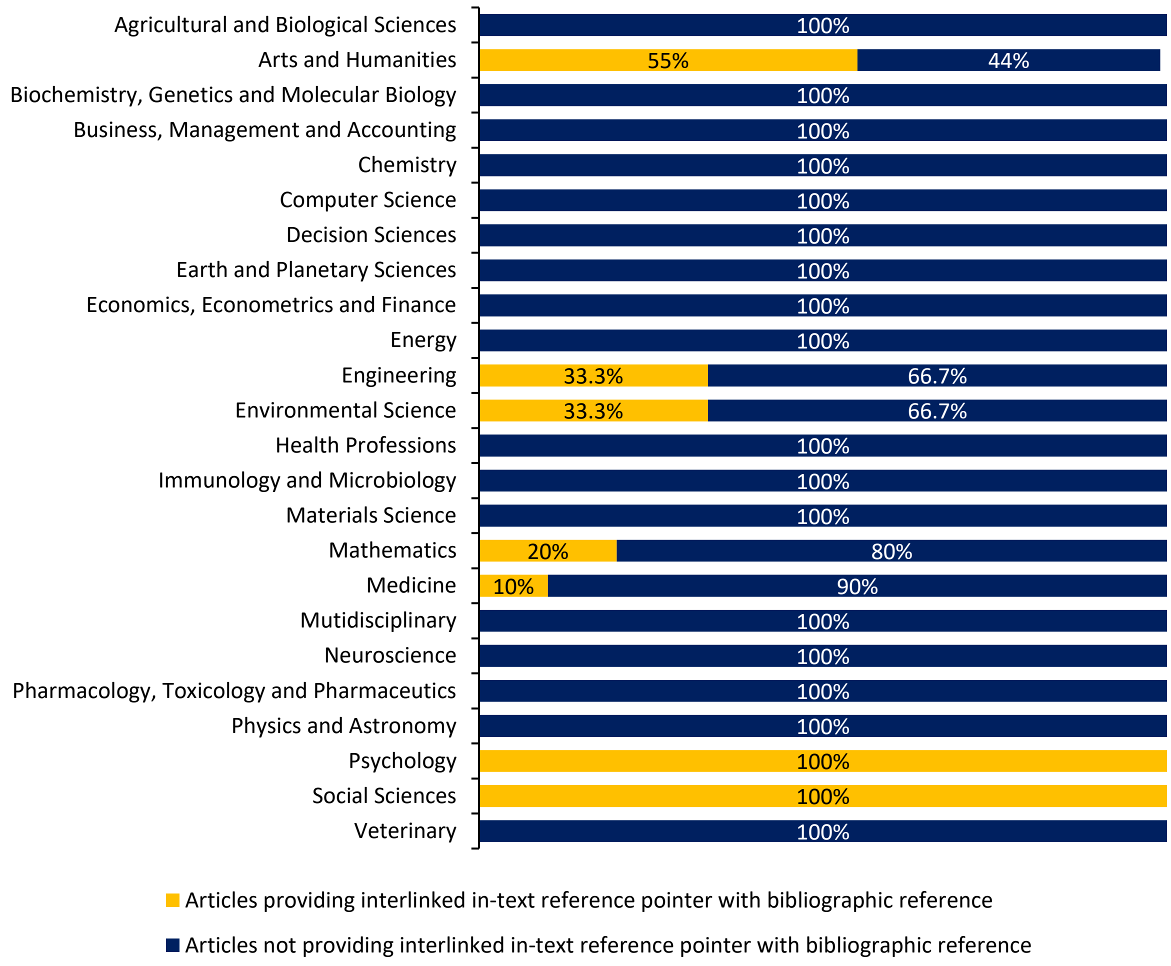
Another feature that brings the correspondence between in-text reference pointers and bibliographic references into evidence is the hypertextually-linked in-text reference pointer. For instance, by clicking on this tool the reader should be sent to the denoted bibliographic reference. The Graphic 14 addresses the percentual rate of articles providing such functionality, per subject area.

Graphic 14. Percentual distribution of articles per subject area providing in-text reference pointers hypertextually linked to the respective bibliographic references in the bibliographic references lists



The reverse way also can be true, i.e., when the reader is expected to be sent to the in-text reference pointer in the text body by clicking on the correspondent bibliographic reference in the bibliographic reference list. We called this functionality *interlinked in-text reference pointer with bibliographic reference*, which is addressed in Graphic 15. It is important to clarify that providing in-text reference pointers hypertextually linked to the respective bibliographic references in the bibliographic references lists is a mandatory functionality for the interlinked in-text reference pointer with bibliographic reference. So, all data considered in Graphic 15 is a micro view of the portion of the sample providing in-text reference pointers hypertextually linked to the bibliographic references (i.e., the Blue bars in Graphic 14).

Graphic 15. Percentual distribution of articles per subject area providing interlinked in-text reference pointer with bibliographic references



Some subject areas showed null rates for both functionalities, namely Agricultural and Biological Sciences, Chemistry, Computer Sciences, Engineering and Pharmacology, Toxicology and Pharmaceutics. In contrast, 100% of articles from Health Professions, Decision Sciences, Immunology and Microbiology, Neuroscience, Energy Mathematics, and Physics and Astronomy subject areas provide this feature. On average, 61.83% of articles offer such a feature. From this total, 11.40% of articles provide interlinked in-text reference pointer with bibliographic reference. From a macro view, Life sciences showed a more frequent adoption of this feature, with 84% of articles.

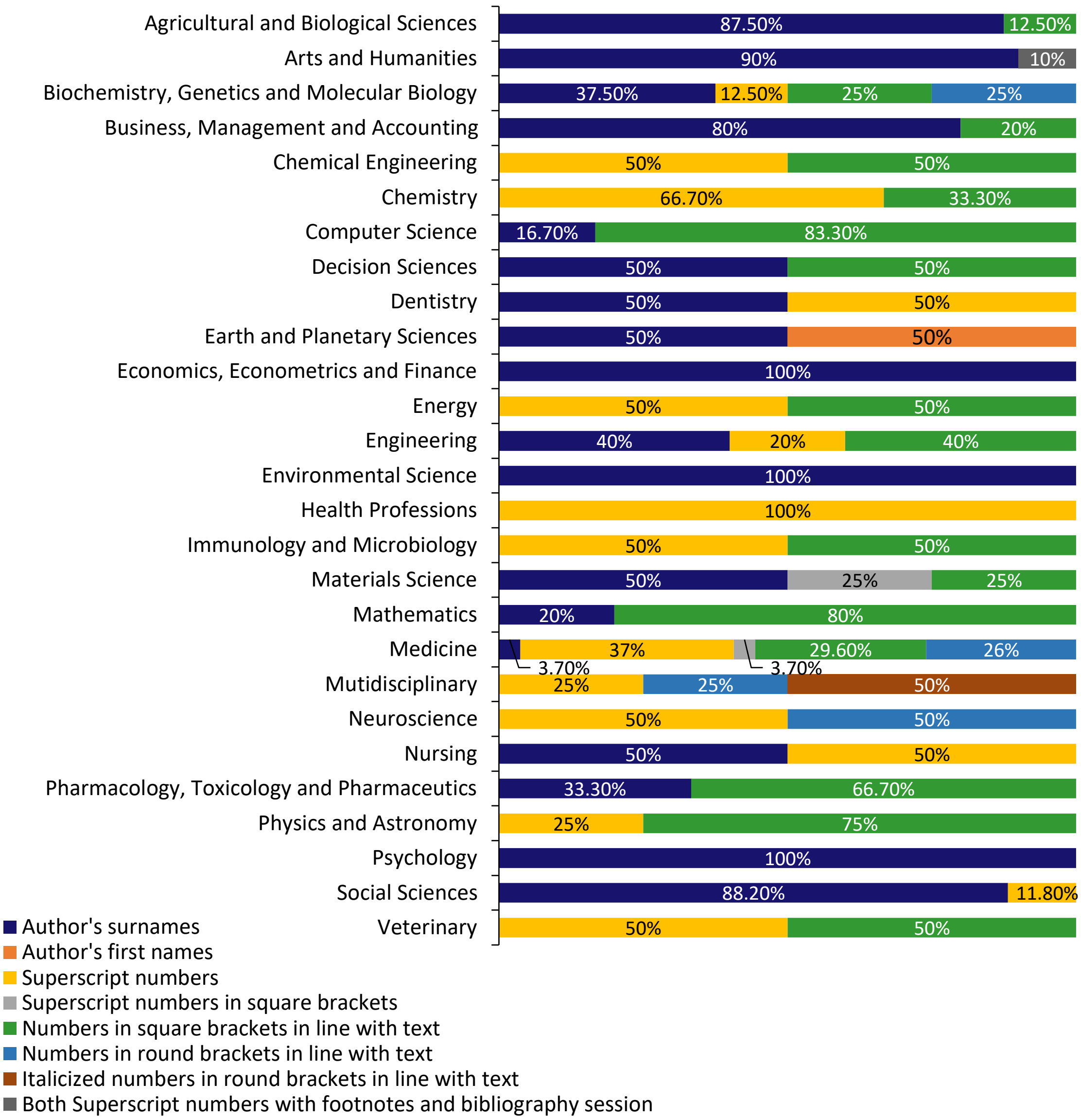
Besides the proper correspondence between in-text reference pointers and bibliographic references, another feature that can facilitate access to the cited works is the provision of DOIs, DOI URLs, or generic URLs within bibliographic references.

Health Professions and Physics and Astronomy's articles stand out as those in which the provision of these kinds of metadata is more frequent, with 92% of bibliographic references providing URL and 29% providing DOIs or DOI URLs for Health Professions and 57% of bibliographic references providing URL and 88% providing DOIs or DOI URLs for Physics and Astronomy. A noteworthy subject area is Chemical, in which we did not detect any of the metadata in question. Considering exclusively the provision of DOIs and DOI URLs, Physics and Astronomy, and Psychology assume a prominent position, with 88% and 78% of bibliographic references compliant with these criteria, respectively.

Non-textual content, i.e., photos, maps, drawings, and several ways of representing information other than the written format, also constitute citable contents, regardless of their embodiment (i.e., printed or digital) and the type of publication (books, articles, etc.) which represent them.

Graphic 16 addresses the structure of in-text reference pointers across the disciplines. With exception of a few subject areas – i.e., Economics, Econometrics and Finance, Environmental Science, Health professions, and Psychology – there is no uniformity in the way in-text reference pointers are used in text bodies, neither across subject areas nor across subject categories.

Graphic 16. Percentual distribution of articles, according to the structure of in-text reference pointers across subject areas



Biochemistry, Genetics and Molecular Biology, Engineering, and Medicine are among the subject areas with the greater variety of in-text reference pointers formats. According to Graphic 7, those disciplines are also among the ones adopting the widest variety of reference styles.

The provision of the page numbers where an external cited content (mentioned or quoted) can be found in the cited work is approached in Graphic 17. In-text reference pointers referring to mentions providing pagination data and the in-text reference pointers referring to quotations not providing pagination data were considered as non-conformities, since we did not detect any reference style referring to the journals composing our sample clearly recommending such posture. Graphic 17 considered only the subject categories in which we observed non-conformities concerning the data provided in the in-text reference pointers referring to quotations, i.e., in-text reference pointers referring to mentions, providing the pagination where the mentioned content can be found in the cited work and, in-text reference pointers referring to quotations not providing the pagination where the mentioned content can be found in the cited work. Because of that, the following disciplines were disregarded from Graphic 17: Biochemistry, Genetics and Molecular Biology, Chemical Engineering, Chemistry, Computer Science, Decision Science, Dentistry, Earth and Planetary Sciences, Energy, Immunology and Microbiology, Neuroscience, Pharmacology, Toxicology and Pharmaceutics, and Veterinary.

Graphic 17. Percentual distribution of in-text reference pointers referring to mentions and quotations considering the provision of the page numbers of the cited passages in the cited works

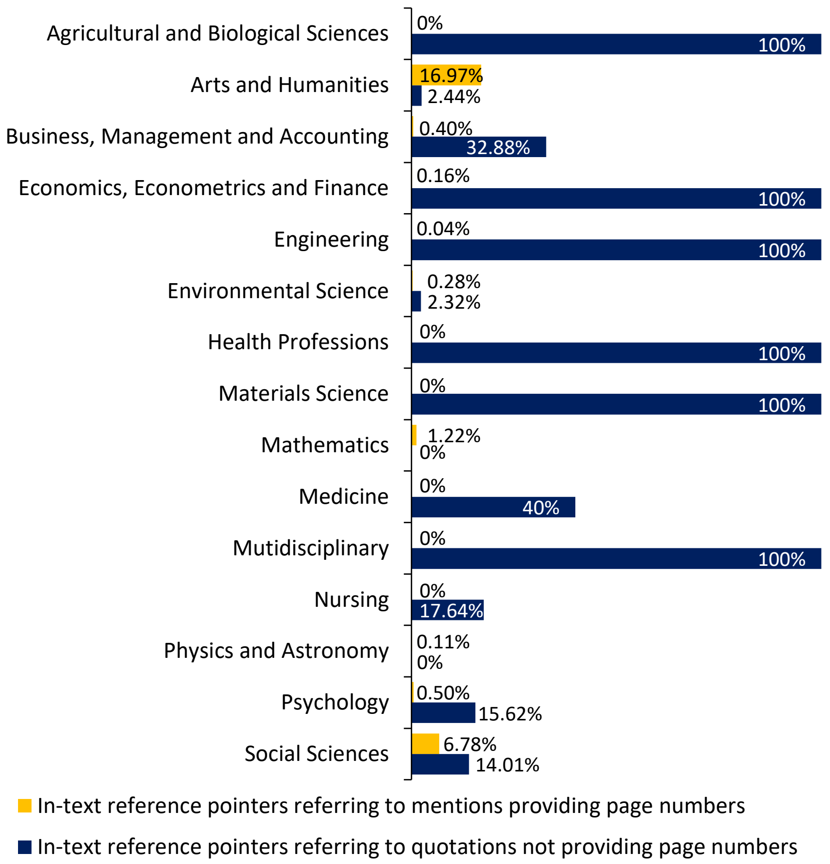


Chart 10 introduces the variety of markups adopted by articles to identify the length of quotations within text bodies and to distinguish them from the self-authored content.

Chart 10 - Percentual distribution of articles per subject area according to the markups adopted for run-in quotations and long quotations

	Run-in quotations							**Long quotations**					
	1	2	3	4	5	6	7	A	B	C	D	E	F
Agricultural and Biological Sciences	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Arts and Humanities	47.06%	0%	2.94%	41.18%	0%	0%	0%	32.35%	2.94%	17.65%	0%	0%	0%
Biochemistry, Genetics and Molecular Biology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Business, Management and Accounting	73.33%	0%	0%	6.67%	6.67%	0%	0%	26.67%	0%	0%	0%	0%	0%
Economics, Econometrics and Finance	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Engineering	83.33%	0%	0%	16.67%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Environmental Science	40%	0%	0%	40%	0%	0%	0%	40%	0%	0%	0%	0%	0%
Health Professions	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Materials Science	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Medicine	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5.88%
Multidisciplinary	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Nursing	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Physics and Astronomy	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Psychology	90%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%
Social Sciences	70.69%	0%	0%	24.14%	1.72%	0%	0%	18.97%	0%	12.07%	0%	3.45%	0%

Legend for Chart 10

1	Double quotation marks
2	Double quotation marks and italic font
3	Double quotation marks some italics and some not italics
4	Single quotation marks
5	Part in single, part in double quotation marks
6	Italic normal font size
7	Part in double quotation marks, part in italic normal font size

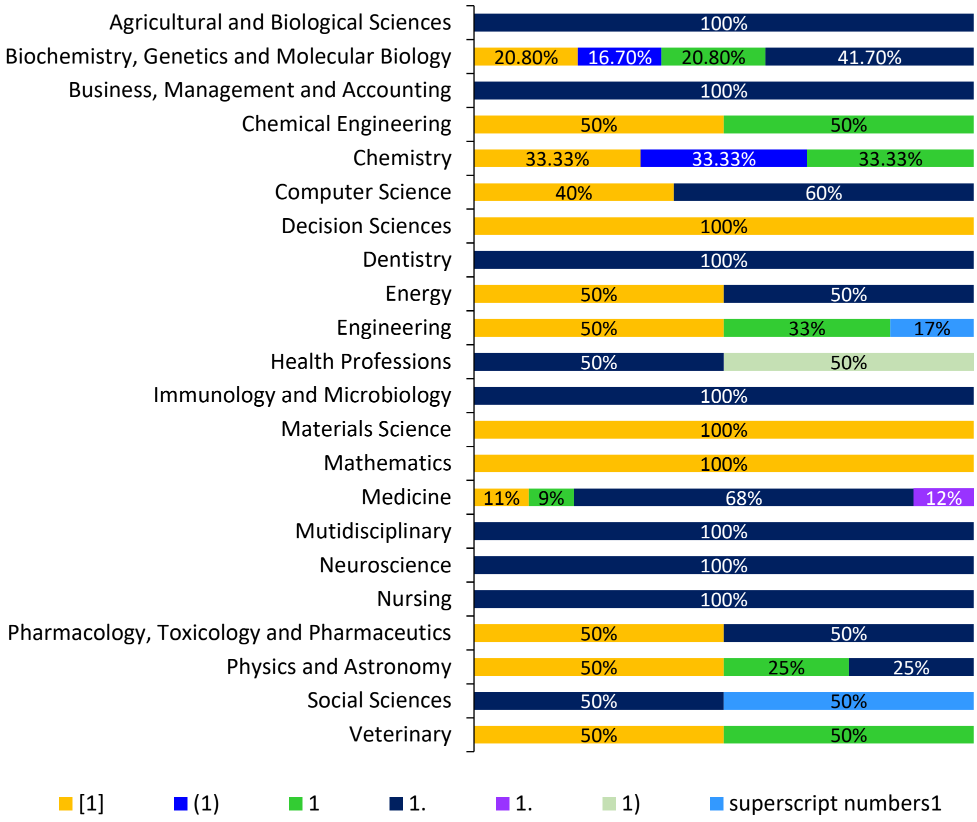
A	Indented in a smaller font size
B	Indented smaller font size, part italics, part not italics
C	Indented in normal size font
D	Indented, in double quotation marks and italic normal font size
E	Indented, in double quotation marks and italic smaller font size
F	Aligned text in a smaller font size

The criteria used for determining the use of markups for long quotations were not regular. For instance, we noticed articles adopting indentation as markup for quotations longer than 80 words; others considered the length of the quotation (in lines) as a parameter for determining the use of indentation or not. Some articles have indented quotations longer than two lines, while others have indented quotations longer than 4 lines. A similar situation was observed for the adoption of other formatting resources, i.e., font size and italics characters.

In some cases (i.e., 1 journal in Social Science, 1 in Business and Accounting, and 2 in Arts and Humanities) both single and double quotation marks were considered as markups for delimiting quotations within the same article. Another point observed is that there is no uniformity even among the unusual scenarios. For instance, some in-text reference pointers concerning mentions in the same article provide the pagination of the mentioned excerpt in the cited work and others do not.

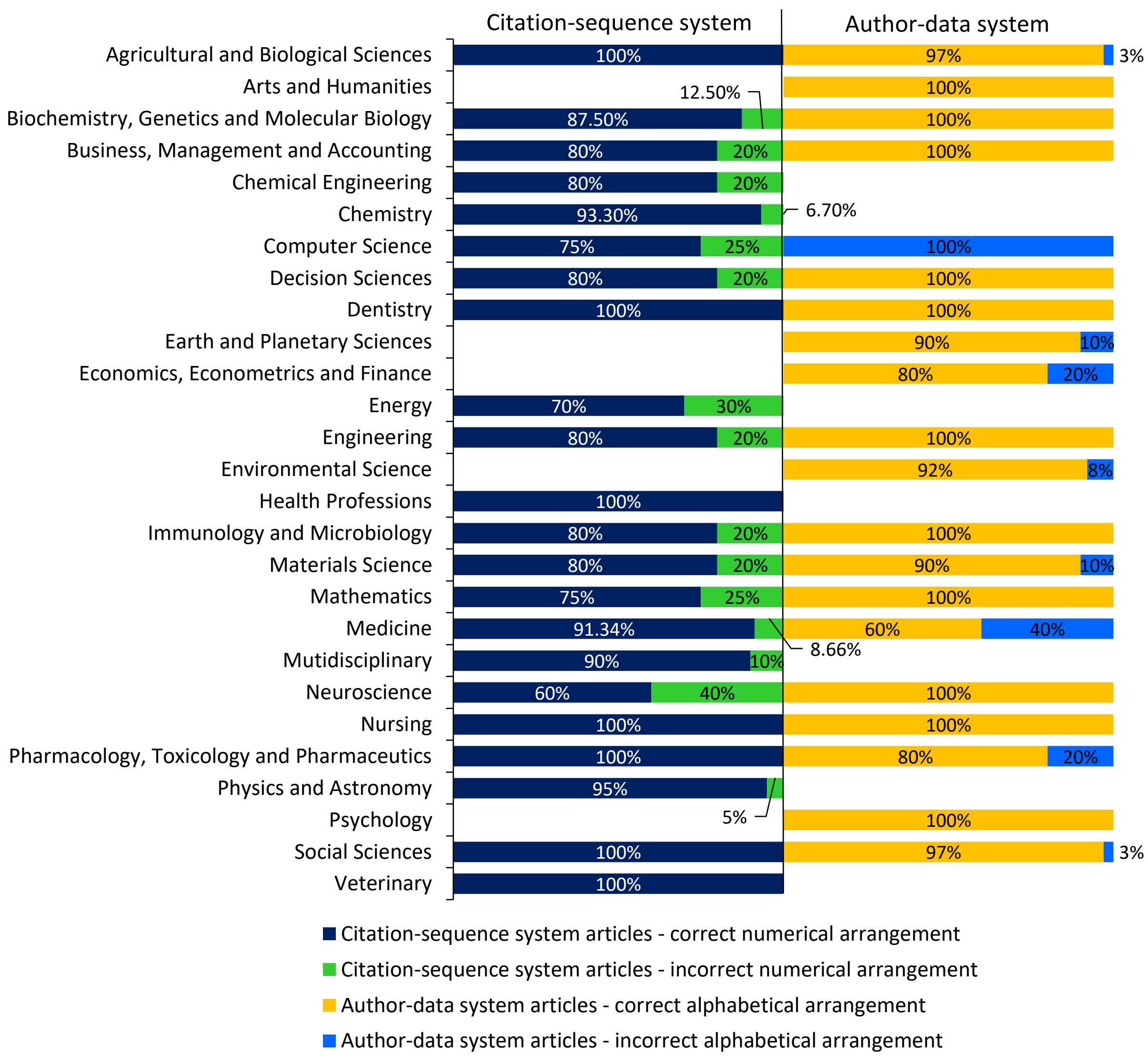
The format of the numeral designating bibliographic references in the bibliographic references lists of articles adopting the citation-sequence system is not uniform. Such variations are shown in Graphic 18, which does not consider Arts and Humanities, Earth and Planetary Sciences, Economics, Environmental Science, and Psychology subject areas, since none of their articles composing the same adopted the citation-sequence system.

Graphic 18. Percentual distribution of articles per subject area, adopting the citation-sequence system, according to the format of the numeral accompanying bibliographic references in the bibliographic reference lists



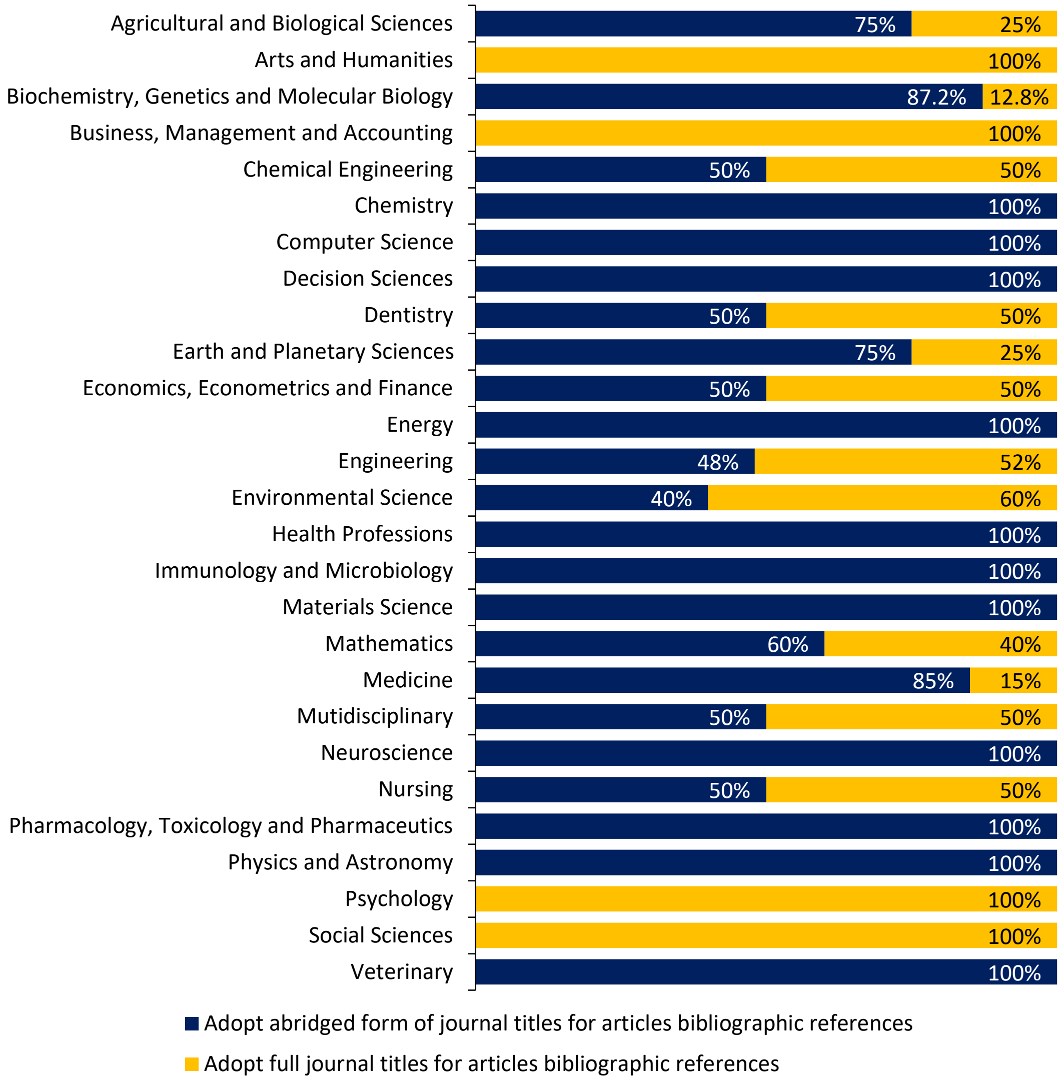
Other important connecting element between in-text reference pointers and bibliographic references which is rarely approached by reference styles, is the bibliographic references list assortment. Usually, bibliographic references included in citation-sequence system articles are ordered numerically in ascending order and the numbers attributed to each bibliographic reference correspond to the in-text references pointers referring to mentions and quotations from the work represented by such particular bibliographic reference under the corresponding numeral in the bibliographic references list. Still considering articles adopting citation-sequence system, the numbers accompanying the bibliographic references in the bibliographic reference lists do not always correspond to the format of those denoting them through in-text reference pointers in the text body, what also may configure an obstacle for the proper correspondence between them. As for articles adopting author-date system, in which the bibliographic references are alphabetically ordered, sorting errors may prevent the reader from correlating in-text reference pointers to the respective bibliographic reference, if the alphabetization mistakes are not perceived by the reader. Graphic 19 addresses the percentual of articles per subject category in which we observed non-conformities in those issues. It should be clarified that the numbers concerning author-date system rates for Computer Science subject area refers to errors perceived within articles from the unique journal of this discipline adopting such citation system. Graphic 19 indicates that, on average, 10.48% and 7.93% of articles adopting the citation-sequence system and author-data system, respectively, do not provide a proper numerical/alphabetical sorting in their bibliographic references list.

Graphic 19. Percentual distribution of articles per subject area, considering the sorting of bibliographic references within bibliographic references list



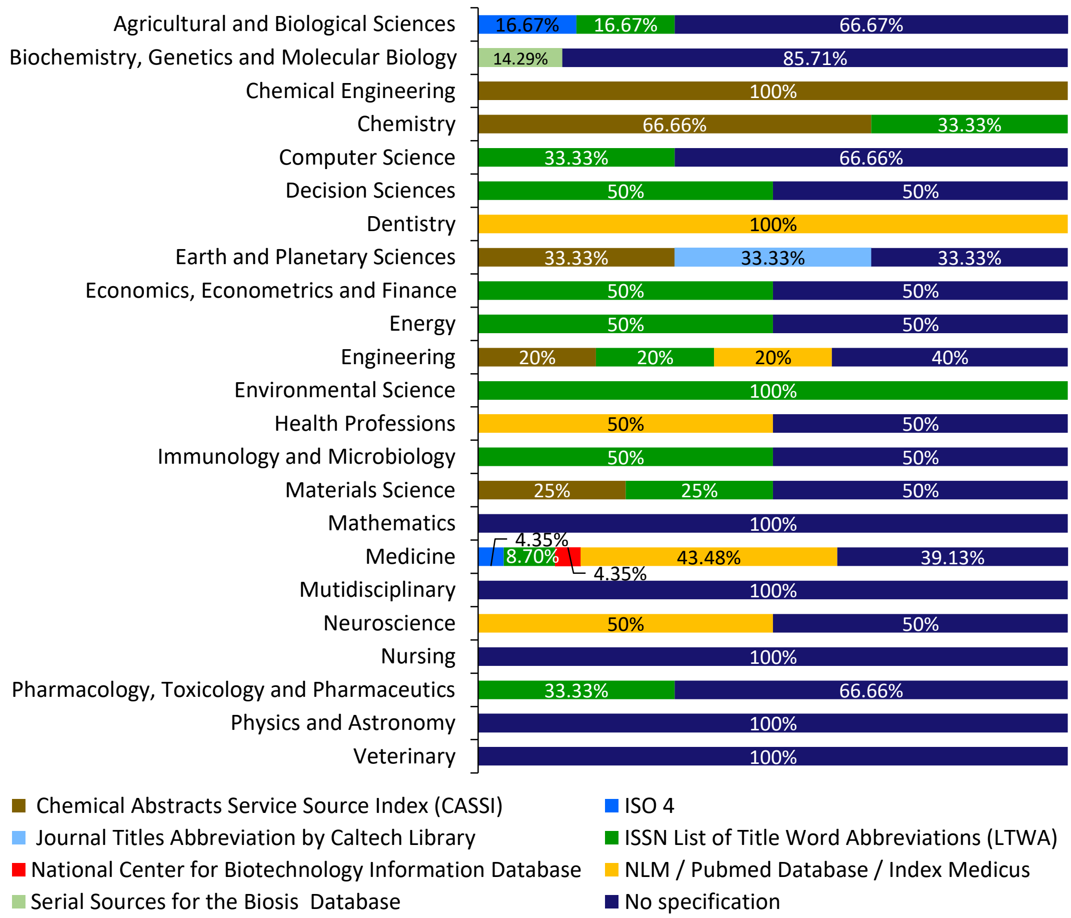
Concerning how the titles of cited journals are transcribed within bibliographic references, the abridged format is adopted by 67.41% of journals. From the subject category perspective, Health Sciences, Life Sciences, and Physical Sciences tend to adopt abridged versions of journals' titles within bibliographic references, while Social Sciences' journals tend to give journals' titles in full. Only 14.81% of disciplines totally consider the full title of articles: Arts and Humanities, Business, Management and Accounting, Psychology and Social Sciences, as shown in Graphic 20.

Graphic 20. Percentual distribution of journals according to the format in which cited journals' titles are transcribed in bibliographic references: in full or in abridged format



Graphic 21 focuses on articles adopting the abridged format of cited journal titles within the bibliographic reference and, therefore, disciplines whose 100% of journals adopt full journal titles in bibliographic references were disregarded. Data represent the rates of journals per discipline according to the source in which the abbreviations for cited journal titles considered within bibliographic references are based on. In 65.22% of the subject areas, different abbreviation sources for journal titles were detected across journals of the same area. However, 46.23% journals in our sample did not specify the source from where the abbreviations used in the bibliographic references of their articles are based on.

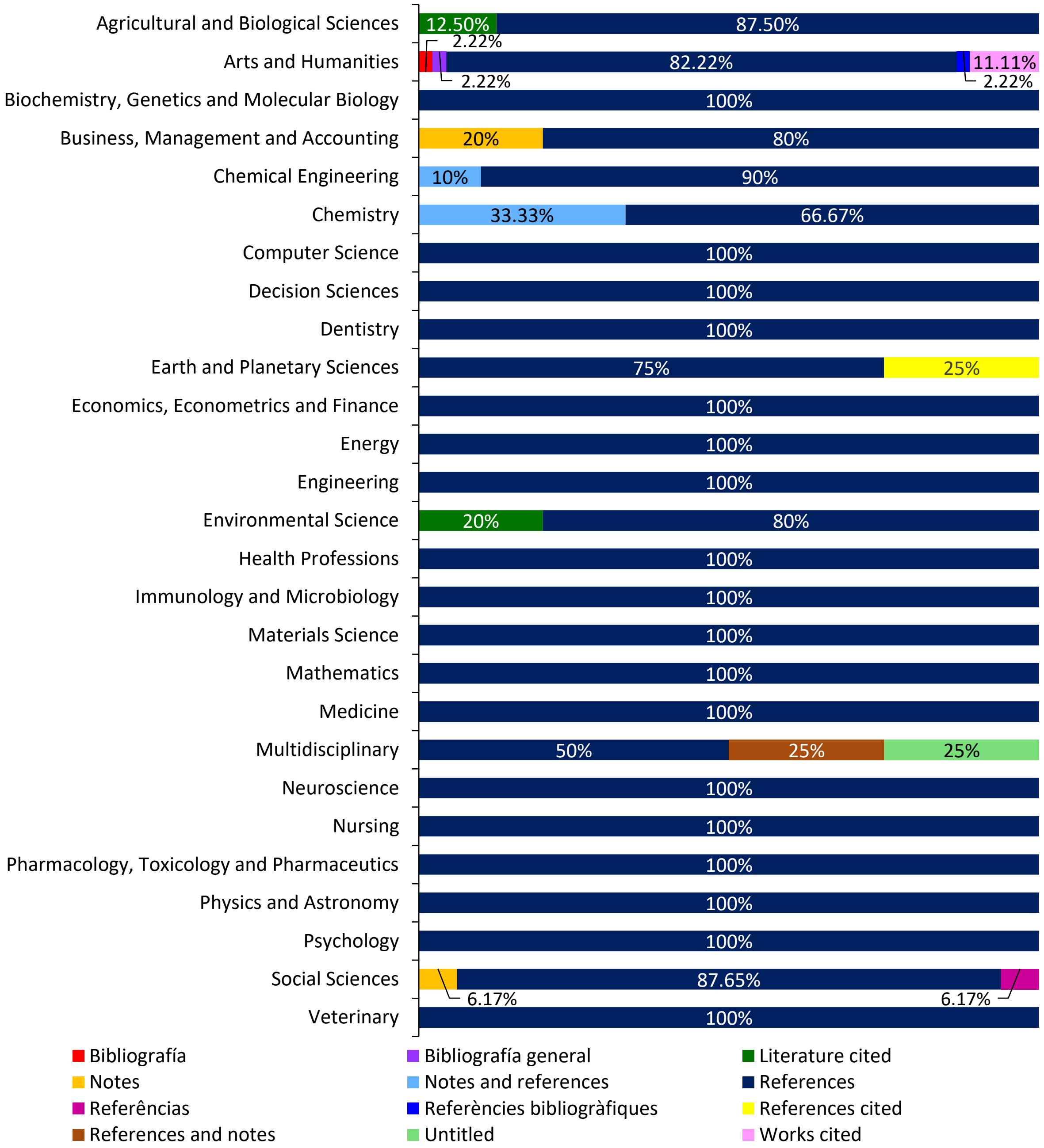
Graphic 21. Percentual distribution of journals considering the source adopted for abbreviations of journal titles of cited articles within bibliographic references



The lack of standardization is also reflected even within the title of the bibliographic references section in which were detected 12 variations across articles of our sample. On average, 93% of the articles entitle their bibliographic references session as “References”. It worth mentioning the particular cases of an Arts and Humanities article adopting both the section “Referències bibliogràfiques” and the section

“Bibliografia” (probably due to the non-English language used for articles’ text), and a Chemical Engineering journal in which part of its articles name the reference section “References” and the other part name it “Notes and References”, as visually represented in Graphic 22.

Graphic 22. Percentual distribution of articles per subject area, according to the title of the section containing bibliographic references



Besides providing or recommending the adoption of the guidelines of a specific reference style, 53.01% of publishers, on average, simultaneously suggest authors to use reference managers to format bibliographic references of their works. Chart 11 shows the rate of the recommendation of each reference manager per subject area. Some journals simultaneously recommend multiple reference styles, some recommend none, and others refrain from mentioning a particular one (i.e., “any reference manager” column), although recommending the use of reference managers. The 50% of Materials Science journals mentioned in the “Not identifiable” column refers to journals in which the instructions provided to authors do not even clarify whether the journal encourages the use of reference managers or not.

Chart 11 - Percentual distribution of journals by subject areas according to the reference manager recommended to authors by publishers (considering only journals recommending reference managers to authors)

Subject Areas / Reference managers	Products supporting CSL Styles	Bibstyle[26]	BibTex[27]	Endnote	Mendeley	Reference Manager	Zotero	Any reference manager software	Not identifiable
Agricultural and Biological Sciences	20%	0%	0%	80%	20%	20%	20%	0%	0%
Arts and Humanities	0%	0%	0%	100%	0%	0%	0%	0%	0%
Biochemistry, Genetics and Molecular Biology	0%	0%	0%	100%	50%	0%	50%	0%	0%
Business, Management and Accounting	50%	0%	0%	50%	50%	0%	0%	0%	0%
Chemical Engineering	0%	0%	0%	100%	0%	0%	0%	0%	0%
Chemistry	33%	0%	0%	67%	33%	0%	0%	0%	0%
Computer Science	33%	0%	33%	33%	33%	0%	0%	0%	0%
Decision Sciences	50%	0%	0%	50%	50%	0%	0%	0%	0%
Dentistry	0%	0%	0%	100%	0%	0%	0%	0%	0%
Earth and Planetary Sciences	25%	0%	25%	25%	25%	0%	0%	0%	0%
Economics, Econometrics and Finance	50%	0%	0%	0%	50%	0%	0%	50%	0%
Energy	100%	0%	0%	0%	100%	0%	0%	0%	0%
Engineering	0%	0%	0%	50%	50%	0%	0%	0%	0%
Environmental Science	33%	0%	0%	67%	33%	0%	0%	0%	0%
Health Professions	0%	0%	0%	100%	0%	0%	0%	0%	0%
Immunology and Microbiology	50%	0%	0%	0%	50%	0%	0%	0%	0%
Materials Science	50%	0%	0%	0%	50%	0%	0%	0%	50%
Mathematics	33%	0%	33%	33%	33%	0%	0%	0%	0%
Medicine	0%	0%	0%	90%	27%	0%	18%	0%	0%
Multidisciplinary	0%	0%	0%	0%	0%	0%	0%	0%	0%
Neuroscience	0%	100%	0%	100%	0%	0%	0%	0%	0%
Nursing	0%	0%	0%	100%	0%	0%	0%	0%	0%
Pharmacology, Toxicology and Pharmaceutics	50%	0%	0%	50%	0%	50%	50%	0%	0%
Physics and Astronomy	0%	0%	0%	0%	0%	0%	0%	0%	0%
Psychology	50%	0%	0%	50%	50%	0%	0%	0%	0%
Social Sciences	0%	0%	0%	100%	0%	0%	20%	0%	0%
Veterinary	50%	0%	0%	50%	50%	0%	0%	0%	0%

[26] Bibstyle is the bibliography style referring to BibTex.

[27] The word, "BibTeX'' stands for a tool and a file format which are used to describe, and process lists of references, mostly in conjunction with LaTeX documents (BibTex.org, c2006). And LaTeX is a marking system or program for publishing high quality typographic documents, specific for the elaboration of scientific texts.

Journals also can provide tools for exporting citation metadata. These tools provide bibliographic references of the published articles, formatted according to specific reference styles or reference standards guidelines (e.g., Chicago and ABNT), simple text, HTML or MS Word file (with no specification of the reference style in which is formatted), machine-readable files (i.e., Endnote or Zotero files) or interchangeable files formats (e.g., CSV and RIS, which are usually readable by reference managers like Endnote, Mendeley and Zotero). It should be mentioned that publishers refer to structured formats (e.g., machine-readable, like RIS) and unstructured formats (e.g., plain text) as being the same things. Chart 12 highlights the most provided formats in which bibliographic references are made available.

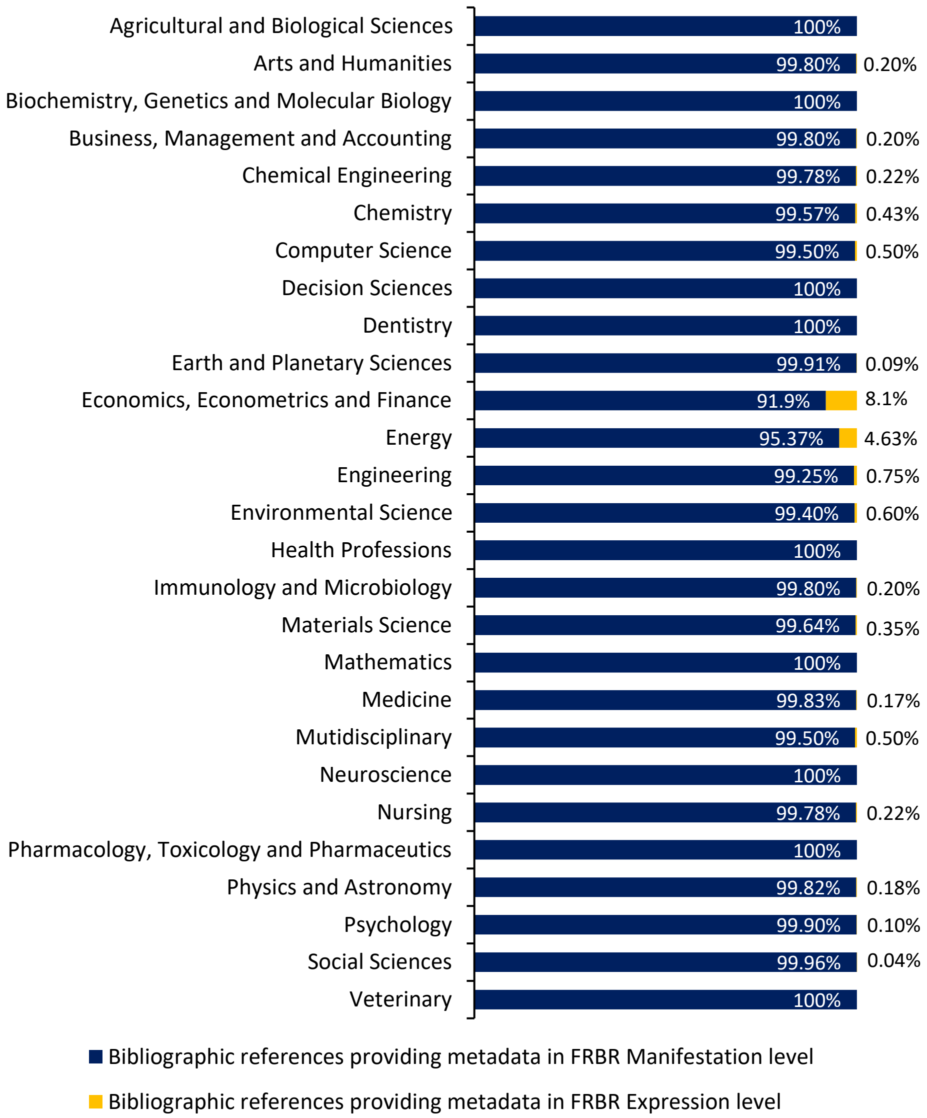
Chart 12 - Exportation files and files formats adopted by publishers for exporting articles' bibliographic metadata

Kind of format	Reference styles or formats provided to export bibliographic references	Percentual average of journals per subject area
Structured format (machine-readable)	BibTex	69.9%
	RIS	68%
	Refworks	56%
	Endnote	41%
	Mendeley	28.6%
	Medlars	24.2%
	Medline / Pubmed	4.9%
Unstructured format (plain text)	Text (a.k.a. simple text or plain text) – refer to journals providing bibliographic references with no specification of the reference system in which it is formatted	60.3%
	Chicago	4%

In Graphic 23, we show how the metadata addressed within bibliographic references comply to FRBR, assuming it as one of the core elements supporting the descriptive representation review. We established a correspondence between the level of description assumed by each bibliographic reference and the FRBR entities concepts, i.e., FRBR Work, Expression, Manifestation, and Item. We associated the corresponding FRBR entity to each bibliographic reference, according to the level of the description provided by the metadata set composing it. The results showed that the metadata set provided by 99.35% of bibliographic references, on average, corresponds to the FRBR Manifestation level. The metadata set considered by the remaining average portion of 0.65% of the bibliographic references corresponds to the

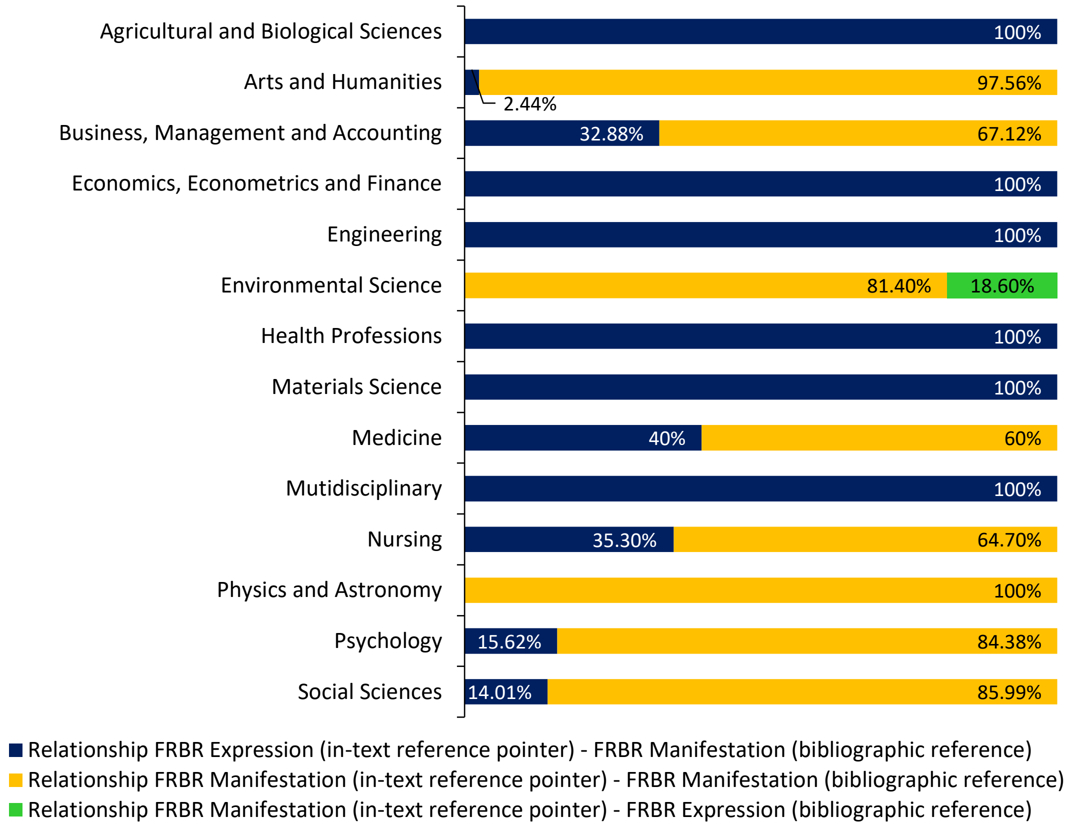
FRBR Expression level of description. No bibliographic references corresponding to the FRBR Work and FRBR Item levels of description were identified within our sample.

Graphic 23. Percentual distribution of bibliographic references per subject area, considering metadata provided by them and the correspondence with FRBR entities concept



Along the same line, the metadata set composing in-text reference pointers referring to quotations were analyzed using FRBR and comparing the FRBR Entities associated to such in-text reference pointer with the FRBR entity associated to the denoted bibliographic references. Considering articles with quotations, we noticed a slight predominance (53% of all the in-text reference pointer – bibliographic reference pairs) in the relation of FRBR Expression (for in-text reference pointers) with FRBR Manifestation (for the denoted bibliographic reference). Detailed data on this matter is shown on Graphic 24. Articles from Biochemistry, Chemical Engineering, Chemistry, Computer Science, Decision Science, Dentistry, Earth and Planetary Sciences, Energy, Immunology, Mathematics, Neuroscience, Pharmacology, and Veterinary subject areas were not introduced in Graphics 23 and 24 since we did not detect any quotations within those text bodies.

Graphic 24. Percentual distribution of in-text reference pointers accompanying quotations and their relationship with the denoted bibliographic references from FRBRized point of view[28]



Mentions and quotations are the core elements which materializes the citation network. From this perspective, the analysis considered the effectiveness of data provided by in-text reference pointers referring to mentions and quotations and

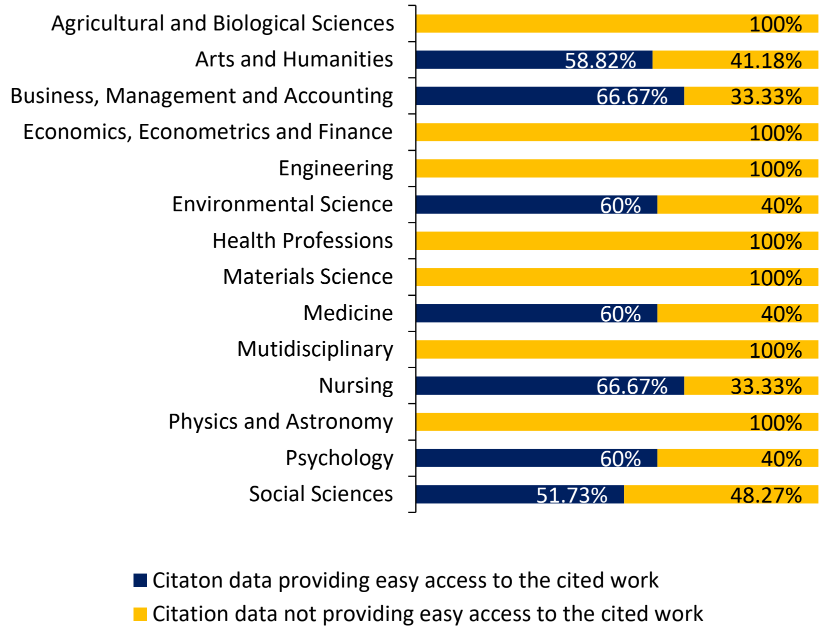
[28] (Biochemistry, Chemical Engineering, Chemistry, Computer Science, Decision Science, Dentistry, Earth and Planetary Sciences, Energy, Immunology, Mathematics, Neuroscience, Pharmacology, Veterinary not considered since we did not identify quotations within articles from such disciplines)

respective bibliographic references. The criteria used in this particular analysis considered, basically, the following functionalities:

a) whether the relationship between in-text reference pointers referring to mentions and quotations and bibliographic references are clear, i.e., whether the main access point considered in the bibliographic references corresponded to the data provided by the in-text reference pointers referring to them;
b) whether the descriptive elements provided by bibliographic references allow the identification of the cited work without consulting to external sources and;
c) whether the URL provided by bibliographic references, in applicable cases, were reliable, i.e., whether they pointed to the cited work itself or to the webpage from where it could be downloaded.

Those three strands represent the basis that make the citation network (more) effective since more than a concept corresponding to the abstract connection between two or more works, citation network is represented and materialized by the links established among different works by mentions, quotations and bibliographic references. Failures in such connections also may interfer in the improper reference to the cited work and, consequently, reflect negatively in the citation network. The results of the analysis considering items a, b and c mentioned above are approached in Graphic 25 which approaches the percentual of articles per subject area in which such analysis identified at least one non-conformity concerning the previously introduced aspects, that is, those providing an easy identification to the cited work. In this context, “easy identification” was assumed as the set of descriptive elements addressed by the in-text reference pointers and bibliographic references sets properly supporting the precise and unmistakable identification of the work containing the quoted passage and their location within the cited work, for the cases in which it was necessary, without the need to proceed complementary seeking in indexes or summaries, or to scroll down an electronic content, or leaf through a printed publication to get to the exactly cited passage. We noticed that 70% of articles having quotations, on average, do not provide easy identification of the external cited contents. Disciplines in which citation metadata were considered totally in conformity to the evaluated aspects were disregarded in Graphic 25.

Graphic 25. Percentual distribution of articles according to the facility of accessing the cited work, considering data provided by in-text reference pointers regarding quotations and bibliographic references sets (considering only disciplines with articles with quotations)



Surely, data shown in Graphic 25 has a subjective characteristic and indexes could show different, had other parameters been considered in the analysis. However, such data suggests that there is a potential research field towards the improvement of the way that citation data are presented in scientific publications, particularly concerning scientific articles.

8.3.Part III – An overview on the most commonly used descriptive elements in bibliographic references

In the first stage of the analysis, we considered the 34140 bibliographic references composing our sample, among with we identified the representation of 36 different types of publications, from with which articles, books and proceedings were, respectively, the first, second and third most cited types of publications across subject areas overall. From the subject category perspective, seven types of publications corresponded to 0.50% or more of the total of the bibliographic references, namely articles (83.55%), books (7.93%), proceedings (2.53%), webpages (1.30%), technical reports (1.17%), working papers/preprints (0.67%) and, conference papers (0.51%).

However, in such portion of the sample with the most cited types of publications from the subject areas' perspective, we identified that such approach did not comprise some types of publications considerably cited by specific disciplines. For instance, grey literature was the eighth most cited type of work across disciplines corresponding an overall percentual rate of 0.47% of total bibliographic references. On the other hand, such type of publication was the third most cited among arts and humanities articles and the fourth most cited among Chemical Engineering, Decision Sciences and Mathematics articles. That is to say that, considering only the overall most cited types of publications from the subject category perspective, in fact, do not proper represent the real citing habits scenery across subject areas. Because of this, the analysis comprised the seven most cited types of publications in each particular subject area, to assure that the coverage of the analysis includes the most cited types of publications also from the subject area's perspective.

8.3.1.Part III – An overview on the most cited types of publications across disciplines

The Chart 13 shows the percentual rates of cited works, per type of publication, observed within the articles of our sample. Each row in Chart 13 represents the cited types of publications per subject area, organized in a decrescent order from the most to the less cited types of publications. Blue cells represent the most cited types of publications across subject categories overall. Yellow cells represent those types of publications comprised among the seven most cited ones in particular subject areas which were not included in the overall subject categories approach (i.e., the citing rates for those specific types of publications did not reach the minimum levels to be classified among the seven most cited types of publications across disciplines overall. In contrast, they were observed among the most cited types of publications within the discipline to which the row where it is located refers to). Red cells represent bibliographic references with which the descriptive elements provided were not enough to identify the type of publication being referenced. White cells represent the less representative types of publications per subject area from citing rates perspective and which, therefore, were not included in this analysis.

Chart 13 - The most used types of publications cited per subject area[29]

Subject Areas	Types of publications being cited by articles under each subject area																		
Agricultural and Biological Sciences	Articles 87.51%	Books 7.88%	Proc. 1.30%	Web pages 0.81%	Technical reports 0.72%	Grey literature 0.40%	Standards 0.40%	Non Identified 0.31%	Online databases 0.31%	Manuals, guides and toolkits 0.18%	Working papers 0.04%	Conf. papers 0.04%	Softwares and Appl. 0.04%	Informative materials 0.04%					
Arts and Humanities	Articles 45.83%	Books 45.38%	Grey literature 1.66%	Web pages 1.48%	Proc. 1.20%	Non Identified 1.20%	Not published Docs. 0.86%	Forthcom-ing articles 0.46%	E-books 0.34%	Conf. papers 0.23%	Audio records 0.23%	Online databases 0.17%	Technical reports 0.17%	Web videos 0.17%	Forthcom-ing book chapters 0.17%	Motion pictures 0.11%	Ms. 0.11%	Speeches 0.11%	Manuals, guides and toolkits and Softwares and Appl. 0.12%[30]
Biochem., Genetics and Molecular Biology	Articles 98.52%	Books 0.94%	Web pages 0.27%	Proc. 0.13%	Technical reports 0.04%	Manuals, guides and toolkits 0.04%	Working papers 0.04%												
Business, Mngmt. and Accounting	Articles 76.74%	Books 13.52%	Working papers 2.56%	Web pages 2.38%	News. and magazines 1.95%	Technical reports 1.34%	Proc. 0.49%	Conf. papers 0.49%	Non Identified 0.18%	Grey literature 0.12%	Personal Comms. 0.12%	Web videos 0.06%	Ms. 0.06%						
Chemical Eng.	Articles 92.92%	Proc. 3.54%	Books 1.11%	Grey literature 1.11%	Non Identified 0.66%	Conf. papers 0.44%	Book series 0.22%												

[29] Appl. means Applications / Biochem. means Biochemistry / Comms. means communications / Conf. means conference / Docs. means Documents / Econ. means Econometrics / Eng. means Engineering / Environ. means environmental / Govt. means Governmental / Mngmt. means management / Ms. means manuscripts / Mutidiscip. means multidisciplinary / News. means newspapers / Proc. means proceedings / Memo means memorandum / Microbiol. means microbiology / Neurosci. means Neuroscience / Pharmaceut. means Pharmaceutical / Pharmacol. means Pharmacology / Pho. means photographs / Presen. means presentation.

[30] Includes Manuals, guides and toolkits and Softwares and applications, each one representing 0.09% of the types of publications cited by articles composing Arts and Humanities subject area.

Subject Areas	Types of publications being cited by articles under each subject area												
Chemistry	Articles 96.91%	Books 1.77%	Non Identified 0.59%	Web pages 0.59%	Grey literature 0.29%	Book series 0.15%	Technical reports 0.15%						
Computer Science	Articles 58.76%	Proc. 18.31%	Conf. papers 8.53%	Books 6.93%	Web pages 2.93%	Non Identified 1.87%	Working papers 0.98%	Data sheets 0.62%	Grey literature 0.36%	Technical reports 0.18%	Not published Docs. 0.18%	Forthcom-ing articles 0.09%	Online databases 0.09%
Decision Sciences	Articles 58.18%	Proc. 25.93%	Books 10.51%	Grey literature 1.17%	Technical reports 1.17%	Web pages 0.93%	Non Identified 0.70%	Conf. papers 0.47%	Working papers 0.47%	Online databases 0.47%			
Dentistry	Articles 98.95%	Books 0.79%	Forthcom - ing articles 0.26%										
Earth and Planetary Sciences	Articles 89.71%	Books 4.15%	Technical reports 3.11%	Conf. papers 0.78%	Proc. 0.52%	Grey literature 0.35%	Web pages 0.35%	Softwares and Appl. 0.35%	Online databases 0.26%	Working papers 0.17%	Forthcom-ing articles 0.09%	Non Identified 0.09%	Manuals, guides and toolkits 0.09%
Economics, Econ. and Finance	Articles 74.84%	Non Identified 8.10%	Books 6.13%	Working papers 4.38%	Technical reports 2.19%	Forthcom - ing articles 1.53%	Proc. 1.31%	Web pages 0.66%	Non idenfified 0.66%	Conf. papers 0.22%			
Energy	Articles 70.50%	Books 5.83%	Technical reports 5.66%	Web pages 5.32%	Non Identified 5.15%	Proc. 4.97%	Grey literature 1.54%	Online databases 0.69%	Working papers 0.17%	Conf. papers 0.17%			

Subject Areas	Types of publications being cited by articles under each subject area																
Engineering	Articles 76.87%	Books 6.67%	Proc. 6.39%	Web pages 2.93%	Technical reports 2.11%	Non Identified 1.63%	Conf. papers 1.09%	Grey literature 0.95%	Working papers 0.68%	Standards 0.27%	E-books 0.14%	Patents 0.14%	Forthcom-ing articles 0.07%	Data sheets 0.07%			
Environ. Science	Articles 76.66%	Books 12.54%	Technical reports 4.01%	Web pages 3.30%	Non Identified 0.90%	Proc. 0.65%	Working papers 0.52%	Conf. papers 0.26%	Grey literature 0.26%	E-books 0.19%	Softwares and Appl. 0.19%	Online databases 0.13%	Data sheets 0.06%	Not published Docs. 0.06%	Manuals, guides and toolkits 0.06%	Web videos 0.06%	Forthcom - ing book chapters 0.06%	Slides presen. 0.06%
Health Professions	Articles 92.99%	Books 6.27%	Web pages 0.37%	Manuals, guides and toolkits 0.37%													
Immunology and Microbiol.	Articles 94.82%	Books 4.80%	Non Identified 0.19%	Forthcom-ing articles 0.19%													
Materials Science	Articles 92.31%	Books 3.91%	Non Identified 0.83%	Web pages 0.71%	Proc. 0.71%	Working papers 0.47%	Conf. papers 0.36%	Grey literature 0.36%	Standards 0.24%	Technical reports 0.12%							
Mathematics	Articles 89.10%	Books 9.31%	Proc. 0.53%	Grey literature 0.53%	Web pages 0.18%	Technical reports 0.18%	Forthcom-ing articles 0.18%										
Medicine	Articles 90.06%	Technical reports 3.22%	Web pages 2.77%	Books 2.00%	E-books 0.43%	Forthcom - ing articles 0.24%	Non Identified 0.22%	Online databases 0.17%	Proc. 0.15%	Informative materials 0.15%	Softwares and Appl. 0.11%	Manuals, guides and toolkits 0.09%	Legislation 0.07%	Conf. papers 0.04%	Govt. Official Publications 0.04%	Grey literature 0.02%	Patents 0.02%	Memo. 0.02%
Mutidiscip.	Articles 94.08%	Books 3.01%	Patents 1.20%	Non Identified 0.50%	Technical reports 0.40%	Web pages 0.40%	Proc. 0.20%	Working papers 0.20%									

Subject Areas	Types of publications being cited by articles under each subject area																		
Neurosci.	Articles 96.04%	Books 1.65%	Web pages 0.99%	Online databases 0.66%	Forthcom-ing articles 0.33%	Working papers 0.17%	Softwares and Appl. 0.17%												
Nursing	Articles 81.14%	Books 12.05%	Technical reports 2.50%	Web pages 1.36%	Non Identified 0.91%	Grey literature 0.91%	Book series 0.68%	Forthcom-ing articles 0.23%	Web videos 0.23%										
Pharmacol., Toxicology and Pharmaceut.	Articles 98.54%	Books 0.73%	Proc. 0.24%	Technical reports 0.12%	Grey literature 0.12%	Online database s 0.12%	Patents 0.12%												
Physics and Astronomy	Articles 88.56%	Working papers 6.18%	Books 2.82%	Proc. 0.91%	Grey literature 0.45%	Technical reports 0.36%	Web pages 0.36%	Non Identified 0.18%	Forthcom-ing articles 0.09%	Softwares and Appl. 0.09%									
Psychology	Articles 85.03%	Books 11.52%	Technical reports 0.78%	Web pages 0.78%	Grey literature 0.65%	Working papers 0.26%	Softwares and Appl. 0.20%	Proc. 0.13%	Forthcom-ing articles 0.13%	Not published Docs. 0.13%	News. and magazines 0.13%	Non Identified 0.07%	Manuals, guides and toolkits 0.07%	Conf. papers 0.07%	Personal Comms. 0.07%				
Social Sciences	Articles 57.73%	Books 27.46%	Web pages 4.31%	Technical reports 3.13%	Non Identified 1.18%	E-books 0.85%	Grey literature 0.77%	Conf. papers 0.61%	News. and magazines 0.55%	Online databases 0.42%	Proc. 0.37%	Motion pictures 0.35%	Working papers 0.31%	Ms. 0.20%	Web videos 0.18%	Audio records 0.11%	Not published Docs., and Manuals, guides and toolkits[31] 0.18%	Pho. 0.07%	Other types of publications[32] 0.22%
Veterinary	Articles 92.56%	Books 4.46%	Web pages 0.89%	Grey literature 0.60%	Working papers 0.60%	Personal Comms. 0.60%	Proc. 0.30%												

Chart 13 evinces, primarily, two things: first, that articles and books are the most cited types of publications for most disciplines, although some of them like Social Sciences subject category, tend to cite a greater variety of types of publications in contrast to other subject areas, like Dentistry for example. Second, the types of publications supporting discussions across disciplines may vary.

By analyzing data showed in Chart 13, we identified the most cited types of works, considering the perspective of all the 27 subject areas involved in this study, namely: articles, book, proceedings, webpages, technical reports, working papers/preprints and, conference papers, representing the most cited types of publications overall and, grey literature, standards, unpublished documents, forthcoming articles, e-books, manuals, guides and toolkits, newspapers and magazines, book series, data sheets, softwares and applications, online databases, patents and, personal communications which, although not included in the most cited publications overall from subject areas perspective (blue cells in Chart 13), were considerably cited and, therefore, understood as relevant publication types for specific disciplines, such as the case of Arts and Humanities, in which Grey literature was the third most cited type of publication, while from the overall subject areas perspective, the same type of publication was the ninth more cited.

8.3.2.Part III – An overview on the bibliographic elements most used for bibliographic references composition: “the starred metadata set”

The 33786 bibliographic references concerning the most significative types of publications from the perspective of subject areas composing our sample (blue and yellow cells in Chart 13), were individually analyzed to identify the descriptive elements composing bibliographic references of the most cited types of publications, whose results were compiled in Chart 14. As stated in the Chart 13, the blue cells in the first column of Chart 14 stand for the most cited types of works across subject categories overall. As for yellow cells stand for the types of publications not observed among the most cited types of publications overall but that meanwhile were considered relevant for particular disciplines.

The percentual indexes showed in the rows of Chart 14, starting from column C, Row 2, represent the rate of articles per subject category in which at least one bibliographic reference in the bibliographic references list provides the descriptive element to with the respective column refers. By providing a descriptive element, i.e., the work's title in English or the ISBN number, for example, we assumed that authors demonstrate an understanding concerning the importance of providing such data in bibliographic references and their role in the identification of the cited works. Under the same view, although the uniformity in the descriptive elements composing each bibliographic reference in a bibliographic references list is the ideal scenery from standardization perspective, we noticed that this is not what happens in all cases. For instance, we observed bibliographic references list containing two or more bibliographic references referring to the same type of publication, in which one provides a specific bibliographic element (e.g., the ISBN), and the others, do not. In such cases, we assumed that by providing the bibliographic element in one bibliographic reference element, the author demonstrates the understanding that such metadata is relevant for the identification of the cited work and that the non-provision of this in the remaining bibliographic references was due to the inexistence or the unavailability of such metadata, instead of as an act of fault or negligence on the part of the authors.

The columns in Chart 14 represent each single descriptive element identified within analyzed bibliographic references, according to the RDA guidelines concerning the core elements for describing an information resource[33]: the title, the statement of responsibility, the edition statement, the numbering of serials, the production statement, the publication statement, the series statement, the identifier for manifestation, the carrier type, and the extent.

The RDA core elements for describing resources were selected according to the FRBR assessment of the value of each attribute and relationship in supporting the following user tasks[34]:

- identify and select a manifestation (FRBR);

[33] Based on the rule number 0.6.5 of RDA. Available from: https://original.rdatoolkit.org/.

[34] The user tasks considered in this study were originally presented by the three entity-relationship models composing the FRBR Family: The Functional Requirements for Bibliographic Records (FRBR), the Functional Requirements for Authority Data (FRAD) and the Functional Requirements for Subject Authority Data (FRSAD).

- identify works and expressions embodied in a manifestation (FRBR);
- identify the creator or creators of a work (FRBR);
- find an agent associated with a resource (FRAD);
- identify an agent (FRAD);
- find one or more subjects and/or their appellations associated with a work (FRSAD);
- identify a subject and/or its appellation (FRSAD) and;
- explore relationships between subjects and/or their appellations (FRSAD).

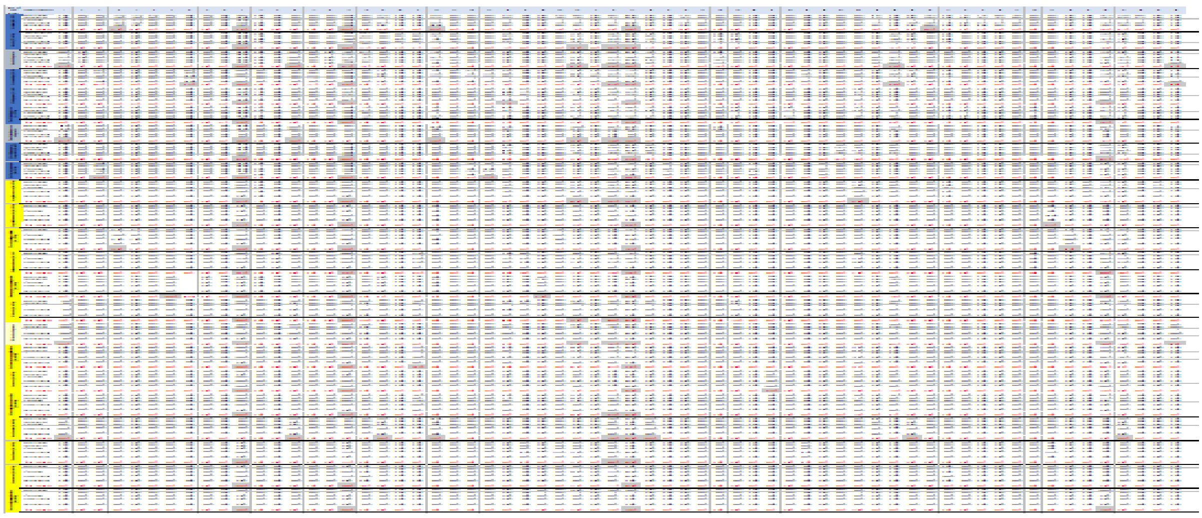
Those bibliographic elements found in the bibliographic references that did not fit to any of RDA Core elements were categorized as general notes, online availability notes or, miscellaneous, as the case may be.

Grey rows in Chart 14 represent the subject categories in which we did not observed any mention or quotation referring to the type of work corresponding to that row (yellow and blue cells in the first column of Chart 14).

The percentual values showed in Chart 14 provide the overall rates of articles per subject category in which we observed at least one bibliographic reference providing the descriptive element referred in the respective column of the chart. Such data evinces that the rate of provision of descriptive elements within bibliographic references may vary not only among different subject categories overall but also among disciplines composing the same subject category. For instance, bibliographic references referring to book chapters not always provide the title of the cited chapter. Although such discrepancies not necessarily configure major problems from the perspective of the identification of the cited work, it reveals a lack of standardization in the way in which such data are provided in bibliographic references considering articles from different disciplines.

The exam of the bibliographic references of our sample showed that a portion of them, particularly speaking of books, e-books, and technical reports, clearly referred to specific parts of the cited works i.e., chapters. To provide a more accurate analysis of the descriptive metadata composing those bibliographic references, we considered such particularities individually in Chart 14 and such data can be seen in the rows identified in light yellow and light blue in the first column. Including, for such cases, indexes referring to the percentual rate of articles with bibliographic references providing the title of the chapter title are presented in the column 1 of Chart 14, which refers to "chapter title", as for the indexes referring to the percentual rate of articles with bibliographic references providing the title of the work containing the cited chapter is provided in the column 11 which refers to the "work's title in the original language".

Chart 14 - Percentual average distribution of articles, according to the descriptive elements provided by its bibliographic references, by type of cited works and by subject category: The starred metadata set (click here or in the image below to see the full chart[35])



Legend for Chart 14

Title	1 Chapter title 2 Chapter title in English (when original title is in another language) 3 Conferences' title 4 Journals' title (abridged format) 5 Journals's title (full format) 6 Journals's title in English (for titles in other languages) 7 Newspaper/magazine title 8 Proceedings' title 9 Session title 10 Works' subtitle in original language 11 Works' title in original language 12 Works's title in English (when original title is in another language)
Statement of responsibility	13 Author full name 14 Chapter author 15 Proceedings' editor

[35] As it contains an expressive volume of data, the dimensions of Chart 14 made it impossible for it to be included directly in the text body. The full picture is available at https://zenodo.org/record/4759403/files/Chart%2014%20-%20Starred%20metadata%20set.xlsx?download=1

	16 Translator 17 Work's author or editor
Edition statement	18 Edition number 19 Issue number 20 Revision number 21 Version number 22 Volume number
Numbering of serials	23 Series number
Production statement	No occurrencies
Publication statement	24 Conference date 25 Conference place 26 Date of citation (date of access) 27 Date of last update / revision 28 Day of publication 29 Month of publication 30 Place of publication 31 Proceedings date of publication 32 Publisher (or granting institutions for thesis and dissertations) 33 Year or date of publication
Series statement	34 Series title
Identifier for manifestation	35 Abstract number 36 Article ID within publisher's webpage 37 Article number part note 38 Chapter number 39 ISBN number 40 Paper number 41 Patent number 42 Technical report number 43 Work number 44 Work number within the conference 45 Working paper number
Carrier type	46 Content type / media type / carrier type (general material designation in AACR2)
Extent	47 Abridged work pagination lenght (e.g. 80-9) 48 Cited work pages range (e.g. 80-89) 49 Work's first page number (e.g. 80) 50 Work full pagination length (e.g. 80-89) 51 Work's total number of pages (e.g. 80 p.)
General notes	52 Work's language note 53 Supplemental issue note 54 Special issue note 55 Supplementary content note 56 General notes 57 Unpublished note[36] 58 In press note 59 Database system number
Online availability notes	60 Hypertext hyperlink (URL) 61 DOI string or DOI URL 62 Online availability note[37] 63 Institutional link or university department
Miscelaneous	64 Latin expression "in" (i.e., for book chapters or conference papers in a proceedings)

[37] This bibliographic element refers to some bibliographic references providing a short statement clarifying that the cited work is available online, but do not provide the URL from where it can be retrieved. All the raw data for this research is available at Santos (2021).

Basically, Charts 13 and 14 together evince that there is no strict uniformity on the descriptive elements considered in bibliographic references referring to the same types of publications across subject areas and categories. And this suggests that descriptive elements have different importance levels across disciplines, according to the type of publication being described. Bibliographic references referring to proceedings were the ones in which we observed the most considerable variation among the metadata set composing each bibliographic reference. For instance, bibliographic references addressing proceedings, in fact, refer to works (like conference papers, speeches, resumes and poster' exhibitions) usually presented in a scientific event, like congresses and symposiums. However, in 10% of articles of our sample we observed bibliographic references addressing proceedings, in which the title of the work effectively cited within the text body was not provided. On the other hand, we observed bibliographic references clearly addressing the conference paper itself and, in such cases, we noticed that the conference title was not provided in the bibliographic reference in 9.72% of Social Sciences articles and 2.14% of Physical Sciences articles.

Although not being their main purpose, bibliographic references also can act like sources of information and, from this perspective, the efforts on providing (at least) the necessary metadata for the proper identification of the referred publications are both worthwhile and necessary as a means of indirectly subsidizing the location and retrieving of cited works, by providing the necessary metadata for bibliographic seeking in external sources, like bibliographic catalogs and bibliographic databases. However, we noticed that such kind of metadata are not always provided. For instance, in 27% of articles from Physical Sciences, we noticed bibliographic references addressing webpages, whose title was not provided. Although such bibliographic references usually provide a hypertext hyperlink or other path for electronic access for the cited publication, i.e., a DOI string or DOI URL, and although such tools are usually free of charge, they should not be considered as substitutes for titles themselves.

Other point evinced by Chart 14 is that different types of publication have different and characteristics from diverse natures. For instance, articles metadata usually includes, among other data, the title of the journal in which they were published, with the respective volume and/or issue numbers, whereas conference papers metadata usually includes at least both the paper and the conference title. Likewise, such issues are also valid considering other types of publications. Therefore, the description of different types of publications may demand different types of metadata, which not necessarily play the same role on the identification of the cited work and, because of that, may have different levels of importance in terms of facilitating the task of identifying the cited work and, therefore, such issues should be considered by metadata treatment tools, like the ontologies.

Chart 14 also evinced that part of the bibliographic references providing means of electronic access to the cited works, i.e., hypertext hyperlinks, do not provide the data in which the content available in the cited online source was consulted. This may represent issues on later retrieving of such content because unlike press sources of information, like books, whose content may not be modified after their release (at least considering works from the same edition), online sources are susceptible to amendments and might even become unavailable at any time, without prior notification. Because of this, including the date of consultation to the cited content in the bibliographic reference is a way to warn and inform the reader against the occasion in which the cited content was available and discharge the author in case of later changes to the cited content.

Despite the existence of thousands of reference styles and standards to guide the use and interpretation of bibliographic metadata in a uniform way, we observed that the representation of the information is approached differently across subject areas and categories. That is to say that the same type of publication may assume different levels of descriptions in different disciplines. This suggests failures in the reference styles' purposes concerning their role on standardizing bibliographic references in large scale. Such variation is shown in Chart 15 - Determination of the most usually used metadata in bibliographic references per subject category and type of publication. Data on such table represent a micro view from data on Chart 14.

In fact, data shown in Chart 15 represent the bibliographic elements considered by at least one bibliographic reference included in the bibliographic reference list in 50% or more of the articles composing the article's sample for each subject category. That is to say that Chart 15 consider the most used descriptive elements within bibliographic references which will be called “starred metadata set” from now on and are explicitly represented in the blue cells. Technically speaking, Chart 13 partially support the answer to the research question number 8 (RQ 8).

Chart 15 - Determination of the most usually used metadata in bibliographic references per subject category and type of publication: The "starred metadata set"

Type of cited publication	Subject categories	Correspondent column number in Chart 14	Descriptive elements composing bibliographic references
Articles	Health Sciences	4, 11, 17, 22, 33, 50	Journal's' title (abridged format), works' title in original language, work's author or editor, volume number, year or date of publication, work full pagination length (e.g., 80-89)
	Social Sciences	5, 11, 17, 19, 22, 33, 50	Journal's title (full format), works' title in original language, work's author or editor, issue number, volume number, year or date of publication, work full pagination length (e.g., 80-89)
	Life Sciences	4, 11, 17, 22, 33, 36, 50	Journal's' title (abridged format), works' title in original language, work's author or editor, volume number, year or date of publication, article ID within publisher's webpage, work full pagination length (e.g., 80-89)
	Physical Sciences	4, 11, 17, 22, 33, 50	Journal's' title (abridged format), works' title in original language, work's author or editor, volume number, year or date of publication, work full pagination length (e.g., 80-89)
	Multidisciplinary	4, 11, 17, 22, 33, 36, 50	Journal's' title (abridged format), works' title in original language, work's author or editor, volume number, year or date of publication, article ID within publisher's webpage, work full pagination length (e.g., 80-89)
	Average	**11, 17, 22, 33, 50**	**works' title in original language, work's author or editor, volume number, year or date of publication, work full pagination length (e.g. 80-89)**
Books	Health Sciences	11, 17, 18, 30, 32, 33	Works' title in original language, work's author or editor, edition number, place of publication, publisher, year or date of publication
	Social Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Life Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Physical Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Multidisciplinary	11, 17, 32, 33	Works' title in original language, work's author or editor, publisher, year or date of publication
	Average	**11, 17, 30, 32, 33,**	**Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication**

Book chapters	Health Sciences	1, 11, 14, 17, 30, 32, 33, 48, 64	Chapter title, works' title in original language, chapter´s author, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Social Sciences	1, 11, 14, 17, 30, 32, 33, 48, 64	Chapter title, works' title in original language, chapter´s author, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Life Sciences	1, 11, 14, 17, 30, 32, 33, 48, 64	Chapter title, works' title in original language, chapter´s author, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Physical Sciences	1, 11, 14, 17, 30, 32, 33, 48, 64	Chapter title, works' title in original language chapter´s author, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Multidisciplinary	1, 11, 14, 17, 30, 32, 33, 38, 48, 64	Chapter title, works' title in original language, chapter´s author, work's author or editor, place of publication, publisher, year or date of publication, chapter number, cited work pages range (e.g., 80-89), Latin expression "in"
	Average	**1, 11, 14, 17, 30, 32, 33, 48, 64**	**Chapter title, works' title in original language, chapter's author, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"**
Proceedings	Health Sciences	3, 11, 17, 33, 48, 64	Conference's title, works' title in original language, work's author or editor, year or date of publication, cited work pages range (e.g. 80-89), Latin expression "in"
	Social Sciences	8, 11, 17, 30, 32, 33, 48, 64	Proceeding's title, works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Life Sciences	8, 11, 17, 32, 33, 48, 64	Proceeding's title, works' title in original language, work's author or editor, publisher, year or date of publication, cited work pages range (e.g. 80-89), Latin expression "in"
	Physical Sciences	8, 11, 17, 32, 33, 48, 64	Proceeding's title, works' title in original language, work's author or editor, publisher, year or date of publication, cited work pages range (e.g. 80-89), Latin expression "in"
	Multidisciplinary	3, 8, 11, 17, 32, 33, 48, 64	Conference's title, proceeding's title, works' title in original language, work's author or editor, publisher, year or date of publication, cited work pages range (e.g., 80-89), Latin expression "in"
	Average	**11, 17, 33, 48, 64**	**Works' title in original language, work's author or editor, year or date of publication, cited work pages range (e.g. 80-89), Latin expression "in"**
Webpages	Health Sciences	11, 17, 26, 60	Works' title in original language, work's author or editor, date of citation (date of access) hypertext hyperlink (URL)

	Social Sciences	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Life Sciences	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Physical Sciences	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Multidisciplinary	11, 17, 60	Works' title in original language, work's author or editor, hypertext hyperlink (URL)
	Average	**11, 17, 60**	**Works' title in original language, work's author or editor, hypertext hyperlink (URL)**
Technical reports	Health Sciences	11, 17, 26, 33, 60	Works' title in original language, work's author or editor, date of citation (date of access), year or date of publication, hypertext hyperlink (URL)
	Social Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Life Sciences	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Physical Sciences	11, 17, 33	Works' title in original language, work's author or editor, year or date of publication
	Multidisciplinary	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**
Technical reports chapters	Health Sciences	1, 11, 14, 22, 30, 32, 33, 60	Chapter title, works' title in original language, chapter author, volume number, place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Social Sciences	1, 11, 14, 30, 32, 33	Chapter title, works' title in original language, chapter author, Place of publication, Publisher, Year or date of publication
	Life Sciences	No citations	No citations
	Physical Sciences	No citations	No citations
	Multidisciplinary	No citations	No citations
	Average	**1, 11, 14, 30, 32, 33**	**Chapter title, works' title in original language, chapter author, Place of publication, Publisher, Year or date of publication**
Working papers / preprints	Health Sciences	11, 17, 26, 30, 32, 33, 60	Works' title in original language, work's author or editor, date of citation (date of access), place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Social Sciences	11, 17, 33, 45, 60	Works' title in original language, work's author or editor, year or date of publication, working paper number, hypertext hyperlink (URL)

	Life Sciences	11, 17, 26, 33, 61	Works' title in original language, work's author or editor, date of citation (date of access), year or date of publication, DOI string or DOI URL
	Physical Sciences	11, 17, 33, 60	Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)
	Multidisciplinary	11, 17, 32, 33, 60	Works' title in original language, work's author or editor, publisher, year or date of publication, hypertext hyperlink (URL)
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**
Conference papers	Health Sciences	3, 11, 17, 33	Conference's title, works' title in original language, work's author or editor, year or date of publication
	Social Sciences	3, 11, 17, 25, 33	Conference's title, works' title in original language, work's author or editor, conference place, year or date of publication
	Life Sciences	3, 11, 17, 25, 33	Conference's title, works' title in original language, work's author or editor, conference place, year or date of publication
	Physical Sciences	3, 11, 17, 25, 33	Conference's title, works' title in original language, work's author or editor, conference place, year or date of publication
	Multidisciplinary	No citaitons	No citaitons
	Average	**3, 11, 17, 33**	**Conference's title, works' title in original language, work's author or editor, year or date of publication**
Grey literature[38]	Health Sciences	11, 17, 30, 32, 33, 46	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)
	Social Sciences	11, 17, 32, 33, 46	Works' title in original language, work's author or editor, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)
	Life Sciences	11, 17, 30, 32, 33, 46	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)
	Physical Sciences	11, 17, 30, 32, 33, 46	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 32, 33, 46**	**works' title in original language, work's author or editor, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)**
Unpublished works	Health Sciences	No citations	No citations
	Social Sciences	11, 17, 33, 57	Works' title in original language, work's author or editor, year or date of publication, unpublished note

[38] "Publisher" in this type of publication should be understood as the university in which the academic work was produced.

	Life Sciences	No citations	No citations
	Physical Sciences	11, 17, 32, 33, 57	Works' title in original language, work's author or editor, publisher, year or date of publication, unpublished note
	Multidisciplinary	No ciations	no ciations
	Average	**11, 17, 33, 57**	**Works' title in original language, work's author or editor, year or date of publication, unpublished note**
Forthcoming articles	Health Sciences	4, 11, 17, 33, 58, 61	Journal's' title (abridged format), works' title in original language, work's author or editor, year or date of publication, In press note, DOI string or DOI URL
	Social Sciences	5, 11, 17, 58	Journals's title (full format), works' title in original language, work's author or editor, In press note
	Life Sciences	4, 11, 17, 22, 29, 33, 58, 60	Journal's' title (abridged format), works' title in original language, work's author or editor, volume number, month of publication, year or date of publication, In press note, hypertext hyperlink (URL)
	Physical Sciences	4,11, 17, 33, 58	Journal's' title (abridged format), works' title in original language, work's author or editor, year or date of publication, In press note
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 58**	**Works' title in original language, work's author or editor, volume number, month of publication, In press note**
Online databases	Health Sciences	11, 17, 21, 26, 33, 60	Works' title in original language, work's author or editor, version number, date of citation (date of access), year or date of publication, hypertext hyperlink (URL)
	Social Sciences	11, 17, 32, 33, 60	Works' title in original language, work's author or editor, publisher, year or date of publication, hypertext hyperlink (URL),
	Life Sciences	11, 17, 21, 33	Works' title in original language, work's author or editor, version number, year or date of publication
	Physical Sciences	11, 17, 21, 32, 33, 46, 60, 61	Works' title in original language, work's author or editor, version number, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2), hypertext hyperlink (URL), DOI string or DOI URL
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**
Newspapers and magazines	Health Sciences	No citations	No citations
	Social Sciences	7, 11, 17, 28, 33, 60	Newspaper/magazine title, works' title in original language, work's author or editor, day of publication, year or date of publication, hypertext hyperlink (URL)
	Life Sciences	No citations	No citations
	Physical Sciences	No citations	No citations
	Multidisciplinary	No citations	No citations

	Average	**7, 11, 17, 28, 33, 60**	**Newspaper/magazine title, works' title in original language, work's author or editor, day of publication, year or date of publication, hypertext hyperlink (URL)**
E-books	Health Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Social Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Life Sciences	11, 17, 26, 30, 32, 33, 60	Works' title in original language, work's author or editor, date of citation (date of access), place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Physical Sciences	11, 17, 18, 26, 33, 39, 60, 61	Works' title in original language, work's author or editor, edition number, date of citation (date of access), year or date of publication, ISBN number, hypertext hyperlink (URL), DOI string or DOI URL
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**
E-books chapters	Health Sciences	1, 11, 14, 17, 30, 32, 33, 48	Chapter title, works' title in original language, chapter author, work's author or editor, place of publication, publisher, year or date of publication, cited work's page range (e.g., 80-89)
	Social Sciences	1, 11, 17, 30, 32, 33, 64	Chapter title, works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, Latin expression "in"
	Life Sciences	No citations	No citations
	Physical Sciences	1, 11, 14, 17, 26, 33, 60, 64	Chapter title, works' title in original language, chapter author, work's author or editor, date of citation (date of access), year or date of publication, hypertext hyperlink (URL), Latin expression "in"
	Multidisciplinary	No citations	No citations
	Average	**1, 11, 17, 33**	**Chapter title, works' title in original language, work's author or editor, year or date of publication**
Softwares and applications	Health Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Social Sciences	11, 17, 30, 32, 33, 46	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, content type / media type / carrier type (general material designation in AACR2)
	Life Sciences	11, 17, 21, 26, 30, 32, 33, 60	Works' title in original language, work's author or editor, version number, date of citation (date of access), place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Physical Sciences	11, 17, 21, 33, 60	Works' title in original language, work's author or editor, version number, year or date of publication, hypertext hyperlink (URL)
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**

Patents	Health Sciences	11, 17, 33, 41	Works' title in original language, work's author or editor, year or date of publication, patent number
	Social Sciences	No citations	No citations
	Life Sciences	11, 17, 33, 41, 46	Works' title in original language, work's author or editor, year or date of publication, patent number, content type / media type / carrier type (general material designation in AACR2)
	Physical Sciences	11, 17, 30, 33, 41, 48	Works' title in original language, work's author or editor, place of publication, year or date of publication, patent number, cited work pages range (e.g. 80-89)
	Multidisciplinary	11, 17, 33, 41, 60	Works' title in original language, work's author or editor, year or date of publication, patent number, hypertext hyperlink (URL)
	Average	**11, 17, 33, 41**	**Works' title in original language, work's author or editor, year or date of publication, patent number**
Manual, guides and toolkits	Health Sciences	11, 17, 30, 32, 33, 60	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Social Sciences	11, 17, 30, 32, 33	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication
	Life Sciences	11, 17, 32, 33	Works' title in original language, work's author or editor, publisher, year or date of publication
	Physical Sciences	11, 17, 21, 32, 33	Works' title in original language, work's author or editor, version number, publisher, year or date of publication
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 32, 33**	**Works' title in original language, work's author or editor, publisher, year or date of publication**
Book series	Health Sciences	1, 14, 19, 22, 32, 33, 34, 47, 49, 61	Chapter title, chapter author, issue number, volume number, publisher, year or date of publication, series title, abridged work pagination length (e.g. 80-9), work's first page number (e.g. 80), DOI string or DOI URL
	Social Sciences	No citations	No citations
	Life Sciences	No citations	No citations
	Physical Sciences	1, 14, 19, 22, 32, 33, 34, 49, 61	Chapter title, chapter author, issue number, volume number, publisher, year or date of publication, series title, work's first page number (e.g. 80), DOI string or DOI URL
	Multidisciplinary	No citations	No citations
	Average	**1, 14, 19, 22, 32, 33, 34, 49, 61**	**Chapter title, chapter's author, issue number, volume number, publisher, year or date of publication, series title, work's first page number (e.g.80), DOI string or DOI URL**
Data sheets	Health Sciences	No citations	No citations
	Social Sciences	No citations	No citations
	Life Sciences	No citations	No citations

	Physical Sciences	11, 32, 33	works' title in original language, publisher, year or date of publication
	Multidisciplinary	No citations	No citations
	Average	**11, 32, 33**	**Works' title in original language, publisher, year or date of publication**
Standards	Health Sciences	No citations	No citations
	Social Sciences	11, 17, 33	Works' title in original language, work's author or editor, year or date of publication
	Life Sciences	11, 17, 30, 33	Works' title in original language, work's author or editor, place of publication, year or date of publication
	Physical Sciences	11, 17, 18, 33, 51	Works' title in original language, work's author or editor, edition number, year or date of publication, work's total number of pages (e.g. 80 p.)
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 33**	**Works' title in original language, work's author or editor, year or date of publication**
Personal communications	Health Sciences	11, 17, 30, 32, 33, 60	Works' title in original language, work's author or editor, place of publication, publisher, year or date of publication, hypertext hyperlink (URL)
	Social Sciences	11, 17, 28, 33, 46, 60	Works' title in original language, work's author or editor, day of publication, year or date of publication, content type / media type / carrier type (general material designation in AACR2), hypertext hyperlink (URL)
	Life Sciences	No citations	No citations
	Physical Sciences	No citations	No citations
	Multidisciplinary	No citations	No citations
	Average	**11, 17, 33, 60**	**Works' title in original language, work's author or editor, year or date of publication, hypertext hyperlink (URL)**

8.3.3. Part III – What are the SPAR Ontologies?

The Semantic Publishing and Referencing Ontologies, a.k.a. SPAR Ontologies, form a suite of orthogonal and complementary OWL 2 DL ontology modules for the creation of comprehensive machine-readable RDF metadata for every aspect of semantic publishing and referencing: document description, bibliographic resource identifiers, types of citations and related contexts, bibliographic references, document parts and status, agents' roles and contributions, bibliometric data, and workflow processes. (PERONI; SHOTTON, c2021?).

In fact, the Semantic Publishing and Referencing (SPAR) are composed of four complementary and interoperable ontologies modules. Each module is composed of a set of ontologies[39]:

a) Ontologies for describing Bibiograpihc resources and their parts
   I. FaBiO (the FRBR-aligned Bibliographic Ontology)
   II. FRBR-DL (the Essential FRBR in OWL2 DL Ontology)
   III. DoCO (the Document Components Ontology)
   IV. DEO (the Discourse Elements Ontology)
   V. The DataCite Ontology
b) Ontologies for describing citations of scholarly resources
   I. CiTO (the Citation Typing Ontology)
   II. BiRO (the Bibliographic Reference Ontology)
   III. C4O (the Citation Counting and Context Characterisation Ontology)
c) Ontologies for describing the publishing workflow
   I. PRO (the Publishing Roles Ontology)
   II. PSO (the Publishing Status Ontology)
   III. PWO (the Publishing Workflow Ontology)
   IV. SCoRO (the Scholarly Contributions and Roles Ontology)
   V. FRAPO (the Funding, Research Administration and Projects Ontology)
d) Ontologies for metrics and statistics for bibliographic resources
   I. BiDO (the Bibliographic Data Ontology)
   II. FiveStars (the Five Stars of online Journal Articles Ontology)

[39] The detailed description and the source for accessing the ontologies themselves are available in Peroni and Shotton (2018).

Together, the 15 ontologies composing the four modules of SPAR Ontologies intend to provide the core elements for describing all aspects of the comprehensive publishing domain, which correspond to much more than citing and referencing domains. FaBiO is the ontology that refers to the FRBR entities describing the type of publications, e.g., books, journals, proceedings and their contents, e.g., books chapters, articles, and conference papers that contain or can be referred to bibliographic references. (PERONI; SHOTTON, 2019).

FaBiO Ontology is based on the WEMI model[40], which triggered the reflections on the conceptual and methodological bases supporting the descriptive representation, showed to be the most comprehensive ontology composing the SPAR Ontologies. In this context, FaBiO Ontology was selected for an analysis for determining its suitability for describing bibliographic elements usually considered in bibliographic references.

8.3.4.Part III – FaBiO Ontology applicability in the description of bibliographic references' metadata

FaBiO's terms are divided into 5 categories: Classes, Object properties, Data properties, Named individuals and, Annotation properties:

a) FaBiO's Classes usually refer to bibliographic element (which are called "objects" by FaBiO) which are referred at work or expression levels (e.g., report class), their realizations (e.g., report document class) and their storage methods and general publication characteristics (e.g., analog storage medium class or hardback class).
b) FaBiO's object properties, are dedicated to describing the relations between Works, Expressions, Manifestations, and Items, represented by each FaBiO object.
c) FaBiO's Data properties are dedicated to describing specific identification data (e.g., DOI, ISBN, ISSN), which can be particularly useful in assuring the proper identification of publications.
d) FaBiO's named individuals are dedicated mainly to describe data concerning the storage media of publications.

[40] WEMI model means FRBR Work, Expression, Manifestation and Item model.

e) FaBiO's annotaiton properties are mainly based on the DCMI Metadata Terms[41] and on The OWL 2 Schema vocabulary (OWL 2)[42] and include terms for the description of properties and relations not approached in the previous categories.
f) FaBiO's terms categories, like the "contributor" term designated for the description of any entity that has contributed for the creation/production of a work.

A detailed description of FaBiO's terms is available in Spar Ontologies documentation available in the SPAR Ontology website [43]. The Ontology itself is available in full in Peroni and Shotton (2019).

Chart 16 evinces the results of the level of coverage of SPAR ontologies, particularly FaBiO Ontology, for describing the "starred metadata" identified in Chart 15. The first column of Chart 16 represents the "starred metadata set" (actually we expanded the bibliographic elements range beyond the "starred metadata"set for this analysis, to have a broader view on the applicability of FaBiO's Ontology to the bibliographic references description). In the second column we grouped FaBiO's terms according to the applicability for describing the descriptive element corresponding to the respective row (first column of the Chart 16). In the third and last column there is the data concerning the types of publications composing the "starred metadata set" (addressed in the first column of the Chart 16), whose bibliographic references referring to them provide the descriptive element to which the respective row of the Chart 16 refers to. The highlighted FaBiO's terms indicate those applicable to the category of descriptive elements to which the respective row of the Chart 16 refers to and were, in fact, observed within our sample. For instance, the FaBiO Class "critical edition" can be associated to the descriptive elements describing the edition of a publication. However, none of the bibliographic references in our sample indicated a critical edition as a descriptive element. So, such term was considered applicable to the category of the descriptive element "edition" but, not to any metadata observed in our sample and, therefore, it was not typed in bold characters.

[41] Available from DCMI Usage Board (2020).
[42] The content of this ontology is based on Tables 6.1 and 6.2 in Section 6.4 of the OWL 2 RDF-Based Semantics specification, available at http://www.w3.org/TR/owl2-rdf-based-semantics/.
[43] Available from http://www.sparontologies.net/.

Chart 16 - The suitability of FaBiO ontology for describing the “starred bibliographic metadata set”

Descriptive elements composing bibliographic references	FaBiO’s entities attending to the descriptive element aspects	Types of publications whose bibliographic references of the sample address the correspondent descriptive element
Chapter title	***Has title*** ***Title***	Book chapters, technical reports chapters, e-book chapters, book series
Chapter title in English (when original title is in another language)	***Alternate title, Has translated title***	No occurrencies filling the “starred metadata” set requisites
Conferences’ title	***Has title*** ***Title***	Proceedings and conference papers
Journals’ title (abridged format)	***Journal*** ***Has short title***	Articles and forthcoming articles
Journals’s title (full format)	***Journal*** ***Has title*** ***Title***	No occurrencies filling the “starred metadata” set requisites
Journals’s title in English (for titles in other languages)	***Alternate title, Has translated title***	No occurrencies filling the “starred metadata” set requisites
Newspaper/magazine title	***Has title*** ***Title***	Newspapers and magazines
Proceedings’ title	***Academic proceedings*** ***Has title*** ***Title***	Proceedings
Session title	***Has title; Section*** ***Title***	No occurrencies filling the “starred metadata” set requirements
Works’ subtitle in original language	***Has subtitle, Has translated subtitle***	No occurrencies filling the “starred metadata” set requisites
Works’ title in original language	***Journal article, Magazine article***	Articles, books, book chapters, proceedings, webpages, technical reports, technical reports

	Has title ***Title***	chapters, working papers / preprints, conference papers, grey literature, unpublished works, forthcoming articles, online databases, newspapers and magazines, e-books, e-book chapters, softwares and applications, patents, manual, guides and toolkits, data sheets, standards, personal communications
Works's title in English (when original title is in another language)	***Alternate title***	Proceedings
Author full name	***Has creator*** ***Creator***	No occurrencies filling the "starred metadata" set requirements
Chapter's author	***Has creator*** ***Creator***	Book chapters, technical reports chapters, e-book chapters, book series
Proceedings' editor	***Has creator*** ***Creator***	No occurrencies filling the "starred metadata" set requirements
Translator	***Contributor***	No occurrencies filling the "starred metadata" set requirements
Work's author or editor	***Has creator*** ***Creator***	Articles, books, book chapters, proceedings, webpages, technical reports, working papers / preprints, conference papers, grey literature, unpublished works, forthcoming articles, online databases, newspapers and magazines, e-books, e-book chapters, softwares and applications, patents, manual, guides and toolkits, standards, personal communications
Edition number	Critical edition ***Has edition***	Books, e-books, standards
Issue number	***Journal issue, Magazine issue, Newspaper issue, Periodical issue*** ***Has edition; Has issue identifier***	Articles, book series
Revision number		No occurrencies filling the "starred metadata" set requisites
Version number	***Has version identifier***	Online databases, softwares and applications, manual, guides and toolkits
Volume number	***Journal volume, Periodical volume*** ***Has volume idenfitier***	Articles, technical reports chapters, forthcoming articles, book series

Series number	***Has number***	No occurrencies filling the “starred metadata” set requirements
Conference date	***Date***	No occurrencies filling the “starred metadata” set requirements
Conference place		Conference papers
Date of citation (date of access)	***Has access date***	Webpages, technical reports, working papers / preprints, online databases, e-books, e-book chapters, softwares and applications
Date of last update / revision	***Date last updated; Has modification date***	No occurrencies filling the “starred metadata” set requirements
Day of publication	***Has date; Has publication date***	Newspapers and magazines, personal communications
Month of publication	***Has date; Has publication date***	Forthcoming articles
Place of publication	***Has place of publication***	Books, book chapters, proceedings, technical reports chapters, working papers / preprints, grey literature, e-books, e-book chapters, softwares and applications, patents, manual, guides and toolkits, standards, personal communications
Proceedings date of publication	***Has date***	No occurrencies filling the “starred metadata” set requisites
Publisher	***Has publisher***	Books, book chapters, proceedings, technical reports chapters, working papers / preprints, grey literature, unpublished works, online databases, online databases, e-books, e-book chapters, softwares and applications, manual, guides and toolkits, book series, data sheets, personal communications
Year or date of publication	***Has date, Has publication date, Has publication year***	Articles, books, book chapters, proceedings, webpages, technical reports, technical reports chapters, working papers / preprints, conference papers, grey literature, unpublished works, forthcoming articles, online databases, newspapers and magazines, e-books, e-book chapters, softwares and applications, patents, manual, guides and toolkits, book series, data sheets, standards, personal communications
	Date	
Series title	***Book series,*** Expression collection, ***Series***	Book series

Abstract number	***Has identifier***	No occurrencies filling the “starred metadata” set requirements
Article ID within publisher’s webpage	***Has electronic article identifier; Has identifier; Has PubMed identifier***	Articles
Article number part note		No occurrencies filling the “starred metadata” set requirements
Chapter number	***Has number***	No occurrencies filling the “starred metadata” set requirements
ISBN number	***Has ISBN***	E-books
Paper number	***Has electronic article identifier; Has identifier***	No occurrencies filling the “starred metadata” set requirements
Patent number	***Has patent number***	Patents
Technical report number	***Has identifier***	No occurrencies filling the “starred metadata” set requirements
Work number	***Has identifier***	No occurrencies filling the “starred metadata” set requirements
Work number within the conference	***Has identifier***	No occurrencies filling the “starred metadata” set requirements
Working paper number	***Has identifier***	No occurrencies filling the “starred metadata” set requirements
Publication format specification	abstract, ***academic proceedings***, addendum, algorithm, analog item, analog manifestation, analog storage medium, announcement, anthology, API, application profile, archival document, archival document set, archival record, archival record set, ***article***, artistic work, ***audio document***, authority file, ***bachelor's thesis***, ***bibliographic database***, bibliographic metadata, biography, ***blog***, ***blog post***, ***book***, ***book chapter***, ***book review***, ***book series***, book set, ***brief report***, call for applications, case for support, case for support document, ***case report***, ***catalog***, ***chapter***, citation metadata, ***clinical case report***, clinical guideline, clinical trial design, ***clinical trial report***, collected works, comment, complete works, ***computer application***, ***computer file***, ***computer program***, concept, concept scheme, ***conference paper***, ***conference poster***, ***conference proceedings***, controlled vocabulary, correction, corrigendum, cover, critical edition, ***data file***, data management plan, data management policy, data management policy document, data repository, ***database***, database management system, ***dataset***, definitive version, deliverable, deliverable report, demo paper, diary, ***digital item***, digital manifestation, ***digital***	All types of publications composing the “starred metadata”, except unpublished works and data sheets

	storage medium, ***directory***, discipline dictionary, ***doctoral thesis***, document repository, dust jacket, e-mail, editorial, entity metadata, ***entry***, erratum, ***essay***, examination paper, excerpt, executive summary, experimental protocol, ***expression***, ***expression collection***, ***figure***, ***film***, folksonomy, Gantt chart, grant application, grant application document, hardback, image, in brief, ***in-use paper***, index, ***instruction manual***, instructional work, item, item collection, ***journal***, ***journal article***, journal editorial, ***journal issue***, ***journal news item***, ***journal volume***, laboratory notebook, lecture notes, legal opinion, letter, library catalog, literary artistic work, ***magazine***, ***magazine article***, magazine editorial, ***magazine issue***, ***magazine news item***, manifestation, manifestation collection, manuscript, ***master's thesis***, meeting report, metadata, metadata document, ***methods paper***, ***microblog***, ***micropost***, minimal information standard, model, ***movie***, moving image, musical composition, nanopublication, ***news item***, ***news report***, ***newspaper***, ***newspaper article***, newspaper editorial, ***newspaper issue***, ***newspaper news item***, notebook, notification of receipt, novel, obituary, ontology, ontology document, opinion, ***oration***, page, paperback, ***patent***, patent application, patent application document, ***patent document***, ***periodical***, ***periodical issue***, periodical item, ***periodical volume***, ***personal communication***, ***Ph.D. symposium paper***, play, poem, policy, policy document, position paper, ***poster paper***, ***postprint***, ***preprint***, presentation, ***press release***, ***print object***, ***proceedings paper***, product review, project metadata, project plan, proof, proposition, questionnaire, quotation, rapid communication, ***reference book***, reference entry, ***reference work***, ***relational database***, reply, ***report***, ***report document***, reporting standard, ***repository***, ***research paper***, ***resource paper***, retraction, ***review***, ***review article***, ***review paper***, scholarly work, screenplay, script, ***series***, short story, song, sound recording, ***specification***, ***specification document***, spreadsheet, ***standard operating procedure***, still image, ***storage medium***, structured summary, subject discipline, subject term, ***supplement***, ***supplementary information file***, ***systematic review***, table, table of contents, taxonomy, ***technical report***, ***technical standard***, term dictionary, ***textbook***, thesaurus, ***thesis***, timetable, ***trial report***, triplestore, tweet, uncontrolled vocabulary, vocabulary, vocabulary document, vocabulary mapping, vocabulary mapping document, ***web archive***, ***web content***, ***web manifestation***, ***web page***, ***web site***, ***white paper***, wiki, wiki entry,	

	wikipedia entry, work, work collection, work package, workflow, ***working paper***, ***workshop paper***, ***workshop proceedings*** ***Has embodiment***; ***Has format***; ***Is embodiment of***; ***Is manifestation of***; ***Is stored on***; ***Stores*** Analog magnetic tape; ***CD***; Cloud; Digital magnetic tape; Digital versatile disc; Film; Floppy disc; HD; ***Internet***; Intranet; ***Paper***; RAM; Solid state memory; Vinyl disk; ***WWW***	
Abridged work pagination lenght (e.g. 80-9)	***Has page range***	Book series
Cited work pages range (e.g. 80-89)	***Has page range***	Book chapters, proceedings, e-book chapters, patents
Work's first page number (e.g. 80)	***Has starting page***	Book series
Work full pagination length (e.g. 80-89)	***Has page range***	Articles
Work's total number of pages (e.g. 80 p.)	***Has page count***	Standards
Work's language note	***Has language***	No occurrencies filling the "starred metadata" set requirements
Supplemental issue note		No occurrencies filling the "starred metadata" set requirements
Special issue note		No occurrencies filling the "starred metadata" set requirements
Supplementary content note	Addendum, ***Supplement***, ***Supplementary information file*** ***Is part of***	No occurrencies filling the "starred metadata" set requirements
General notes	Hardback Comment	No occurrencies filling the "starred metadata" set requirements
Unpublished note		Unpublished works
In press note		Forthcoming articles
Database system number	Has ArxXiv identifier; ***has electronic article identifier; Has identifier***	No occurrencies filling the "starred metadata" set requirements

Hypertext hyperlink (URL)	***Web manifestation, Web page, Web site***	Webpages, technical reports, technical reports chapters, working papers / preprints, forthcoming articles, online databases, newspapers and magazines, e-books, softwares and applications, personal communications
	Has URL	
DOI string or DOI URL	***Has DOI***	Working papers / preprints, forthcoming articles, online databases, e-books, e-book chapters, book series
Online availability note		No occurrencies filling the "starred metadata" set requirements
Institutional link (university department)		No occurrencies filling the "starred metadata" set requirements
Latin expression "in" (i.e. for book chapters or conference papers in a proceedings)	***Has part; Is part of***	Book chapters, proceedings, e-book chapters

Legend for Chart 16

Cell colors	FaBiO's Ontology cagetories
	Classes
	Object properties
	Data properties
	Named individuals
	Annotation properties
	No FaBiO terms applicable

The results of crossing the terms that compose FaBiO Ontology with the descriptive elements identified as the “starred metadata set” showed that some few terms of FaBiO Ontology did not match any descriptive element considered in the Chart 16, since they suggest being used in specific descriptive contexts. Such terms, which integrates the FaBiO category “Annotation properties” are: description (A short statement containing a brief description of the content of the publication, like a summary or an abstract); label (A statement that represents the concept and essential nature of the term (DCMI Usage Board, 2020), Preferred Namespace Prefix (the preferred namespace prefix to use when using terms from VANN vocabulary[44] in an XML document (DAVIS, 2005)), Preferred namespace URI (the preferred namespace URI to use when using terms from VANN vocabulary in an XML document (DAVIS, 2005)), Rights (refers to data concerning legal rights under a particular publication), See also (a term used for establishing relations between two or more objects), Version info (used for describing the version of the ontology adopted for writing a bibliographic description). The bibliographic elements for which we did not found a correspondent FaBiO term are represented by the red cells in the Chart 16.

We also observed that some types of publications may have more than one date to be considered, from the bibliographic description perspective. For instance, the date which a particular event took place and the date the respective annals/proceedings were published are not necessarily the same and, such different metadata can negatively influence both the identification and the retrieval of the publications represented by the bibliographic references. In bibliographic references concerning conference proceedings, the provision of both the date the event took place and the date the annals/proceedings were published are required, according to several reference styles’ guidelines, like ABNT reference style for example. The aforementioned descriptive elements (and others) can be (preferably) automatically collected and interpreted by reference managers but, without registering the differentiations between such dates in the description of bibliographic data at the semantic level, reference managers should not be provided with the necessary elements to properly establish the relationships between the bibliographic metadata

[44]VANN means “A vocabulary for annotating vocabulary descriptions”. Available from: https://vocab.org/vann/#preferredNamespacePrefix

and the piece of information they represent by themselves, i.e., without the human intervention. Therefore, the chances that reference managers provide incorrect or incomplete bibliographic references are greater. However, we did not observe a way for properly describe the chronological differences approached above, using FaBiO's terms and, therefore, this sets up an improvement suggestion for FaBiO Ontology.

FaBiO Ontology has proven to be sufficient to properly describe most of the descriptive elements referred in Chart 15. However, the analysis evinced that there is still a lack of FaBiO's terms to describe specific types of publications that, although they are not among the most commonly cited ones by the scientific community, they still constitute citable publications and, therefore, such aspects should be considered by the instruments dedicated to bibliographic description. It is the case of engravings, lithographs, documents in press or unpublished publications and television programs. Here we have another improvement suggestion for FaBiO Ontology.

9. DISCUSSION

In this section, we discuss the previously introduced results focusing on answering the nine research questions (RQ 1 – RQ 9) previously introduced in the section 2 – The problem. In particular, the discussion is organized in three parts. In the first part, we address RQ 1 – RQ 3, while in the second part we discuss RQ 3 – RQ 6. In the third and last part of this session, we addressed RQ 6 – RQ 9. Below, a reminder of the research questions which this discussion focuses on:

a) Part I
   - RQ 1.Considering current bibliographic tools like reference styles and reference manager softwares, were they effective on fully addressing the issues pointed out by the study made by Sweetland in 1989?
   - RQ 2.Are there other possible causes for errors in citing and referencing other than those specified by Sweetland?
   - RQ 3.What impacts are to be expected by readers on retrieving information from citing and referencing metadata, considering the current descriptive representation revision and the potential differences between the level of description adopted by bibliographic catalogs and bibliographic references?

b) Part II
   - RQ 4 - Considering the changes in the production, storage, retrieval, and use in the information's universe, do Sweetland's (1989) claims remain updated?
   - RQ 5 - Are there current possible causes for citing and referencing errors other than those pointed by Sweetland's study?
   - RQ 6 - Which are the possible impacts of the FRBR approach regarding descriptive representation on information retrieval from citing and referencing metadata perspective?

c) Part III
   - RQ 7 Which are the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations and bibliographic references, considering different types of cited works?
   - RQ 8 Is there a common metadata set used across the disciplines for describing cited works within bibliographic references?

- RQ 9 Do SPAR Ontologies comply with bibliographic elements composing bibliographic references, particularly the “starred metadata set”?

9.1. Part I - Citing and referencing habits among Medicine and Social Sciences journals

This session addresses the discussions on the analysis of data presented in sessions 7.1.1 and 7.1.2, referring to the findings concerning the observations on the articles from Medicine and Social Science subject areas, and aims, primarily, to answer RQs 1 up to 3.

9.1.1.Part I - Issues in citing and possible causes for errors

90.23% of Medicine bibliographic references refer to articles published in journals. This is probably due to the workflow of journals which is relatively more dynamic compared to books, which favors fast discussions in health sciences. Instead, in Social Sciences domains, aspects of sociological, historical, cultural, political, chronological, anthropological, and geographical nature, directly influencing research trends are approached both in articles and in other types of publications, such as books. This justifies the greater variety of typology of works cited by social sciences articles in comparison to medicine articles according to data shown in Graphic 3.

While usually published in restricted access journals, Medicine articles of our sample showed that they cited freely available articles more than restricted access ones, even when such cited articles were published in restricted access journals. However, from a Social Sciences perspective, we noticed the opposite behavior, since Social Sciences articles in our sample tended to cite restricted access articles without providing any metadata for their freely available version (if any).

Providing URLs and hyperlinks within bibliographic references was a frequent behavior in Medicine articles. However, the accuracy of the location pointed by such URLs/hyperlinks was more reliable in Social Science articles. Indeed, the number of times an URL/hyperlink pointed to a wrong Web location, e.g., to an online bibliographic catalog instead of the cited work itself, was higher in Medicine articles.

This suggests two things. First, Social Science authors' awareness of promoting the access to freely available content may result in the higher reliability of the hypertext links provided in their bibliographic references. Second, we did not find detailed instructions regarding the provision of URLs and DOI hyperlinks within bibliographic references in reference styles (RQ 1). In particular, while usually encouraging authors to provide these bibliographic data, the reference styles analyzed generally did not indicate clear guidelines regarding the description procedures neither on the location where those hyperlinks should point to, such as the actual file containing the cited work or to the landing page from where it may be downloaded (RQ 2).

These behaviors suggest a resistance – by authors, editors, or both – on providing enough metadata in the bibliographic reference to facilitate the access to the cited works (RQ 2). Although the articles' content and its accuracy are authors' responsibilities, there did not seem to be an adequate verification of the bibliographic references before their publication, which may suggest that the efforts devoted by publishers of the articles in our sample in providing trustful bibliographic data in their articles, especially concerning bibliographic references, were not as meaningful as those related to content quality. In this sense, Sweetland (1989, p. 300) considers that "the role of citations is not taken very seriously by the scientific community". Cronin (1982, p. 71) complements that "journal editors and referees could pay greater attention to the quality and quantity of references".

Our analysis showed the adoption of 20 different reference styles within the journals of the sample. According to Graphic 4, the average number of reference styles adopted by journals composing the sample is 45% higher in Social Sciences journals than that observed within Medicine journals. Such diversity in the adoption of reference styles in the same subject area weakens the argument that the existence of multiple reference styles is justified by the specific needs in different disciplines (GRATZ, 2016), specific rules for a particular audience (BARBEAU, 2018) and by tradition (BARBEAU, 2018; BIBME, 2017; GRATZ, 2016), for instance:

> "social sciences tend to use current research, so the publication date of a source is very important. For this reason, people in the social sciences tend to use APA style; APA style puts the date before all other information, aside from the author's name, which makes it much easier for researchers in that field to find valid, up-to-date information." (UCM WRITING CENTER, 2016).

This huge variety of reference styles disfavors the uniform bibliographic metadata description in bibliographic references since different reference styles can recommend different ways of representing the same information. As stated by Sweetland, “complaints about lack of uniformity are common in the literature, whether from authors or librarians”. The author complements that “given the variety of formats for citation and the lack of any real agreement among journals or authors, the chance of misunderstanding is high” (SWEETLAND, 1989, p. 298).

Some journals recommend more than one bibliographic reference managers within the instructions provided to authors for writing their manuscript. Updating the format used by such reference managers to return bibliographic references according to the reference styles’ guidelines can prevent journals from receiving differently formatted articles. Theoretically speaking, that could decrease the editorial work on checking bibliographic metadata accuracy.

From the reference manager administrator’s perspective, the more the number of reference styles available is huge, the greater the challenge of creating algorithms for recognizing and stylesheets for formatting bibliographic references appropriately. Besides, an incredible effort is needed to update such algorithms and stylesheets when, for instance, particular reference styles have variations on their citing systems (BARBEAU, 2018) and multiple versions and editions of the same reference style are introduced in time.

The diversification of guidelines may confuse researchers. A clear example of this situation is the name used by publishers to refer to reference styles in their homepages. According to the data we gathered, we noticed several variations – e.g., AMA Reference Style was often mentioned as AMA Manual of Style, Harvard Reference Style was also referred to as Harvard Reference System, and Vancouver Reference Style was also referred as ICMJE and Citing Medicine.

We also noticed that the name attributed to the section containing bibliographic references in Social Sciences articles may change. For example, sections containing bibliographic references are named differently within articles adopting the 16th edition of Chicago reference style: 93.8% name it “References” and 6.2% name it “Notes”. Besides, two different citation styles were observed on this same sample slice: citation-sequence adopted by 87.6% of articles and author-date adopted by 6.2% of articles. The remaining portion of articles adopts both citation systems. No misuses of adopted reference styles were observed in medicine articles, considering, specifically these aspects.

We speculate that, in time, failures regarding the interpretation of bibliographic guidelines description cannot be exclusively attributed to the availability or to the simultaneous use of multiple reference styles (RQ 2). Indeed, our analysis revealed that reference styles content could be clearer, including the ones authored or adapted by journal publishers. We observed shortcomings and omissions, as discussed in the results session, which may increase the probability of making mistakes like those introduced in Figures 12-17, as well as those pointed by Sweetland (1989), as approached in Chart 1.

Most of the Medicine journal publishers in our samples adapted reference style guidelines to their particular purposes. However, even with these changes, the reference styles remained vague and imprecise in some respects and may result in errors in mentioning, quoting, and referencing cited works (RQ 2).

Similarly, although Social Sciences publishers generally did not recommend any adaptation to the existing reference styles adopted by their journals, often the guidelines for writing and formatting bibliographic references and in-text reference pointers related to mentions and quotations were not clear and easy to understand. For instance, while the most adopted reference styles focused on instructions for bibliographic references content and formatting, they generally did not provide guidelines for describing all the types of publications cited in the articles in our sample such as engravings and lithography (RQ 1). These shortcomings, also mentioned by Sweetland (1989), often are the reason that denies a precise identification of the cited works by a reader. This situation seemed particularly relevant in Social Sciences articles, which cited many types of publications.

Also, 5.9% of the Social Sciences journals did not adopt explicitly a reference style and, instead, recommended the authors to consult the bibliographic references of the journal sample issue and to consider them as a model for writing and formatting bibliographic references. However, the coverage of the types of cited works in such sample issue was not complete. In these cases, the author had no alternative unless writing the bibliographic reference according to his/her own beliefs on what could be the best way of describing the cited works. This may explain some cases in which required metadata for the identification of the cited work are missed (RQ 2).

Along the same lines, we observed that some of the reference styles provided by the publishers of the Medicine articles in our sample, which adopted a citation-sequence system, did not mention the logic under the bibliographic references should be sorted in the bibliographic reference lists, thus creating even more confusion for the author who wrote them. Also, we observed that four variations of the style of the numerical character of such bibliographic references adopted in citation-sequence system in Medicine articles, while for Social Science articles there were found only two variations.

Regarding how quotations are marked up within the article content, we found cases within Social Science articles in which both double and single quotation marks were adopted for run-in quotations, while long quotations were indicated using indentation which was, sometimes, accompanied by italicized characters and double quotation marks. Besides, we noticed that there were no shared rules used to identify a quotation as run-in or long. When specified, the main strategies adopted in the reference styles of the journal in our sample were:

1. to classify quotations considering the total numbers of words quoted (usually, 80 words at maximum for run-in quotations);
2. to classify quotations according to the length of the quoted passage in the citing work (usually, 3 or 4 lines at maximum for run-in quotations).

Considering the bibliographic references included in Medicine articles in our sample, the names of journals in which cited articles were published are provided in abridged formats, which are usually based on lists providing standardized abbreviations to journal names. In the meantime, we observed five different sources from where these abbreviations can be extracted. In principle, the same journal title can be abridged in at least five different ways, which reinforce the claim by Sweetland (1989, p. 298) in which “differences in journal title abbreviation have been commonly noted as a particular source of error” (RQ 2). Figure 18 reproduces one of the bibliographic references of an article composing our sample (UZUNALLI *et al.,* 2019). In this example, the abbreviation “Ann” can be understood by the reader as “Annals”, “Annual”, “Annalen”, “Annales”, and so on. As an additional source of ambiguities, the guidelines for authors provided in the journal’s webpage did not mention the source from which journal titles abbreviations should be based on.

Figure 18 -A bibliographic reference with an abridged representation of the title of the cited journal.

Fanning AS, Anderson JM. Zonula Occludens-1 and -2 Are Cytosolic Scaffolds That Regulate the Assembly of Cellular Junctions. Ann N Y Acad Sci. 2009; 1165:113–20. https://doi.org/10.1111/j.1749-6632.2009.04440.x. [PubMed]

Thus, abridging journal titles within bibliographic references in Medicine articles seemed not to be fully accomplished. One possible strategy to adopt to disambiguate the journal would be to include complimentary information within bibliographic references, e.g., the cited journal’s ISSN, to assure uniform interpretations of journal titles abbreviations.

Our analysis also revealed a high percentage of articles not including the source of non-textual content mentioned in the textbodies, e.g., tables and figures, especially in Medicine articles. Although non-textual content is also citable, we observed that reference styles rarely provide guidelines on how to proceed with the presentation and description of this kind of content.

Summarizing, all the information presented so far allowed us to conclude that the issues raised by Sweetland’s’ study in 1989 are still valid today (RQ 1). We do not need thousands of reference styles if we do not have clear guidelines on how to inject bibliographic metadata into bibliographic references. In this context, neither the standardization nor the technologies developed were able to fully support the creation and management of bibliographic metadata to write bibliographic references. Reference styles do not provide sufficient and clear information to authors regarding the procedures of citation, metadata description, and formatting. Looking at the articles in our sample, reference styles seemed to be a list of suggestions used by authors and publishers to support their own decisions regarding referencing and citing rather than a concise and precise instrument of guidance compliant with standardized behaviors on citing habits.

As anticipated by Galvão (1998), our analysis confirmed that the terminology used to refer to bibliographical concepts was ambiguous. Terms like “reference styles” and “citation styles”, and “bibliography”, “bibliographic references”, “references”, and “notes” referred to the same thing on different occasions, thus confirming Sweetland’s statements regarding the “lack of uniformity in the literature” and “of real agreements among journals and authors” (SWEETLAND, 1989, p. 298).

The rate of errors in citations in respected scientific journals in Medicine and Social Sciences is high (RQ 2). Sweetland (1989) states that there is little consensus about who should be responsible for correcting citations: publishers may think it is up to the authors while authors would like to have referees and editors to double-check them.

In fact, there are no explicit rules or clear statements on the roles expected to be played by different agents in the editorial field (i.e., authors, editorial boards, and publishers), nor behavioral patterns concerning citing and referencing matters among editorial market agents. For instance, some publishers provide an expert service for normalization and, in such cases, submissions are usually accepted in any formatting format. In other cases, publishers, that accept submissions in any format, provide authors with the adopted reference styles’ guidelines just in case of acceptance of the submitted paper for publication. Again, in other situations, publishers indicate either a particular reference style or a set of guidelines and then leave the primary responsibility for normalization on the authors. In this situation, the limits concerning the purview of the of agents who are expected to directly act on normalization matters are tenuous and, therefore, not precise. This probably contributes to the confusion on the definition of the role to be played by the main core agents involved in the production of scientific content, as previously stated by Sweetland, who complements:

> no one, except perhaps librarians, seems to care very much about the problem. […] We spend considerable time and effort in training catalogers in both the theory and the methods of descriptive cataloging. It would be good to spend at least some effort on training all information workers [including authors and publishers] in the theory and methods of the citation. (SWEETLAND, 1989, p. 301-302).

9.1.2.Part I – The conceptual representation of bibliographic references, mentions and quotations

Toward the end of the 20th Century, librarians began discussing possible changes in the description, access, and encoding of bibliographic information. Typically, especially in bibliographic catalogs which were based on the Anglo-American

Cataloging Rules (AACR) (ANGLO-AMERICAN… 2002), documents were described out of context and their descriptions usually referred to a particular edition published by a specific publisher. In modern times, this approach has been no longer sustainable and compliant with the fulfillment of the functions of bibliographic catalogs in the new information scenario (JOUDREY *et al.*, 2015; TILLET, 2003b; OCLC, c2020).

Between 1992-1995, the IFLA Study Group on Functional Requirements for Bibliographic Records (FRBR) developed an entity-relationship based on a conceptual model for describing bibliographic records for all types of materials. This conceptual model should not only consider the function of the catalogs, which should enable users to find, identify, select, obtain bibliographic resources and navigate within the catalog, but also allow the performance of user tasks associated with bibliographic resources and the conceptual model of the bibliographic universe: the entities, their relationships, and attributes (TILLET, 2003a, 2003b; IFLA, 2009). FRBR determines that items (embodiment of works in physical or electronic publications) must be described in a context in a manner sufficient to relate the item to the other items comprising the work using the four-level bibliographic structure: FRBR Work, FRBR Expression, FRBR Manifestation and, FRBR Item (OCLC, c2020).

Although IFLA Functional Requirements family (FRBR, FRAD, FRSAD) was consolidated by the IFLA Library Reference Model (IFLA LRM) in 2017, we decided to consider the FRBR entities concepts in this study. FRBR concepts are needed to construct a theoretical background that supports further discussions concerning citing and referencing matters – indeed IFLA LRM concepts are more comprehensive than those of FRBR family. In addition, FRBR was the first concrete IFLA initiative that have changed descriptive representation approaches on information and, theoretically, may be considered the starting point of the distancing between the facets which compound descriptive representation.

Even adopting new perspectives, the essence of descriptive representation in bibliographic catalogs still focus on describing a publication as stored in a particular information support, be it physical or electronic. Indeed, despite all these new aspects introduced by FRBR, cataloging, and the preparation of bibliographic references, according to IFLA Study Group on the Functional Requirements for Bibliographic Records (2009), should aim at:

a) identifying all the background information that have supported an author's ideas conception, or the information which reading is being recommended by him as a complimentary content of his ideas;
b) using the information contained in the bibliographic reference (e.g., an author, the title, or a journal) to select an entity that is appropriate to one's needs;
c) finding the entities described.

Since the new approach of descriptive representation materialized in the FRBR guidelines considers document content (FRBR Expression) rather than the format in which it is embodied (i.e., FRBR Manifestation or, more concretely, FRBR Item), the relationship among mentions, quotations, bibliographic references, and in-text reference pointers seems to collide to some degree.

Quotations are the transcription of excerpts of textual contents available in the cited works (FRBR Expressions), while mentions are the author's written interpretations concerning a textual content of the cited works (again FRBR Expressions) that should convey somehow the original idea of the cited author (FRBR Work). Instead, bibliographic references usually focus on one particular embodiment (FRBR Manifestation) which could be available in a single and physical (or electronic) exemplar of a publication (FRBR Item), e.g., the copy of a PDF file.

In this context, it should be considered that a FRBR Expression can be (and actually is) published in different information supports and formats. Since most of the scientific production is available on the Internet (sometimes free of charge) and since this content can assume different FRBR Manifestations, referring to specific FRBR Manifestations within bibliographic references may restrict the possibilities of access for the reader, if he does not have the perception (reading the bibliographic reference) that certain content to which he does not have access to (e.g., in the publisher website) may be available in a different FRBR Manifestation (e.g., in a preprint server). Besides that, in-text citations usually refer to the content of a bibliographic resource, that is, the FRBR Expression of a FRBR Work, while the respective bibliographic references usually refer to its embodiment (FRBR Manifestation). This richness represents also an obstacle to facilitating and improving the ways of accessing information, which also corresponds to one of the purposes of descriptive representation.

In our sample of Medicine and Social Sciences articles we found that the guidelines above were not always followed (RQ 3). In particular, the descriptive elements provided in the main part of in-text reference pointers that accompany quotations pointed to the FRBR Manifestation layer of the cited work due to the presence of the pages where the quoted text is contained. However, a smaller portion of such in-text reference pointers did not provide any descriptive element for referring to the FRBR Manifestation layer and actually seemed to relate to the pure content of the cited works (i.e., their FRBR Expressions). Besides, we found that the main part of bibliographic references referred to a particular FRBR Manifestation of the cited works. However, there were a few cases where it was not possible to identify even the FRBR Expression of the work defined by a bibliographic reference. For instance, the bibliographic reference “World’s Work. 1909. "The March of Events.". World’s Work, December.” found in one of the articles analyzed did not provide enough descriptive information to classify it properly.

Usually, the reference styles of the journals considered in our analysis encourage authors to provide page numbers in the in-text reference pointers referring to quotations, which can be used to locate the excerpt within the cited work if we strictly consider the particular FRBR Manifestation of the cited article referenced by the particular bibliographic reference. Although considering a broader context where each cited work can be fully characterized according to FRBR, that scenario can be restrictive, since does not allow one to consider the possible various kinds of print-like embodiments (printed within a volume, PDF in the publisher website, PDF in an institutional repository, etc.) the work may assume. In addition to that, we found there are no clear guidelines when the cited works cannot be paginated, such as in HTML versions of articles, speeches, digital media content, and tridimensional objects.

The evolution that has characterized the universe of information sciences may have brought some additional challenges in specific processes, such as the bibliographic normalization activity. In this work, bibliographic normalization was addressed as a facet of descriptive representation. Therefore, it has been inevitable to consider the evolutions and improvements that have been carried out in the cataloging domain, including the introduction of FRBR. Conceptual changes in cataloging should potentially impact the way bibliographic metadata is written and managed, the relations between in-text reference pointers, bibliographic references and the cited works, and bibliographic catalogs. Cataloging description level concept under FRBR, which corresponded to the descriptive comprehensiveness, needs to consider different aspects of the same work, which may switch the interpretation of the metadata available in the catalog. This is something to be considered in bibliographic normalization activities, to understand and, eventually, foresee how these changes can impact the way bibliographic references should be presented (RQ 3).

9.2.Part II – Citing and referencing habits across all scholarly disciplines

This session addresses the discussions on the analysis of data presented in session 7.2. referring to the findings concerning the observations on the articles from all the 27 SCImago's subject areas, and aims, primarily, to answer RQs 4 up to RQ 6. The first part of this session will focus on the RQs 4 and 5, while the second part will focus on the answering RQ 6:

RQ 4 - Considering the changes in the production, storage, retrieval, and use in the information's universe, do Sweetland's (1989) claims remain updated?

RQ 5 - Are there current possible causes for citing and referencing errors other than those pointed by Sweetland's study?

RQ 6 - Which are the possible impacts of the FRBR approach regarding descriptive representation on information retrieval from citing and referencing metadata perspective?

9.2.1. Part II – The same data, several representations

According to the suggestions provided by Ranganathan's Laws (ZABEL; RIMLAND, 2007), it is crucial to assure that the metadata embedded in the bibliographic references and in-text citations are provided to readers in a comprehensive way to allow them to easily identify the cited works. Such needs by themselves require and justify the efforts on bibliographic metadata standardization.

There are around 1,500 different bibliographic styles and others 7,000 derived ones, "thus covering more than 8,500 different publication venues (broadly construed to include universities and departments)" (KARCHER; ZUMSTEIN, ca. 2018). The Citation Style Language project (CSL, 2020?), which maintains a repository with more than 9,500 citation styles available in an interchangeable format, encourages publishers to standardize the use of reference formats across journals, and to adopt only a few different reference formats (CSL, 2020?). However, our analysis showed that each knowledge area adopts different styles within their periodical publications, as also observed by Galvan et al. (2017 apud CARRIÓN *et al*, 2017), and that there is a variation in the adoption of reference styles even among journals from the same discipline. For instance, we detected 11 different reference styles among journals from the Medicine subject area (Graphic 7).

According to data showed in Chart 7 and considering the widely adopted reference styles, Vancouver was the most adopted among journals from Health Sciences and Physical Sciences, APA 6th ed. was the most adopted one among Social Sciences journals, and Chicago (no specific edition) was the one most adopted among Life Sciences journals. The percentage of journals adopting their "own reference styles" surpassed all those concerning other reference styles. Indeed, 100% of Multidisciplinary journals adopt their "own reference styles".

In theory, having such a volume of journals adopting a reference style under the same label, i.e., "own reference style", may give a false impression of a "more appropriate" behavior concerning normalization purposes, compared with the remaining subject categories. In fact, the label "own reference style" denotes either a set of citing and/or referencing guidelines authored by journals publishers or the customized versions of existing reference styles (usually those widely accepted by the scientific community, like APA, Chicago, and Vancouver as approached in Graphic 7). So, although clustered under the single label "own reference style", the reference styles addressed in this category actually refer to a lack of standardization, since the reference styles composing this category are generally different from each other and contribute to increasing the number of thousands of existing reference styles (RQ 5). Thus, disciplines adopting "own reference style" should be considered the less engaged with the normalization matters, since the lack of standardization is one of the main contributing factors for citation errors (SWEETLAND, 1987) (RQ 4). Besides, customization is expensive (HOFFMAN 2009) and disagree with standardization purposes, since a customized version of a reference style, indeed, configures a new reference style (RQ 5).

The wide variety of reference styles adopted across disciplines, with different (and sometimes insufficient and/or unclear) instructions for bibliographic metadata structuring, as considered in Graphic 7, supports Galvan's *et al*. (2017 apud CARRIÓN *et al*, 2017) arguments that the presentation of bibliographic references is one of the most frequent problems found in the scientific literature. Besides, the multiple coexisting editions of some reference styles, i.e., AMA, APA, and Chicago represent issues in the normalization field. For instance, we were not able to identify which edition of Chicago was the one most adopted among Life Sciences journals. Some reference styles are likely to be revised from time to time and, such updates are released as new versions or editions of the main reference styles, but sometimes, such revisions do not include major increments which justify the release of a new edition of it. Besides confusing authors, those versions of "more of the same" favor the occurrence of differences between bibliographic references addressing the same type of publication but considering different editions of the same reference style (RQ 4). Thus, it is crucial that journals adopting one of those multiple versions of reference styles specify the precise edition of the adopted reference style authors are expected to consider while formatting bibliographic metadata of their works. Nevertheless, our analysis suggested that it is not the way it usually happens. For instance, 25.93% of Medicine journals adopt the AMA reference style. From this portion of the sample, 7.41% of journals indicates the AMA 9th edition, 11.11% indicates the AMA 10th edition, and 7.41% do not provide accurate instructions concerning which edition of such reference style authors should consider, as addressed in Graphic 7 (RQ 5). As compensation for such negative effects, the very least that can be expected from each of these reference styles is to provide clear and precise information covering the maximum facets concerning citing and referencing data, including aspects related to the type of publications of cited works – even if our analysis has shown that this is not the case.

9.2.2.Part II – The disadvantages of reinventing the wheel

Our analysis showed a tendency that reference styles authored by publishers usually do not provide clear guidelines. From Graphic 11, we can speculate about a direct relation between journals adopting reference styles authored by their respective publishers and the rate of reference styles classified as not clear, i.e., those not providing clear formatting guidelines for citing and referencing metadata. Such results suggest a tendency that reference styles authored by publishers provide citing and referencing formatting guidelines in a relatively lower degree of clarity, in comparison with those widely adopted ones (RQ 5). In general, we observed a lack of details on how to reference specific types of publications, such as online grey literature, and a lack of metadata regarding cited works freely available online. It was also observed that the reference styles provided by publishers rarely provide instructions on referencing and citing and this may contribute to errors on this matter (RQ 4). For instance, Chemical Engineering, Neuroscience, and Nursing subject areas were the most accurate disciplines in this context, all providing clear and complete guidelines. Indeed, none of them adopt “own reference styles” (Graphics 7 and 11) but rather well-known and shared guidelines. Therefore, providing own reference styles seems to be a counterproductive practice since it does add ambiguities, often does not clarify how to cite, and reference appropriately, and does not provide a comprehensive approach to address these matters (as evinced in Graphic 11). Thus, they do not favor the author’s work. The elaboration of yet another personal reference style configures it as a duplicated work, since usually its descriptive content is limited, and its purposes could be addressed more appropriately by the well-known and shared reference styles (RQ 5). As Sweetland (1989) claimed, publishers could be more zealous on behalf of standardization of bibliographic metadata matters (RQ 4).

It is worth mentioning that our analysis did not consider issues related to the content of the well-known reference styles. That is to say that, although journals indicating those documents were classified as “providers of clear citing and referencing guidelines”, this statement can mask a different scenario, i.e., being a well-known reference style should not be considered as a synonym of being a clear reference style, since the guidelines provided by those reference styles may be unclear as those provided by the ones authored by journals publishers. Therefore, further studies broadening and deepening those approaches are encouraged.

From the data gathered, we also noticed a lack of commitment of some publishers to standardization issues when they instruct authors on how to proceed citing and referencing bibliographic metadata by asking them to follow the citing and referencing patterns from articles published in previous issues of the journal or from sample articles provided by publishers. There are also the cases in which authors cannot identify the reference style adopted by the journal since the link provided by the publisher, which should point to the reference styles' guidelines, did not work. In other cases, journals accept submissions in any formatting style and, in case of acceptance for publication, citations and bibliographic references should be further formatted by the publisher's editorial team - no mention about the reference style adopted. Finally, some publishers do not provide information concerning the reference style adopted by their journals within the instructions to authors made available on their webpages – not to mention the case of a particular journal recommending authors to consult the bibliographic references lists of the articles published in the previous issues of the journal and consider them as a model for formatting their own bibliographic references. For such cases, we understand that besides not providing a particular reference style (not even one authored by the own publisher) and besides not providing a single article to be used as a model, those publishers take the risk of reiterate errors introduced by the authors in previous journal issues (SWEETLAND, 1989) (RQ 4). Therefore, in all these cases, authors have no choice but to format in-text reference pointers and bibliographic references following examples provided in sample articles which, of course, may not provide models for all the types of publications cited. Indeed, authors may need to cite different types of works than those referenced in the sample articles, and they have to proceed with such a formatting task using their intuition (RQ 5). Considering all these issues, "journal editors and referees could pay greater attention to the quality and quantity of references" (CRONIN, 1982, p. 76) (RQ 4).

9.2.3.Part II – Limited expression

Besides not providing clear and comprehensive guidelines, some "own reference styles" still limit the maximum number of bibliographic references allowed per article.

However, the number of 30 bibliographic references usually determined by those reference styles is smaller than the average number of bibliographic references approached in bibliographic reference lists of the journals of the correspondent subject areas. This suggests that, in those cases, limiting the maximum number of bibliographic references allowed per article may also limit the needs of expression of some authors. For instance, except for Computer Science, all the subject areas showed average use of more than 30 bibliographic references per article (Graphic 9). Also, we noticed cases of articles not observing the maximum number of bibliographic references allowed per article, thus going against the journal's editorial policies. Three distinct aspects are suggested from this scenario. First, the editorial policies are not always properly observed (and it seems to be okay for publishers). Second, limiting the number of bibliographic references per article do not necessarily limit the author's citing habits (for instance, we noticed an article from Agricultural and Biological Sciences with 58 bibliographic references, while the maximum allowed by the publisher was 36). Third, editorial policies are not effective in all the cases, that may denote a weakness in publishers' authority on granting the compliancy with particular editorial requirements. For instance, the Materials Science subject category, which registered the highest number of journals limiting the maximum number of bibliographic references within bibliographic references lists (75%), showed an average of 42.25 bibliographic references per article, which is 40.83% above the limit of 30 bibliographic references per article (RQ 5). It is worth mentioning that, in our study, we did not consider whether the presence of limits for the number of bibliographic references per article may have impacted the average number of mentions and quotations per article. Further studies are needed to address this aspect properly.

We did not identify a standard behavior on citing habits across articles of the same discipline. Instead, we noticed huge differences considering the maximum and the minimum number of bibliographic references in the articles and, therefore, those numbers should not be considered as trustful metrics for defining citation habits across subject areas. For instance, Medicine articles of our sample addressed from 7 up to 157 bibliographic references per article, which indicates huge degree of heterogeneity that may be even biased by the journals limiting the maximum number of bibliographic references per article, which may contribute to mask the citation habits' outlook. As shown in Graphic 9, 67.6% of articles had between 21 and 60 bibliographic references per article, while the 21.8% of articles had between 61 and 130 bibliographic references per article.

9.2.4.Part II – A (very) brief epistemological approach

Galvão (1998) states that Librarianship does not propose new concepts but, instead, import them from other disciplines. By not establishing a debate on which concepts are imported and which names are used to refers to such concepts, Librarianship has not established a necessary epistemological break for the constitution of its own conceptual framework. According to Galvão (1998), in the historical path of librarianship and documentation, concepts and denominations from other sciences or disciplines have always been imported and/or adapted to obtain solutions to practical problems. In this sense, concepts, and denominations of some theories of Administration, Linguistics, Logic, Communication and other areas were imported. However, aiming at an immediate use of these concepts and denominations, there was no understanding/problematization of the methodological concepts and descriptive concepts and a questioning about the implications of the use of concepts and denominations coming from different areas of knowledge. Thus, in this area, the break with common sense seems to be very fragile, or almost nonexistent.

An example of Librarianship's epistemological problems, previously introduced by Galvão (1998), is the heterogenicity among titles addressed to the bibliographic reference sessions within articles (Graphic 22). This scenario goes beyond standardization issues and suggests misunderstandings concerning epistemological and conceptual matters concerning bibliographies, bibliographic references, references, and notes (RQ 5). Once more, we noticed non-standardized procedures, even within articles of the same journal (RQ 5).

An example of such statements can be noted on the conceptual approaches of the terms “citation styles” and “reference styles” (also referred to as “style guides”), which are frequently referred to by publishers as being synonymous. In theory, citation styles are a set of guidelines concerning the proper way of using mentions and quotations, considering their length (as a criterion for defining long quotations and run-in-quotations), markups (as a criterion for delimiting the beginning and ending of a quoted passage), formatting (as a mechanism of assigning a visual identification to mentions and quotations), and the structuring of in-text reference pointers. Reference styles, in turn, corresponds to the tools providing instructions on how to structure and format bibliographic metadata in bibliographic references and their organization into the bibliographic reference lists. In this framework, citation styles should focus on citing (i.e., mentioning and quoting and related apparatus) procedures and bibliographic references should focus on referencing procedures. However, we noticed that publishers usually do not provide appropriate information for citation styles. Instead, citing issues are seldom and superficially addressed within reference styles' contents, which often contain different interpretations for equivalent or similar rules. For instance, quotations are usually distinguished within text bodies by using markups, including indentations which, by their turn, are usually defined according to the length of the quoted passage, which may vary from journal to journal. Some journals consider the extent of the quoted passage according to the number of quoted lines in the cited work. Some others consider the number of quoted lines in the citing work. Still, others consider the number of quoted words. Publishers seem to consider mentions, quotations and in-text reference pointers as separate elements when they are part of the same whole that complement each other. Also, besides not providing clarifications concerning such differences, publishers still format quotations in their articles in different ways, as shown in Chart 10, suggesting a disagreement concerning the conceptual parameters for classifying a quotation as long or short and, consequently, the criteria on when to use indentation or quotation marks to markup them (RQ 5).

9.2.5.Part II – Mentions and quotations markups

Despite the importance of clearly differentiate self-authored statements from the quoted ones, we detected cases whose equally lengthened quotations were differently marked up within the same article. Because issues concerning markups and formatting matters are rarely addressed within reference styles, there are no parameters to evaluate whether the choice of each journal on this matter is correct or not (RQ 5).

From the in-text reference pointers' perspective, the scenario is similar. Reference styles rarely provided clear and appropriate instructions to deal with in-text reference pointers. In the same line, some reference styles and standards are flexible on the consideration of both author-date and citation-sequence systems, e.g., ABNT Standard and Chicago Reference Style. For these cases, it is up to the publisher to define which citation system should be adopted by their journals and guide authors to their proper use – something that, according to the findings of the research, does not usually happen (RQ 5).

9.2.6.Part II – The citation systems: citation-sequence and author-date

Graphic 19 evinces a balanced scenario between journals adopting the author-date citation system and the citation-sequence system.

However, it is worth highlighting the 0.2% of the articles (corresponding to 5.9% of Social Science articles adopting The Chicago Manual of Style, 16th ed.), adopting both author-date and citation-sequence systems, within the same articles. In those cases, mentions and quotations are both marked up with superscript numbers, denoting the bibliographic reference of the cited work, which is provided in a footnote (citation-sequence system). Simultaneously, the bibliographic reference lists consider the same bibliographic references addressed in the footnotes, in an alphabetical assortment (author-date system). Such behavior denotes unpropped behavior of the publisher regarding citation matters which goes against standardization purposes on facilitating the correlation between bibliographic metadata and the works they address (RQ 5).

The actual open access scenario is gradually introducing changes into journal editorial processes, although some traditions remain in force, such the case of most Health Sciences journals, which adopt citation-sequence system (LÓPEZ CARREÑO; MARTÍNEZ MÉNDEZ, 2015). However, such changes arise questions on which of those traditions remain necessary and justifiable considering where descriptive representation is going nowadays.

Although adopting completely different ways of presenting data and bibliographic metadata, articles should not be assumed as unproperly formatted before being faced with the guidelines of the reference style supposed to comply with. That is to say that, considering the thousands of existing reference styles, it can be appropriate to present bibliographic metadata in different ways if they are compliant with any bibliographic style (RQ 4). One contributing factor to the vastness of the citation styles is the fact that certain styles have variations on their citing systems, like Chicago style (BARBEAU, 2018) (RQ 4) and, especially in these cases, publishers should devote efforts in explaining to authors the interpretation and selection of the alternatives offered by the reference style, since doing customizations and amendments to such styles do not make the normalizing tasks simpler (RQ 5). It is worth remembering that Social Sciences was the subject area which showed the second highest rates of reference styles adopted and the only discipline showing an article using both citation systems simultaneously, which supports the claim that the excesses in the variety of guidelines and the omission of editors can be disruptive to standardization issues (RQ 4).

9.2.7.Part II – The huge range of cited works

7 of the 36 types of publications –articles, books, proceedings, web pages, technical reports working papers, and conference papers – represent 99% of the type of publications addressed in bibliographic references. From this portion, 91,48% correspond to articles as shown in Graphic 12. However, some types of publications less frequently cited – as the case of government official publications cited by Medicine articles, and photographs, engravings, and lithographs cited by Social Sciences – are not even mentioned by most of the reference styles, even considering the most adopted ones, as addressed in Graphic 12 and Chart 14. Because of the specific features of such less cited types of publications, their explicit and proper representation within bibliographic references become more and more important considering, principally, the accomplishment of the function of allowing the reader to precisely identify of the cited work represented by bibliographic references. It turns out that in some cases, like the ones mentioned above, the clear and appropriate bibliographic description may demand some skills and a minimum background knowledge in Information Science to properly interpret and adapt specific reference styles' guidelines, especially when considering the less cited types of publications,

which according to our analysis is not what normally happens. Besides that, even assuming that authors are committed to doing their best to provide a reliable description of works cited by them, it should not be ignored the observed fact that, in most cases, reference styles do not provide guidelines for the bibliographic description of such less cited types of publications, whether clear or not, what strongly contributes to the writing of bibliographic references not accomplishing to their main function of identifying cited works (RQ 5). Since standard guidelines are not always hardly followed, the proper identification of the type of publication described can remain dubious, as per 0.66% of the bibliographic references of our sample, in which we were not able to identify the type of publication they refer to, as shown in Graphic 12 and Chart 13. Aiming to explicitly indicate specific identification requirements of specific types of publications, such bibliographic references should consider providing at least the descriptive elements like the media type, the carrier type, or content type of the cited work, as suggested in chapters 3.2, 3.3 and 6.10 of RDA, since the level of description considered in bibliographic references usually corresponds to the FRBR Manifestation (IFLA STUDY GROUP…, 2009) as approached in Graphic 23.

9.2.8.Part II – The correspondence between cited works and bibliographic references

Patino Diaz (2005) defines a bibliographic reference as the data that indicate to the reader whose **quote** he is reading and where to find it in its original version, i.e., the cited work. Masic (2013, p. 150) complements that “in scientific circles, the reference is the information that is necessary to the reader in identifying and finding **used sources**”. However, in addition to the “identifying function”, some bibliographic references providing hypertexts links or DOI hyperlinks also accomplish the “finding function” referred to by Masic (2013), although it is primarily up to the library catalogs. At this point we have to complement Patino Diaz’s statement (2005) with the information that bibliographic references not only suggest the identification of the sources of quotes, but also, the sources in which mentions are based on.

However, bibliographic references referring to works not cited along the text body and the reverse situation, i.e., mentions and/or quotations without a corresponding bibliographic reference in the bibliographic reference list, were identified in our sample. A brief comparison between data showed in Chart 9, Graphic 9, and Graphic 13 supports the understanding that the total 34,140 bibliographic references do not necessarily stand for cited works, since 1.2% of this total, on average, represents works not mentioned within the text body. In a reverse way, for 1.1% of in-text reference pointers referring to mentions and quotations, on average, we did not detect a corresponding bibliographic reference in the bibliographic reference list. In such cases, the provision of a single DOI hyperlink could dismiss the provision of all the remaining metadata addressed in a bibliographic reference referring to an article, which could represent a way of saving the author's time on writing bibliographic references and the reader time on identifying, seeking and retrieving such publication. Considering the identification of the cited works as the main function of bibliographic references, both the cases of bibliographic references with no correspondent mentions or quotations within the text body and the cases of mentions or quotations with no correspondent bibliographic references in the bibliographic references list, annul the core function of bibliographic references. In such cases the reader might be prevented from retrieving the cited content, configuring a contradiction to the 5 Ranganathan's Laws (ZABEL; RIMLAND, 2007) (RQ 5).

The provision of data that favors the online access to cited works may be considered a courtesy but, above all, an efficient way to facilitate the identification and access of cited works. However, it should not be ignored that reference styles rarely provide clear and enough instructions on this matter, favoring different interpretations for similar approaches within different reference styles (RQ 5).

Other aspect to be considered concerning the correspondence of bibliographic references and mentions and quotations is the order of the bibliographic reference list in articles adopting the citation-sequence system. Such correspondence should be directly derived from the order of appearance of their related in-text reference pointers in the text body and trusty reproduce it. In this context, the most coherent situation is when both in-text reference pointers and the numbers accompanying the related bibliographic references in the bibliographic references list are subject to identical formatting definitions. However, this is not what happens in all the articles in our sample. Indeed, as shown in Graphic 16 and Graphic 18, the in-text reference pointers and the characters determining numerical order in bibliographic reference lists are usually different from each other in terms of structuring and formatting, which may confuse the reader on matching them and on identifying cited works (RQ 5).

Focusing both on alphabetical and numerical bibliographic reference lists, on average, 18.41% of the sample did not show a proper ordering, as shown in Graphic 19. Starting from this, the relation between in-text reference pointers and bibliographic references and the reader's task of matching them to each other should be approached from two perspectives, according to the citation system used (i.e., author-date or citation-sequence). Considering articles adopting author-date citation system, an incorrect bibliographic reference list sorting may not prevent the reader from identifying the cited work' bibliographic reference, despite the need for seeking for it in the whole list. Instead, considering articles adopting citation-sequence citation system, the identification of the bibliographic reference matching a mention or quotation depends also on the proper correspondence between the in-text reference pointer and the respective numerical indicator in the bibliographic reference list (RQ 5).

9.2.9.Part II – In-text reference pointers structuring

In-text-reference pointers are also subject to the ravages of the multiplicity of reference styles and the superficiality of formatting instructions. We observed that there is no standardization at all neither on metadata addressed within in-text reference pointers referring to mentions and quotations, not on their formatting instructions, as addressed in Graphic 16 and 17. Considering the several ways in-text reference pointers appear in text bodies, a reader may confuse them with other elements. For instance, mathematical and chemical formulas frequently use superscript numerical characters which can easily be confused with in-text reference pointers, since both can share the same format (RQ 5). In addition, we noticed that, in most disciplines, there is not an established standard behavior on how to present in-text reference pointers even considering the choices between author-date and citation-sequence styles, as can be seen on Graphic 17, which are the main categories that guide the definition of the in-text reference pointer structure (RQ 5).

For instance, one of the long-indented quotations in our sample reproducing a passage from Jamie Dreier (KURTH, 2019), is finished by the following in-text reference pointer: “(2014a: 178; also: Korsgaard 2008; Gibbard 1990)”. This example suggests that there are misunderstandings regarding the concept of mentions, quotations, and the proper way of denoting them within works’ text bodies by using in-text reference pointers. Since a quotation is a literal and exact transcription of one or more passages from a cited work into a citing work, it is not possible to simultaneously cite more than one author per quoted passage, unless it is a mention or shared authorship cited work. This is a concerning fact that supports the claim that besides the epistemological issues previously pointed by Galvão (1998), such conceptual unclearness that hangs over Information Science also has a practical effect on the identification of certain elements in scientific works. This may represent a multifaceted problem. First, readers may have difficulties or even be prevented of identifying mentions and quotations within a text. Second such cases may represent research ethical issues, since the unclear or inappropriate identification of mentions and quotations may not provide due credits to the respective cited authors.

9.2.10.Part II – Page numbers provision in in-text reference pointers perspective

The provision of the pagination, within the in-text reference pointers referring to mentions and quotations, where the cited content can be found in the cited work is not uniform since most reference styles do not provide instructions for addressing this aspect (RQ 5). Indeed, we did not notice a common habit among the articles of our sample. For instance, as addressed in Graphic 17, the pagination metadata are sometimes provided when it is considered an optional element, i.e., in-text reference pointers referring to mentions, and it is not always provided when it is considered a mandatory element, i.e., in-text reference pointers referring to quotations, which make harder the task of seeking the original text in the cited work (RQ 5).

Page number is considered a mandatory descriptive element for most of the reference styles providing instructions on this matter. In fact, such data may help the user to locate the precisely quoted passage within cited works but, such functionality only is valid in cases where the reader consults precisely the same FRBR Manifestation used by the author of the citing work, i.e., the same work edition, publisher and/or version, since the pagination may change among different versions and embodiments of a same work – e.g., Web publications usually do not have the same pagination (if any is specified in relation to the same work published in the hardcopy version). This suggests that the provision of page numbers within in-text reference pointers and bibliographic references, is an issue to be thought of, facing the current and continuous changes in the way information is being registered across several information supports, be it analogical, i.e., printed or digital. For instance, according to the FRBR, which is now being considered as a starting point for establishing the trends of bibliographic description, consider that a single Work may assume several and totally different embodiments, without changing their main content (RQ 6).

9.2.11.Part II – The transcription of journals titles in bibliographic references

Our analysis showed that in most journals from Health Sciences, Life Sciences, and Physical Sciences, the titles of the journals in which cited articles were published, are provided in the bibliographic references in the abridged format, as addressed in Graphic 20. Probably, the main reason may be making bibliographic references shorter since some journal titles can be extensive. However, in such cases, it may be difficult for a reader to precisely interpret to which journal such an abridged title refers to. For instance, according to the ISO 4 standard, the journal title “European Physical Journal” should be abbreviated as, “Eur. Phys. J.”. Considering such abbreviation from the perspective of the reader, who is not supposed to know ISO 4 guidelines, the abbreviation “Phys” may be interpreted as “Physics”, “Physical” or “Physician”. Those misinterpretations may represent difficulties on identifying the correct original journal title, especially because the source from where such abbreviations are taken is not provided within articles but only within the instructions for authors, whenever provided (RQ 5).

Lastly, Graphic 21 evinces that 46.2% of journals adopting abridged journal titles, on average, do not provide the source in which such abbreviations should be based, not even to authors (RQ 5) and, within the remaining sample, we detected 7 different recommended sources for journal’s titles abbreviations. Besides favoring different abbreviations for the same journal title, such a range of sources go against the principles of standardization (RQ 4). Even considering the possibility of making bibliographic references shorter, abridging journal titles may be considered a non-sense practice in the era of the electronic universe, especially considering IFLA-LRM approaches on the “data and functionality required by end-users (and intermediaries working on behalf of end-users) to meet their information needs” (RIVA; LE BŒUF; ŽUMER, 2017).

9.2.12. Part II – The use of reference managers for managing bibliographic metadata

According to data shown on Chart 11 authors are usually recommended by publishers to use reference managers for dealing with bibliographic metadata, like Endnote or Mendeley. In theory, reference managers can solve standardization problems within bibliographic references. In practice, using such tools (should) demand careful monitoring of reference styles and, whenever an update or amendment is identified, such data should be immediately introduced within the reference manager’s stylesheets. It should be questioned whether publishers would have such a huge staff to appropriately perform all the editorial tasks plus such additional monitoring (RQ 5).

Yet, some issues on bibliographic metadata description may demand specific expertise in Information Science issues. The issue at stake is to what extent computer tools are replacing such human skills. For instance, considering our sample, Endnote and Mendeley were the reference styles most recommended by publishers across disciplines, according to data shown on Chart 11. However, we did not check the usability of these tools against the accuracy of the bibliographic references provided by them, something we need to claim whether such reference managers are appropriate to handle all the situations or whether we still need the expert human interventions in specific cases. Such issues are encouraged to be approached in further studies.

It should be mentioned that, in some publisher’s webpages, the recommendation to use reference managers is more evident than the use of the adopted reference style itself. This may give a false impression that, by providing a file containing bibliographic metadata compatible with – i.e., readable by – any reference manager or even that by recommending the use of such tools, publishers would become free from providing authors with clear and accurate standardization guidelines. On the other hand, authors may have the false impression that, by using such reference managers recommended by publishers, they are free from checking the bibliographic references provided by the machine, i.e., the reference manager. These situations may bring to the following issues:

a) the reference manager's stylesheet may not be updated according to the last version of the reference style adopted by a particular journal (RQ 5);
b) authors considering different reference managers recommended by a publisher may get different bibliographic references, if both tools are not updated and configured under the same parameters, which is in contrast to the standardization principles (RQ 5);
c) since the excess of information may be as dangerous as its lack, authors may be confused on choosing a reference manager, facing multiple available alternatives;
d) authors may be convinced that, because reference managers proceed the structuring of bibliographic references, they are free from checking the reference manager work or from learning at least the basic concepts on bibliographic metadata normalization (RQ 5).

Generally speaking, we got the impression that publishers commit themselves more fully to providing tools for the automatic writing of bibliographic references than to offering instructional resources on such issues (RQ 5), what seems to confirm Sweetland's (1989) statements that the lack of training in the norms and purposes of the bibliographic citation (RQ 4).

9.2.13.Part II – Exporting citations tools within publisher's webpages

Chart 12 shows that publishers may provide tools for exporting bibliographic metadata of the articles they publish within their journal's webpages. However, different output formats are presented as similar. For instance, some publishers provide bibliographic metadata both in structured formats, i.e., RIS or another machine-readable file, which demands a specific reference manager for decoding them like Endnote, and in textual human-readable formats, for which usually there is no indication of which reference style it is formatted to. This reinforces the understandings regarding the conceptual problems in the bibliographic universe (RQ 5) and shows that publishers could be more careful on bibliographic metadata matters (RQ 4). When providing bibliographic references in textual format, publishers usually do not mention which standard or reference style's guidelines was considered in its formatting. Since authors may simply copy and paste such bibliographic references into their own bibliographic references lists, by collecting such data from several publishers, they might have bibliographic references formatted under different criteria i.e., different reference styles' guidelines, mixed in the same bibliographic reference list (RQ 5).

9.2.14. Part II – Some FRBR approaches

Bibliographic catalogs, like those used in libraries, and bibliographic databases are complementary to the accomplishment of the core function of bibliographic references, since they provide at least the basic bibliographic metadata to allow the proper identification of a cited work. Being one of the core tools considered in the current and ongoing revision of the trends of representative, we traced a parallel between bibliographic references, the related in-text reference pointers, and the Functional Requirements for Bibliographic Description (FRBR) Entities concepts, namely FRBR Work, FRBR Manifestation, FRBR Expression, and FRBR Item.

It should be mentioned that FRBR and the related conceptual models previously developed by IFLA (i.e., FRAD and FRSAD) were consolidated into the IFLA Library Reference Model (IFLA LRM) in 2017 (RIVA; LE BŒUF; ŽUMER, 2017). However, we considered the FRBR concepts in this discussion first, because FRBR marked the beginning of the revision of the representative description and, simultaneously, boosted the distancing between descriptive representation facets, i.e., cataloging and referencing.

In the first approach, we considered the correspondence between the level of description observed in bibliographic references and FRBR Entities (Graphic 23). The results showed that the metadata set provided by bibliographic references usually corresponds to the FRBR Manifestation (99.35% of the cases, on average), and FRBR Expressions (0.65% of the cases, on average). Since a single FRBR Expression can be embodied in different FRBR Manifestations, the metadata specified in bibliographic references considering FRBR Manifestation level of description may limit the reader's search possibilities and, consequently, reduce the chance of accessing such content

(FRBR Expression) regardless the format it may have been published. Also, bibliographic catalogs tend to describe publications according to the FRBR Expression level first, and then to complement the record with data concerning the formats of such publication (i.e., the FRBR Manifestation level). Thus, bibliographic references are not necessarily expected to provide access to the publications they represent. It is true that, by providing a URL or a DOI number or a hyperlink, the bibliographic reference is, in fact providing the access to the represented publication but, this is not a mandatory descriptive element, indeed. In other words, whenever the reader does not have the perception that the core access points to a particular content refer to the FRBR Expression level instead of FRBR Manifestation level, he might not succeed in seeking a particular publication represented by a bibliographic reference (RQ 6).

As well as bibliographic catalogs complement bibliographic references' functions, the fulfillment of in-text reference pointers functions, i.e., the identification of a cited work within a text body, is directly dependent on its proper matching with the correspondent bibliographic reference referencing the cited work. Sometimes, as shown in Graphic 13, some in-text reference pointers associated to mentions or quotations do not match with the correspondent bibliographic reference and, therefore, the relation between them does not remain explicit.

As shown in Graphic 23, the bibliographic metadata described in the in-text reference pointers and the metadata defined in their respective bibliographic references usually do not match the same FRBR level of description. For instance, sometimes the metadata described in the in-text reference pointers refer to the FRBR Expression level of the cited works and may not be helpful to a reader in finding the cited excerpt within the cited work (RQ 6).

9.2.15.Part II – Do bibliographic metadata facilitate access to scientific information?

The rate of bibliographic references whose metadata was not enough to permit the clear identification of the work they reference is a matter of concern which suggests that reference styles usually do not provide clear and comprehensive instructions on how to present bibliographic metadata. Consequently, such guidelines are not properly understood by the authors, who end up providing bibliographic metadata in the way they believe it should be the most proper one, which is usually not. In this perspective, publishers' passive stance on the issues concerning citing and referencing description and normalization, confer them a significant role in compounding the counterproductive scenario on bibliographic normalization matters, which might at least be less problematic from the authors and readers' perspective.

All the previously discussed items support the understanding that, as shown in Graphic 25, 69.7% of in-text reference pointers regarding quotations and the related bibliographic references, on average, do not provide the essential metadata to easily identify the cited works and precisely locate the quoted passages within them. Such a portion of bibliographic references is not being effective in the accomplishment of their functions (RQ 5 and RQ 6).

Álvarez De Toledo (2012) states that scientific styles are methods of writing, structuring, representing, and organizing scientific contents, including mentions, quotations, and bibliographic references. Therefore, citing and referencing habits depend on the guidelines of the adopted scientific style, i.e., there is no universal habit in this scenery. Taylor (2006) states that the task of cataloging, primarily, is to develop and apply standards to create bibliographic records that describe and provide access to information packages. Such statements reinforce the understanding that bibliographic catalogs complement the functions of bibliographic references as information access facilitators. Substituting *cataloguing* in Taylor's notion with *bibliographic reference standardization* makes his statement still valid. The principle of user's convenience assumes that catalogers can objectively determine the user's needs and will know how to customize bibliographic records to meet these needs (HOFFMAN, 2009). However, it should be noted that multidisciplinary is becoming more and more a necessity and not only an added value (MARTINS, 2007). Facing this, it can be assumed that, in contrast to cataloging, citing and referencing matters should not consider the *local user's* but the *global user's* concepts and needs. For instance, there can be no assurance that a medical work will not support a Social Science work. This weakens the statement that the existence of multiple reference styles is needed to fulfill the specific needs of each discipline. Indeed, the multidisciplinary point of view invalidate the local user's point of view and either enlarge or enhances the target audience of a particular work to the whole scientific community. Second, in this context, the users and their real needs become unclear. Within the bibliographic metadata standardization domain, there are no local users and, therefore, the customization of bibliographic references loses its sense. In such a context, questions raised by Hoffman (2009) referring to the cataloging universe, i.e., “how can local users’ needs be met?” or “who is responsible for meeting users’ needs in cataloging?” and “what is the “right” way(s) for cataloging to help users and ensure equitable access to materials?”, do not apply to citing and referencing world (RQ 5). So, the widely adopted reference styles, such as Vancouver and Chicago, would suffice the scientific community needs and expectations on bibliographic matters. Bibliographic references do not address problems and needs but rather provide metadata that allows one to clearly identify a particular work and to seek it within bibliographic catalogs.

9.3.Part III – An approach on ontologies for treating bibliographic metadata

This session addresses the discussions on the analysis of data presented mainly in session 7.3.1 up to session 7.3.4, which refer to the findings concerning the observations on the structure of bibliographic references, considering the most used bibliographic elements used in their compositions, which resulted in the “starred metadata set” and, the analysis of the suitability of the FaBiO Ontology on describing bibliographic elements, mainly those comprised in the “starred metadata set”. This session aims, primarily, to answer RQs 7 up to RQ 9, as follows:

RQ 7 Which are the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations and bibliographic references, considering different types of cited works?

RQ 8 Is there a common metadata set used across the disciplines for describing cited works within bibliographic references?

RQ 9 Do SPAR Ontologies comply with bibliographic elements composing bibliographic references, particularly the “starred metadata set”?

9.3.1. Part III – The citing habits in scientific works: how mentions and quotations connect with bibliographic references lists

Citing and referencing are two complementary research tools that coordinate across the scientific text bodies. Specifically speaking on mentions and quotations, they represent the external content supporting a particular scientific discussion[45]. However, such content only has a meaning when the original cited source is known. While the bibliographic references provide the identification of the external works whose content were somehow mentioned or quoted in the text body, the in-text reference pointers referring to such mentions and quotations link the cited content in the text to the respective bibliographic reference in the bibliographic reference list, as a coordinated and complementary work.

The format in which the in-text reference pointers are provided in the text is usually defined by the citation system adopted. In general, articles adopting the citation-sequence system usually attribute growing numbers (superscript or not), to mentions and quotations according to the order in which they appear in the text. Such numbers should correspond to the numerically and ascending sorted bibliographic reference list, usually placed at the end of the cited excerpts. As for the articles adopting author-data system, in-text reference pointers referring to mentions and quotations usually provide at least the last name of the first author and the year of publication of the cited work.

The approaches of reference styles on citation systems and the procedures for coordinating in-text reference pointers, mentions, quotations and bibliographic references are usually more superficial than those approaching guidelines for writing and sorting bibliographic references in the bibliographic references list (which actually are a little bit far from being enough, as suggested along the previous discussions). As a possible consequence, we noticed issues which may represent obstacles for the fluid connection between the text body and the bibliographic references list and the external publications supporting a particular discussion.

Graphics 1 up to 25 show the most evident situations in which citing and referencing data could be provided in a clearer and/or most appropriate way and evince that

[45] In this context, supporting are being considered as the content which substantiate an argument, since mentions and quotations also may be considered to argue against a specific ideology or concept.

Information Science, publishers and all the agents involved with the production and management of products of information still have a long path to walk through, both on standardization matters and on the effectiveness of reference data provided in text bodies. For instance, Graphic 10 report a portion of Social Science articles in which we observed the simultaneous adoption of both the author-date and the citation-sequence systems, while Graphic 16 report the variation in which in-text reference pointers are being considered in text bodies. At this point we highlight the half of the sample representing Earth and Planetary Sciences Journals in which in-text reference pointers referring to mentions and quotations provide the first author's names, while the main access point for the respective bibliographic references in the bibliographic references list is the cited works author's last name. This would not represent a big issue, if all those in-text pointers were hypertextually linked to the respective bibliographic references but, as can be seen on Graphic 14, this is not the case. For instance, still considering the Earth and Planetary Sciences Journals, in which this situation was more evident, we noticed that 5% of the articles providing the first author's names in in-text reference pointers are hypertextually linked to the respective bibliographic reference.

Hypertext hyperlinks are helpful tools to evidence the intrinsic connections between the elements mentioned in the text body (mentions and quotations included), and the correspondent related content (like the bibliographic references). Around 49% of articles in our sample provide such feature (Graphic 14). However, from this total, only 15% provide round hypertext hyperlinks linking in-text reference pointers and bibliographic references (Graphic 15). In addition, we noted other issues at this point of the analysis. The bibliographic reference referring to a particular publication is (or should be) unique in a bibliographic reference list and, therefore, all in-text reference pointers that refer to mentions or quotations referring to the content of the same publication should be linked to a single bibliographic reference. The reverse way is not true, since a single bibliographic reference may be linked to several in-text reference pointers referring to it along the text body. So, by clicking in the bibliographic reference, the reader can be sent to any point of the text containing an in-text reference pointer referring to the clicked bibliographic reference which not necessarily will correspond to the exact point of the text which the reader was consulting when he first clicked in the in-text reference pointer which sent him to the bibliographic reference list. It would be helpful if such discrepancies could be corrected within scientific articles because, after all, such functionalities are kind courtesies from publishers to readers but that can become obstacles to the fluid reading, if they do not work properly.

We also observed that in contradiction with the guidelines of most of the reference styles, or at least those explicitly providing any instruction on this matter, we detected a few portions of mentions in the text bodies of articles composing our sample whose in-text reference pointers provided the page number where the mentioned content could be found in the cited work. Although this is not exactly a problem (sometimes it can be convenient for the reader to have this information actually), such provision does not agree with most of the reference styles guidelines so, this is one of the cases in which “it is not too bad to be against the rules”. The same cannot be said about quotations. Although the retrieval of the cited work is not among the primordial functions expected to be performed by bibliographic references, it is also true that the descriptive elements provided by them support the information seeking tasks. Graphic 17 evinces the percentual of in-text reference pointers referring to quotations which do not provide data concerning the page where the quoted passage can be found in the cited work. It should be highlighted that in 22% of the subject areas, all quotations fit into this scenario, although the provision of data concerning pagination in the cited work in in-text reference pointers referring to quotations is mandatory, according to most of the reference styles. Since quotations correspond to the literally copy of an excerpt of an external work, the pagination becomes more than a simple convenience for the reader in this context, mainly considering that the quotation refers to a specific part of a whole work, and not to its entirety, so, the provision of data concerning the pagination where the quoted passage was originally published in the cited work would be appreciated and required to the identification of a particular excerpt within a publication.

It has been emphasized throughout this discussion the relation between the external contents supporting a discussion, represented by mentions and quotations and, the respective bibliographic references in the bibliographic references list. This presupposes that for each bibliographic reference included in the bibliographic references list, there should be at least one mention or quotation referring to the corresponding publication in the text body. This proves not to happen with all articles composing our sample. For instance, Graphic 13 approaches the percentual of bibliographic references included in the bibliographic references list, for which the represented publication is not mentioned in the text body and, mentions and quotations for which there was not find a correspondent bibliographic reference in the bibliographic reference list. Situations like these broke the link connecting bibliographic references lists to the text.

In addition to there being correspondence between the text and the bibliographic references lists, it is important to provide an appropriate bibliographic references assortment in bibliographic references lists. However, data on Graphic 19 show a negative scenery which highlights that 12% of the bibliographic references lists considering citation-sequence adopting articles and 10% of the bibliographic references lists considering the author-date citation systems articles do not provide a proper bibliographic references assortment. Have the reader not realized that the assortment of the bibliographic reference list is incorrect (be it numerical or alphabetical) the cited work can be prevented from being identified.

All those arguments were discussed to complement the answer to the RQ 7. (Which are the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations, and bibliographic references, considering different types of cited works?). As already mentioned, in-text reference pointers are the connection channel between mentions and quotations in the text body and the bibliographic references included in the bibliographic references lists. The elements composing in-text reference pointers referring to mentions considering articles adopting author-date citation system are, basically, author's surname and the year of the publication of the cited work. In text pointers referring to quotations are added of the page number where the quoted excerpt is contained in the cited work. As for the articles adopting the citation-sequence system, in-text reference pointers referring to mentions are usually represented by cardinal numbers corresponding to the numbers attributed to the bibliographic references in the bibliographic references lists which describes the cited work. However, in 51% of the subject areas, we detected bibliographic references in which the link between in-text reference pointers and bibliographic references were explicitly clear, i.e., could be identified without complementary consults to secondary reference pointers, interpretations on improper bibliographic references sorting or elements provided by in-text reference pointers not corresponding to the main access point to the bibliographic reference. This suggests, first, that something went wrong in the publishing process since the failures previously detected were not corrected previously to the publication of the article. Second, that reference styles seem not to provide clear and complete guidelines concerning the presentation of mentions and quotations and their links to the bibliographic references (or authors are completely ignoring them). Third, the findings of this specific part of the study evinces that there is a lot to be done on citing field to become information clearer.

9.3.2.Part III - Bibliographic references metadata: a brief consideration

Data shown in Chart 12 suggests that articles are the most used channel to communicate scientific findings. In fact, the capillarity of information has a more intense and faster flow in articles, compared to other communication channels such as books, which are traditionally recognized as sources of theoretical information, primarily. On the other hand, books were observed among the three most cited types of publications in all disciplines considered in our sample and this supports and reinforces the understanding that Science is based on theoretical foundation, usually approached in books whose bias is applied to the dissemination of the most recent investigations and findings, usually published in the form of articles. Therefore, the main form of improving knowledge is the cyclical fusion between theory and practice.

Bibliographic references, combined with citation data, play some important roles within the publications in which they are contained (regardless of whether scientific or not) and in the scientific universe at all. First, they represent a way to report the publications containing the contents which supports a particular discussion. Second, they represent a way to clearly differentiate the self-authored content from the external content in a text body, i.e., the mentioned and/or quoted content. Third, they are a way to compile references to other publications on similar or related subjects, which allow them to be considered as a kind of bibliography[46], that is to say, a source of reference data on a

[46] In the context of this study, bibliographic reference refers to the brief description of work which has effectively been mentioned in the text body. As for bibliography is a list of publications with specific common characteristics (e.g., thematic axis, authorship) not cited in the text body but recommended as a supplementary content to a particular context. A more detailed definition of both terms is available from the https://zenodo.org/record/3996578#.YHhUQehKjzY.

particular subject, which may represent a significant way to save the researcher's (or reader's) time in the task of seeking and identifying information which complements each other, as an accomplishment to the predicted fourth Ranganathan's Law: Save the time of the reader (ZABEL; RIMLAND, 2007). Fourth, the joint work developed by bibliographic references, mentions and quotations is the protagonist element that materializes and makes the citation network more evident within scientific publications.

It is evident that citing and referencing data play a very important role within scientific communication. Science is a body of systemized knowledge built based on empiricism. Starting from this assumption, it becomes easier to understand the huge variability in the types of publications cited by the articles composing our sample. As can be seen on Chart 13 and Graphic 3, we found 36 different types of publications within disciplines. Such variety suggests and reveals some citing habits across disciplines. For instance, we noticed a considerable portion of bibliographic references for which we were not able to identify which type of publication it referred to (red cells in Chart 13), considering the data provided in the bibliographic references. This suggests that reference styles are not clear or do not provide enough instructions on how to describe certain types of publications or that authors could be more careful when writing bibliographic references or that publishers could carry out a more accurate revision on citing and referencing data prior to the release of any publication (or a mixture of all of these). In either cases, treating citing and referencing metadata available from publishers' repositories under the semantic perspective could be considered a valuable way of achieving better bibliographic descriptions overall. Better bibliographic references writing = less time spent on identifying cited works and corresponding them to citing works = strengthening connections between publications approaching related contents = strengthening the links which compose the citation networks.

Another highlight concerning Chart 13 is that some disciplines, with emphasis on the humanities, cites a notable huge variety of publications. This suggests first, that the discussions on such disciplines demand more comprehensive approaches and, second, that reference styles adopted by such disciplines should provide more extensive guidelines for describing citing and referencing data i.e., they should provide instructions on how to describe a greater variety of publications, in relation to other disciplines such as Dentistry, in which discussions are more concentrated, since bibliographic references refer to three types of publications. 55% of cited types of publications was considered among the 7 most cited ones across disciplines. At the same time, in theory, this makes the reference styles more concise, it also does not exempt regulatory bodies i.e., the International Standardization and Normalization (ISO) and other institutions which are responsible for the content of any reference style, from the responsibility of providing instructions on the description of all types of citable publications, after all, it is not possible to precisely preview the types of publications which will be cited by a particular work before its effective publication. And in this context, it should be considered that although citing certain types of publications is improbable, it is also not impossible. For instance, in Social Science discipline, we found articles whose bibliographic references referred to Lithography and Engravings. In any of the reference styles we examined, we observed instructions on how to properly describe such types of publications. Scenery like that contributes to the increasing of the rates of unidentifiable bibliographic references as those represented by the red cells in Chart 13 and, as a consequence, can make it difficult (or impossible in some cases), to identify the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations and bibliographic references and here, a cascade effect is triggered: if there are difficulties since the identification of these elements, probably there will not be good expectations about the uniformity of their representation within the articles' text bodies.

Since the variation of types of publications cited by each discipline has a meaning, better bibliographic references should be understood as those which allow the readers to precisely identify a particular cited publication from the metadata they provide. Consequently, they would contribute effectively to the intellectual capital turnover and to the citation networks improvement. This by itself justifies every effort to make bibliographic description more assertive.

Chart 15 is the main element of the third part of this study, from the perspective of the research question 6 (RQ 6), which refers to the basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations, and bibliographic references, considering different types of cited works. First, it is important to clarify that we considered book chapters, technical report chapters and e-books chapters as subdivisions of the main publications they are published in, i.e., books, technical reports and e-books. This subdivision was considered on Chart 15 (and not in the previous considerations) because in this point of the study it is important to identify specific descriptive elements included in the bibliographic references referring to specific parts of such publications.

The analysis of Chart 15 evinces the deficit in normalization field concerning bibliographic references. For instance, considering those bibliographic references referring to articles, we noticed that 71.59% of them provide the title of the journal which has published the cited article in the abridged format. The remaining portion of the sample provide the cited journal title in full. Although being an evidence of lack of standardization on the ways which bibliographic metadata are presented in bibliographic references, such situation not necessarily would prevent the reader to identify, seek and retrieve a particular cited work, however, it should be considered that some search engines (including the bibliographic catalogs), may not recognize the abbreviations of journal titles. Still, it should be considered that there are several sources for defining journals titles abbreviations. The NLM Catalog[47], the Web of Science Journal Title Abbreviations[48], The CAS Source Index (CASSI)[49], The List of Title Word Abbreviations (LTWA)[50] and even the ISO 4:1997 – Information and documentation — Rules for the abbreviation of title words and titles of publications (which also serves as the basis for the establishment of title word abbreviations by the ISSN Network) are some examples of journal titles abbreviation sources. The big issue is that the abbreviation for a particular journal title may diverge considering different sources guidelines, as exemplified in Chart 17.

---

[47] Available from https://www.ncbi.nlm.nih.gov/nlmcatalog/journals/
[48] Available from https://images.webofknowledge.com/images/help/WOS/A_abrvjt.html
[49] Available from https://cassi.cas.org/search.jsp
[50] Available from https://www.issn.org/services/online-services/access-to-the-ltwa/

Chart 17 - The abbreviation of the Journal Anales de la Real Academia Nacional de Farmacia, according to 4 different sources of journal titles abbreviations

	Journal abbreviation source			
	LTWA	**NLM Catalog**	**Web of Science Journal Titles Abbreviations**	**CASSI**
Full journal title	Anales de la Real Academia Nacional de Farmacia	Anales de la Real Academia Nacional de Farmacia	ANALES DE LA REAL ACADEMIA NACIONAL DE FARMACIA	Anales de la Real Academia Nacional de Farmacia
Abbreviation	an. r. acad. nac. farm.	Anal. Real Acad. Nal. Farm.	AN REAL ACAD NAC F	An. R. Acad. Nac. Farm.

By the data observed in Chart 17 it is demonstrated that a single journal title may assume different abbreviated formats, according to the adopted source of journals titles abbreviation. This may have negative consequences both for the precise identification of the referred journal and, consequently, for its retrieval. Having a single and universal source of journal abbreviation, like the ISO 4, would be effective to mitigate such negative aspects in information description.

However, more than having a unique source of journal abbreviation, such source should provide a clear and comprehensive coverage of the citable journal titles. For instance, according to LTWA, words composing journal titles starting with the prefix “Nacion-“ should be abbreviated as “nac.” in “mul” languages. From this orientation, we understand that the words “nacional” (in Portuguese and Spanish), and “nacionalidade” (in Portuguese) should be abbreviated in the same way, according to LTWA guidelines. Other examples are the prefixes annual- (for Portuguese, Spanish and Romanian words) and anuar- (for Romanian and Spanish words) which should be abbreviated as “anu.”.

Therefore, to ensure the correct interpretation of an abbreviation, including from automated resources and search tools perspectives, it should be accompanied by the indication of the source in which it was based on. However, we noticed this is not a reality in scientific universe, since providing or indicating the source in which authors should base the abbreviations for the titles of the journals referred in their bibliographic references lists is not a unanimous practice among publishers. We also did not observe any information in the articles concerning the used source for abbreviating titles, for readers orientation.

Fluid communication demands the sharing of the same language between the sender and the receiver of a message. The same principle applies within the scope of the bibliographic universe and, in the impossibility of sharing a common metadata language, it is essential to different metadata processing tools (i.e., the reference managers and publisher's databases or webpages), to use an exchange protocol that allows the conversion of metadata among them.

Chart 14 also revealed that the titles of the cited works are not always provided by bibliographic references (column 11). It is understandable that, in some cases, e.g., bibliographic references referring to articles, the title of the article itself is not a mandatory element for allowing its identification and retrieval but, on the other hand, it is, in fact, a mandatory element for identifying a cited article from the data provided by the bibliographic reference. However, what should not be disregarded, is that one of the main purposes of bibliographic references is to allow the proper identification of the cited work, as for providing its retrieval can be considered as a courtesy from authors and publishers to readers. In this context, considering bibliographic references referring to articles, we observed that data like the number of the volume and the issue in which the cited article was published are also omitted in some cases. Inevitable to say that the old purpose of using abbreviations for describe information description, which was saving space in the catalog cards, is not totally applicable nowadays in the current juncture of electronic sources and resources, to which such space limitations are, at certain point, circumventable. If using abbreviations and omitting some specific data in bibliographic references are necessary or convenient from diagramming and editing issues, on the one hand, such habits may pose problems to the readers, from whom it cannot be expected to be experts in the interpretation of "librarianshipely" coded data.

Concerning the uniformity of the data set provided by bibliographic references referring to specific types of publications overall, we can notice that in most cases, there is a relative (i.e., poor) uniformity. That is to say that the metadata composing bibliographic references referring to the same type of publications (explicitly shown on Chart 14) vary across disciplines. On the other hand, there are some particularities that cannot be dismissed. For instance, the date of citation (column 26 of Chart 14), which represents the date in which the citing author consulted a particular electronic source available online), is not always provided in bibliographic references. Such data is important to readers, who are warned about the occasion in which a certain content was available in an online source, considering that such types of publications may change or become unavailable with no previous advice and, therefore, may not be available anymore in the occasion of its further retrieval. So, more than some advice for readers, such data is a safeguard for authors, while indicating the specific chronological occasion in which a resource was available in the way it has been mentioned in the citing work. Even with such benefits, the date of citation is not always mentioned in bibliographic references as evinced in bibliographic references for webpages, purposefully mentioned here since it is the only type of publication that is necessarily available online and so, the date of citation is applicable to every single bibliographic reference referring to such type of source. Even so, our analysis showed that 51.31% of articles citing webpages provide the date of citation in at least one of the bibliographic references included in the bibliographic references list. The ideal rate in this context would be 100%.

Including, this is another point in the analysis worth mentioning: it was not uncommon to observe bibliographic references providing metadata in a non-uniform way, i.e., two (or more) bibliographic references referring to webpages in the same bibliographic reference list, one informing the date of citation and the other, not. We observed that this happens frequently but, we did not quantify such rates of occurrence, because it was not among the objectives of this study and so, further investigations on this matter are encouraged. But the fact here is that, first, authors not always provide comprehensive metadata concerning the publications they cite in the bibliographic references included in their works. Second, publishers seem not to care about such issues otherwise, we would not detect sceneries like these so often among scientific publications. Third, since we noticed that authors provide some specific metadata, i.e., the date of citation, only for part of their bibliographic references, we assume that they are aware of the importance of providing clear and comprehensive metadata for the publications they cite in the bibliographic references lists (even because authors are also readers) and if so, why do they not adopt the same level of description for describing all cited works? Is it for negligence? Or unfamiliarity with bibliographic standards?[51].

Yet, it cannot be ignored that although the provision of the location where the cited work can be found online (e.g. an URL, a DOI hyperlink or a DOI number), is not necessarily mandatory from references styles perspectives and, although the most commonly considered purpose of providing such data within bibliographic references is to facilitate the retrieval of the cited works, they also can be considered as an identification element which inclusive, can be “almost” as precise as a ISBN number, for example. “Almost”, because URLs are susceptible to change without previous warning, as mentioned above. And even if there was a previous warning, it would not change the metadata included in the previously published bibliographic references referring to a changed URL which would, from now on, point to an unavailable website, for example. So, surely, providing DOI hyperlinks and DOI numbers (which are immutable) in bibliographic references, is recommendable and preferable in comparison with providing “conventional” URLs. Even so, the point is that the analysis showed that providing URLs is not a common habit across disciplines at all (disregarding webpages and online databases). We also observed that in some bibliographic references providing URLs, such URLs, such locators did not point directly to the cited work. For instance, some URLs pointed to online bibliographic databases (like PubMed, for example), from which the cited work could be downloaded. The ideal situation is the one in which the URL point directly to the cited work. Contradictorily, even in those cases in which the URL did not point directly to the cited work, they still contribute to the identification of the cited work, besides facilitating its retrieval and, therefore, should be provided in bibliographic references, whenever possible or applicable.

Other point to consider is that the uniformity in the elements composing bibliographic references and in the order they are provided, added to the standardized punctuation between one and another element, is a combination that favors the breaking down of language barriers in the interpretation of bibliographic metadata considered in

[51] As we did not explore the answer for those questions in this study, investigating the answers for these questions suggests other research issues for further studies.

bibliographic references. For instance, what follows is the bibliographic reference of the History of the Russian Book Chamber, in Russian:

История Российской книжной палаты, 1917–1935 / Р. А. Айгистов [и др.]. – М.: Рос. кн. палата, 2006. – 447 с. – ISBN 5-901202-22-8.

For a non-Russian speaker who do not know in which bibliographic standard this bibliographic reference is based on, it can be a problem to identify which part of the bibliographic reference refers to the work title, the author and so on. Knowing which bibliographic style was used to write this bibliographic reference, it is possible to identify the metadata through the punctuation before and after each descriptive element, like the date between the first comma and the first slash is the date of birth and death of the author, and so on. So, even though Russian is not understood by the reader, he still can identify the bibliographic elements included in the bibliographic reference in such language. This reinforces the importance of informing the reader about the reference style in which the bibliographic references of the publications are formatted.

Chart 14 evinces the results of the data showed in Chart 13 (actually, it is a micro view of those data), concerning the basic set of metadata (descriptive elements) composing bibliographic references of articles across disciplines, that is, the “starred metadata set”. Basically, the blue cells show the bibliographic elements provided in at least one bibliographic reference included in the bibliographic references lists of the minimum of 50% of articles of each subject area composing subject categories. In other words, such bibliographic elements which we called “starred metadata set” represent the common metadata set used across the disciplines for describing cited works within bibliographic references, which answers RQ 8.

By analyzing these particular data (the “starred metadata set”), we can observe that important elements which would help readers identifying the cited works are usually dismissed by bibliographic references. For instance, the DOI is not among the “starred metadata set” for bibliographic references referring to articles, as ISBN is not among the “starred metadata set” for bibliographic references referring to books and book chapters. Important to say that such metadata, by themselves, are enough to identify and make unmistakable the publications they refer to. In general, the “starred metadata set” identified for some types of publications comprise the necessary descriptive elements to allow the identification of the publication being represented by the bibliographic references considering such metadata. This was the case for Articles, books, book chapters, Webpages, working papers / preprints, conference papers, grey literature, unpublished works, newspapers and magazines, patents, and book series. The overall “starred metadata set” for proceedings did not comprise the title of the proceedings in which the cited work was published nor the title of the conference in which the cited work was presented. However, such data were observed among the “starred metadata set” for proceedings in Health Sciences and Multidisciplinary (in which the “starred metadata set” includes Conference´s title) and Social Sciences, Life Sciences and Physical Sciences (in which the “starred metadata” includes proceeding’s title). Because of this particularity, the “starred metadata set” for proceedings were considered in the compliant “starred metadata set” list. The same logic was applied for Working papers / preprints, in which the “starred metadata set” for some subject categories (Health Sciences, Social Sciences Physical Sciences and Multidisciplinary) provide the URL where the cited work can be found, while Life Sciences provide a DOI number or a DOI URL.

However, the “starred metadata set” for the other types of publications evinced a concerning scenery, since the respective “starred metadata set” do not allow the identification of the cited works by themselves, for several reasons:

a) Technical reports: such documents generally deal with specific matters related to the performance of a particular company (or a group of them) in a predetermined period or describe the processes, progress and/or the results of a scientific research or a research problem. The point is that such publications may assume similar formats to those assumed by books or other types of publications (and so do the bibliographic references representing such types of publications). If the respective bibliographic references do not provide the information concerning the type of publication being represented or even the institution their contents refer to, it can be difficult to identify the cited work properly from the data provided by the bibliographic references, without consulting external sources.
b) Technical reports chapters: In only two subject areas we found bibliographic references referring to technical reports: Health Sciences and Social Sciences.

The considerations on this matter are similar to the points approached in the previous item, referring to technical reports. If not providing information concerning the type of publication being referenced, particularly speaking of technical reports, may represent issues on identifying a technical report, the same logic applies to the chapters contained in such publications. Also, some bibliographic references referring to a chapter composing a technical report, do not provide the title of the chapter being cited. The point here is that both titles (the chapter title and the technical report title) are crucial for the proper identification of the content being mentioned by the citing work.

c) Forthcoming articles; online databases; e-books and e-books chapters: Bibliographic references of our sample referring to forthcoming articles usually did not provide the title of the journal in which the cited articles are supposed to be published. In fact, such omission does not configure an impediment to further location of the article after its effective publication but, from the point of view of the bibliographic reference, it prevents the immediate identification of the cited article considering, above all, that the existence of two or more articles with the same title is unlikely, but not impossible. The same logic is applicable to bibliographic references referring to online databases which do not provide the URL where the cited content can be found. Even, the authorship of these type of publications is not always identifiable and, in those cases, the work's titles, which are usually considered as the main access point to the publication, may have homonyms. Bibliographic references referring to e-books and e-books chapters not providing publisher and URL, respectively, also fit the perspective which in all cases (Forthcoming articles, online databases, e-books and e-books chapters included), some additional information like the journal title or the URL where to find the cited work, is necessary to make the bibliographic reference unique and true to the cited work, what in fact, represents a paradox, considering the provision of a location metadata (i.e. the URL) is required to the bibliographic reference to accomplish to one of its main function which is to identify the cited work.

d) Software and applications; manual, guides, and toolkits; data sheets; standards and personal communications: For all these types of publications, we noted that bibliographic references did not provide any information concerning the nature of the document, previously called “general material designation” (GMD) in AACR2. The point here is that the description of non-conventional types of publications (i.e., publications in formats other than those most cited, like articles, books, technical reports, and others), requires a clear indication of the type of publication being cited to allow its immediate identification from the data provided in bibliographic references. For instance, bibliographic references referring to grey literature usually provide a short note like “master thesis” or “doctoral thesis”, which provide the reader the format of the cited work immediately and, therefore, becomes its identification easier.

So, as a response to RQ 8, we noticed a basic set of descriptive elements provided by in-text reference pointers regarding to mentions, quotations, and bibliographic references, considering different types of cited works, which are represented in Chart 15. This does not mean that such metadata set is the most complete and appropriate one for describing bibliographic resources in the format of bibliographic references. For instance, the previously introduced issues, concerning the cases of the provision of URLs from where the cited works can be download in the respective bibliographic references and the cases of using abbreviations in such descriptions, are examples of improvement points which can be implemented within bibliographic references to become them clearer and more effective. Mapping these and other similar issues and proposing a method for their implementation within scientific publications aiming to improve the scientific communication through bibliographic metadata should be among the Information Science’s concerns. Facing this scenery, the resources of artificial intelligence based in semantic methods tends to be increasingly applied in the production and management of bibliographic metadata. Certainly, this will demand the professional exchange and improvement, particularly from the professionals from Information Science and Computer Science.

9.3.3. Part III - The ontology concept in Information Science – a very brief consideration

We avoided dwelling on the discussion approaching epistemological aspects in this research, considering it does not match the objectives of this study. However, for introducing such issues, it worth mentioning that the term Ontology was originated in Philosophy and is now approached from different perspectives acorss different disciplines (FONSECA, 2007; GRUBER, 1993; ALMEIDA; SOUZA; FONSECA, 2011; LIMA, 2020).

Ontology is a basic description of things in the world (FONSECA, 2007), in an explicit specification of a conceptualization (GRUBER, 1993). The term is borrowed from Philosophy, where an ontology is a systematic account of existence (GRUBER, 1993), a “branch of metaphysics that concerns itself with what exists” (BLACKBURN, 1996, p. 269). As for Computer and Information Sciences, “an ontology refers to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words” (GUARINO, 1998, p. 4).

For the Information Science domains, it can be said that an ontology is a list of concepts or entities inside a specific domain which can be hierarchically structured by means of semantic relationships formally explicit, in a computerized environment (LIMA, 2020). Along the same line, Gruber (1993) points the basic components of ontologies: a) terms and definitions, which explicit and formalize the sense of concepts; b) classes which are organized under a hierarchical conceptual structure; c) relations, which represents the relationships between concepts and classes; d) axioms, which represent premises considered as truthful; e) instances, representing the specificity of data; f) attributes, which describe characteristics of concepts and instances. Lima (2020) complements that such components should have an explicit and formal representation for the ontology to be machine-readable, to control the ambiguity, the synonymy, the hierarchical and associative relationships, the complex relationships (rules and axioms) and are exclusively for technological use.

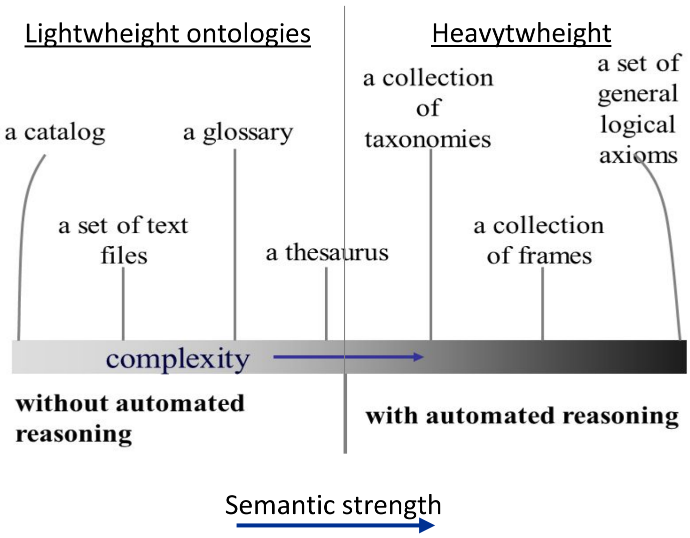
The term ontology, which have been gaining interest and acceptance in broader audiences (LASSILA; MCGUINESS, 2001), was used for the first time in Information Science in the end of the decade of 1990 (VICKERY, 1997; SOERGEL, 1999), especially by researchers in the field of knowledge organization, as a tool for representing knowledge with the express use of semantics, defined by Vickery (1997) as a systematic formalization of definitions of concepts, relationships and rules which captures the semantic content of a specific domain into a machine-readable format.

The concept of attributing terms capable of assigning meaning and context to bibliographic data is not exactly recent for Information Science. Previously to the advent of the Technology, libraries used to employ efforts on describing the contents of documents by using a documentary language with the objective to facilitate the memorization of information in archives, files, databases, and databanks (CHAUMIER, 1988). One of the tools used for such descriptions were the thesaurus, which lists the main concepts/terms in a particular domain and specifies relations between the concepts/terms using only a small number of relation types. This small set of relations may be adequate for information retrieval applications because the focus of a thesaurus is on indexing and searching, but it is not sufficient for more complex or intelligent applications that require knowledge-based inferencing and a detailed representation of domain knowledge (KHOO; NA, 2006). Other librarian technique used for providing an information subject seeking facility is the classification, which is referred by Hjørland (2002), as “basic to our very existence” and “fundamental in our lives” and, therefore, librarians and Information workers should not ignore it but, adopt techniques to meet today’s needs and complement that classificatory techniques are “one of the most powerful access tools that we possess”.

Lassila and Mcguiness (2001) refer to controlled vocabularies as one of the simplest notions of a possible ontology, and to glossaries as potential ontologies specifications and finally that thesauri provide some additional semantics in their relations between terms. Inclusively, Khoo and Na (2007, p. 183) point out that “one major difference between an ontology and a thesaurus is the richer set of relations used in an ontology”. Hjørland (2002) finally state that in Computer Science the term “ontologies” is very popular, and can be considered a modern development in classification research, which justifies the “growing interest in ontologies because of their potential for encoding knowledge in a way that allows computer programs and agent software to perform intelligent tasks on the Web (KHOO; NA, 2007). As for the professionals dealing with information organization, ontology experts can be considered as senior taxonomists or information architects – the evolution of each function would lead the experience in the field of ontology (POLLOCK, 2010).

From the Information Science and Computer Science viewpoints, the ontology is explored as an instrument of knowledge representation, usually, as a computing artifact (LIMA, 2020). This supports the understanding that information management migrated from an analogical and less complex term-based approach of information management to a broader and more complex concept-based approach in electronic environments, with an increasing semantic richness, as visually represented in Figure 19.

Figure 19 -Knowledge and organization systems spectrum



Source: adapted from Lassila and McGuiness (2001) and Smith and Welty (2001).

9.3.4.Part III - FaBiO: The FRBR-aligned Bibliographic Ontology

In general, the metadata set provided by bibliographic references allows the identification of the cited work, although some specific aspects (previously described) suggests that there is still a long way to go on this matter. For instance, machines and search tools can decode some elements composing bibliographic references, like DOI numbers and abbreviations. On the other hand, information producers (including authors) should keep in mind that the accomplishment of bibliographic references purposes requires these to be as much human-readable as possible (what is not necessarily applicable in the ways and formats in which bibliographic metadata are stored and processed by machines). In this context, an alternative for obtaining the best cost-benefit ratio from the perspective of Science progress aiming, primarily, to favor the scientific communication flow and to save researcher's time on describing information (i.e., writing bibliographic references), should be the use of semantic tools for treating bibliographic metadata at the very moment of their conception, i.e., at the publisher's research tools level.

In this context, our study analyzed the SPAR Ontologies, particularly the FaBiO Ontology, as a tool for describe bibliographic metadata in a semantic perspective. One of the first effort to provide Semantic Web description of the publishing domain was the Dublin Core Metadata Terms (DCTerms), followed by the Functional Requirements for Bibliographic records (FRBR), in which SPAR Ontologies are based (PERONI; SHOTTON, 2018, p. 120). However, the FRBR's entity-relationship modelling framework itself was enhanced (together with other IFLA's conceptual models), by the IFLA-LRM (RIVA; Le BŒU; ŽUMER, 2017). That is to say that the SPAR Ontologies themselves might demand a review to incorporate the new (and probably broader) approaches on information description proposed by IFLA-LRM.

Considering FaBiO Ontology as it was at the occasion of conducting this study, the first point to highlight is its structure and the correspondence to FRBR entities and RDA entities. FaBiO is a FRBR-aligned Bibliographic Ontology (PERONI; SHOTTON, 2012a), as for the "Resource Description and Access is a package of data elements, guidelines, and instructions for creating library and cultural heritage metadata that are well-formed according to international models" (RDA Toolkit, c2021c). Actually, RDA guidelines were applied to properly identify each descriptive element observed within bibliographic references in our sample.

Such correspondence (FaBiO, FRBR and RDA) is very opportune. The description of most of the bibliographic references composing our sample corresponds to the FRBR Manifestation level. What is to say that, basically, bibliographic references provide data concerning FRBR Work and Manifestation Entities, enriched by qualified categories and other statements which aggregate relationships among those and other descriptive elements. The set of such entities and their relationships combined result in the full bibliographic description. As for from FaBiO Ontology perspective, the combination of classes with the object properties, data properties and named individuals, results in the semantic approach of a particular publication, through which it is possible for machines to recognize descriptive elements and establish relations among them, what can be helpful not only for searching tools but also for the automatic construction of bibliographic references (i.e., those made by reference managers like Zotero, Mendeley, Endnote and others).

In the traditional descriptive representation perspective, particularly referring to bibliographic records and bibliographic references, the information contained in the publications and represented in bibliographic descriptions are usually approached from the FRBR Manifestation level. However, the work approach should be the focal point of the description, while the other attributes and complementary elements, like authorship and press data, should act like complements that individualize the work in focus, from other similar works. In the semantic approach this perspective is broadly expanded since every single descriptive element may be semantically connected with several other elements. For instance, a single person may author several different works which, by its turn, may assume several expressions and manifestations.

9.3.5.Part III – A FRBRized view of FaBiO Ontology structure

Previously, we briefly mentioned epistemological issues that permeate Information Science. Here we have the same question involving the concepts of entities in FaBiO Ontology and FRBR.

“The FRBR-aligned Bibliographic Ontology (FaBiO) is an ontology for describing entities that are published or potentially publishable [e.g., journal articles, conference papers, books], and that contain or are referred to by bibliographic references.” (PERONI; SHOTTON, 2019) “FaBiO entities are primarily textual publications such as books, magazines, newspapers and journals, and their content such as poems, conference papers and editorials. However, they also include blogs, web pages, datasets, computer algorithms, experimental protocols, formal specifications and vocabularies, legal records, governmental papers, technical and commercial reports, and similar publications, and also anthologies, catalogues and similar collections” (PERONI; SHOTTON, 2012b). That is to say that, from FaBiO Ontology’s perspective, entities represent the type of publication and the type of content contained therein. FaBiO’s documentation refers to entities, classes, and properties and, since there are no explicit statements containing a precise definition of each category of terms from FaBiO’s perspective, the users may understand that such categories refer to the same things when in fact they do not. FaBiO classes are structured according to the FRBR schema of Works, Expressions, Manifestations, and Items. Additional properties have been added to extends the FRBR data model by linking Works and Manifestations (fabio:hasManifestation and fabio:isManifestationOf), Works and Items (fabio:hasPortrayal and fabio:isPortrayedBy), and Expressions and Items (fabio:hasRepresentation and fabio:isRepresentedBy) (PERONI, 2018).

From the FRBR perspective, when someone refers to "a book", he is describing a physical object, with sheets of paper and other characteristics that are his own. For FRBR, this is an item. Also, when someone refers to a book, he may mean a “publication”, knowing its ISBN (International Standard Book Number), inclusive. For FRBR, it is a manifestation. However, if one thinks, "a book, which is a translation of ...", with a particular text in mind, in a specific language, for FRBR it is an expression (MORENO, 2006).

Entities and Attributes has clear definitions in FRBR domains. Entities are defined as the “key objects of interest to users of bibliographic data” (IFLA STUDY… 2009, p. 13), which includes other elements than those mentioned in FaBiO’s perspective, like authors, titles, and subtitles, among others. This suggests that the concept for entities in FRBR is more comprehensive than those considered in FaBiO Ontology. As for attributes are the set of characteristics defining each entity, which serve as “the means by which users formulate queries and interpret responses when seeking information about a particular entity” and “will not necessarily be exhibited by all instances of that particular entity type” (IFLA STUDY… 2009, p. 31). For instance, we will never find an authorship entity associated with an edition attribute.

Attributes of entities “generally fall into two broad categories. There are, on the one hand, attributes that are inherent in an entity, and on the other, those that are externally inputted. The first category includes not only physical characteristics (e.g., the physical medium and dimensions of an object) but also features that might be characterized as labeling information (e.g., statements appearing on the title page, cover, or container). The second category includes assigned identifiers for an entity (e.g., a thematic catalogue number for a musical composition), and contextual information (e.g., the political context in which a work was conceived). Attributes inherent in an entity can usually be determined by examining the entity itself; those that are imputed often require reference to an external source" (IFLA STUDY… 2009, p. 31).

Besides that, Entities in FRBR are subdivided into three groups: The first, comprises the "products of intellectual or artistic endeavor that are named or described in bibliographic records: Work, Expression, Manifestation and Item […]. The second group comprises the entities responsible for the intellectual or artistic content, the physical production and dissemination or the custodianship of such products: person and corporate body. The entities in the third group represent an additional set of entities that serve as the subjects of works. The group includes concept (an abstract notion or idea), object (a material thing), event (an action or occurrence), and place (a location)" (IFLA STUDY… 2009, p. 3).

9.3.6.Part III – Some pathways for FaBiO Ontology enhancements

Although being a FRBR-aligned Bibliographic Ontology, we noticed a different entities structure in FaBiO, in relation to FRBR. This is not exactly an issue, even because the objectives of such instruments are different. However, it would be productive if FaBiO provided clear statements containing the logic used in the subdivision of the terms that compose it and the appropriate definitions and applicability. Such a measure, in addition to clarifying users about the structural aspects of the Ontology, would also mitigate the chances of occurrences of divergences of epistemological nature with other bibliographic instruments with similar purposes. So, this suggests being a point for improving the FaBiO Ontology.

Focusing on the results of the analysis of data compiled in Chart 16, we considered the 4 categories of terms approached in FaBiO Ontology and checked the applicability of the terms contained therein to properly describe the bibliographic elements identified in the bibliographic references included in articles of our sample referring to the most cited types of publications, as shown in Chart 15. Although we identified the "starred metadata set" for each type of publication considering each subject category, we expanded the analysis at this point, beyond the "starred metadata set", in order to get a broader view of the FaBiO Ontology applicability on describing bibliographic elements considered in bibliographic references.

First, we noticed that the structure considered in FaBiO's terms categorization (i.e., Classes, Object properties, Data properties and Named individuals) is compatible with the logic used for defining Expressions according to RDA rules (also considering that both SPAR Ontologies and RDA are FRBR-aligned). That is to say that FaBiO Classes refer to the RDA's concept of "content type" according to which corresponds to "a categorization that reflects the fundamental form of communication in which the content is expressed and the human sense through which it is intended to be perceived." (RDA TOOLKIT, c2021b).

The FaBiO terms categorized under the "Named individuals" label has a common characteristic: all of them refer to the storage medium (digital and analog). Such contents relate to the concept of "carrier type" in RDA, under which it means "a categorization reflecting the format of the storage medium and housing of a carrier in combination with the type of intermediation device required to view, play, run, or otherwise access the content of a manifestation." (RDA TOOLKIT, c2021a). From this perspective, FaBiO Named individuals category could be expanded. For instance, the terms videodisc, microscope slide, object (referring to tridimensional objects), online resource and volume could be considered in such category. This suggests another enhancement for FaBiO Ontology.

Editors, translators, and other agents who contribute to a work, an expression or a manifestation, are considered as relationship elements for the entity "person", according to the RDA's rules. Such relationship elements also may be considered as authorized access points for persons according to the publication and to the context its bibliographic description is intended. However, the FaBiO's term "has creator" is the one that comes closest to proper describe such entities. This support the statement that FaBiO Ontology is not totally suitable for describing statements of responsibility others than the main work's creators, i.e., the authors. The inclusion of terms for expanding the coverage of the statements of responsibility, like translators, illustrators, editors, etc., is another aspect for improving the FaBiO Ontology.

We also identified descriptive elements within the bibliographic references of our sample (in addition to the previously mentioned case of responsibility statements), for which we did not found any FaBiO term suitable for their proper description (corresponding to the red cells in Chart 16). In this analysis, we associated the FaBiO' term “Has creator” with the descriptive element “Proceedings’ editor”, and the term “translator”, to the FaBiO term “Contributor”. The justification for this is that from the descriptive representation perspective, the editor can possibly be considered as one of the main access points for a publication, acting as an "alternative or secondary author". As for translators, they are usually considered as added entries, instead of main access points. Surely, those are adaptations to make both FaBiO’s terms suitable to the bibliographic context, considering our sample. However, ideally, there should be a single term referring to each entity possibly related to a publication. This represents another improvement suggested to be considered by FaBiO Ontology.

We also did not find a suitable FaBiO term for describing some specific notes observed as bibliographic elements in the bibliographic references composing our sample. In general, such notes refer to additional information concerning the cited work, e.g., a note concerning a forthcoming publication, the date and place in which a conference was held or a supplementary content note. Such data usually do not constitute “authorized access points” for a publication, like authors and titles for example, and, the omission of such data in bibliographic references not necessarily will prevent the identification of the referred publications. Even though, it cannot be ignored that such metadata facilitate the identification task, which is expected to be performed by bibliographic references. Thus, the inclusion of some specific notes in bibliographic references should be (and actually is) a common habit and, therefore, a similar behavior would be expected from the instruments for describing bibliographic resources, including SPAR Ontologies. This supports another suggestion as a point of improvement for FaBiO Ontology.

Different types of publications require different metadata for their proper description. For instance, the description of an article usually provides the title, the issue, and the volume of the journal in which such article was published, as for describing books, other bibliographic elements, as the book author the work’s title, the editor’s names and the place of publication are considered more appropriate metadata. Likewise, some types of metadata, like the translator, the edition number, the version, and the series title for example, are not always applicable to all types of publications, nor to all manifestations of the same work. However, we noticed that there is not a common habit on providing such metadata for describing certain types of publications in bibliographic references across subject categories. For instance, considering bibliographic references for articles, only in Social Sciences articles we noticed the provision of the issue number of the journal in which the cited article was published in more than 50% of the articles composing our sample, although such metadata may facilitate its identification within a journal and becomes an eventual seeking for the referenced article easier for the reader. Similar situation was perceived in bibliographic references referring to books. In Multidisciplinary articles, bibliographic references generally did not provide the place of publication of the cited work, in contrast with the articles from the other disciplines. In fact, contemporary end-user search systems, like Online Public Access Catalogs (OPACs) and bibliographic databases offer flexible search capabilities and expand the access points (a.k.a. index terms) of publications overall. Because of this, neither authors nor editors or even publishers can predict which of such access points readers would select to perform seeking using those search systems and, therefore, the omission of descriptive elements within bibliographic references can be considered, at certain point, as a way of limiting the seeking possibilities for the reader.

The results of the analyzes previously discussed configure the answer to the RQ 9, which refers to the suitability of SPAR Ontologies for describing (mainly) the “starred metadata set” composing the bibliographic references of our sample. Based on the findings of the study, it is possible to affirm that FaBiO Ontology is suitable to accomplish with the purposes of descriptive representation concerning bibliographic references. However, as was demonstrated in Charts 13 and 14, the variety of types of publications being cited in the scientific domain is huge (and growing) and this variety volume is directly proportional to the volume of data elements being required to properly describe each publication type. In this scenery, FaBiO Ontology still have an improvement path to get through to maintain and reinforce its suitability and to better serve to the informational needs of scientific community and the bibliographic metadata flow and exchange. Including, such improvements, in addition to potentially favor the massive adoption of SPAR Ontologies by publishers (and other scientific content producers) for describing bibliographic metadata, still may assigns semantic characteristics to bibliographic metadata, which would have a direct impact on strengthening the citation network. As a counterpart, the metadata exchange would be facilitated and, consequently, even the reference managers (e.g., EndNote, Zotero and Mendeley) could benefit from the possibility of providing more liable and consistent bibliographic references to their users, regardless the rules of the reference style under which the bibliographic references are formatted. Yet, by providing bibliographic metadata with semantic properties, authors probably will be exempted from filling in reference managers' fields with the bibliographic information of the publications mentioned in their works, since bibliographic metadata provided by publishes would be readable by reference managers. For instance, providing articles' metadata in their own headers or footnotes is an example of good practices which may contribute to facilitating referencing tasks to any authors eventually citing those works. However, those metadata should be manually inputted in the reference management applications. Had such metadata semantic characteristics, probably an algorithm would be enough to make such data gathering automatically. At the same time this would save researcher's time, still could contribute to the standardization of bibliographic metadata provided in bibliographic references, make bibliographic searches responses more efficient and, minimize the chances of errors caused in manual typing or misinterpretation of bibliographic metadata. Therefore, ontologies, and particularly the FaBiO Ontology, has a lot to contribute to the bibliographic metadata management.

Managing information has become an increasingly complex task as the information sources become more and more diversified and specific. It certainly demands adaptations not only in the instruments of information management but also the qualification of professionals to work in this constantly improving scenario. For instance, ontologies are relatively recent instruments in Librarianship, which has already shown the first signs of adaptation to the new information landscape, such as the publication of the FRBR and subsequent publications (e.g., The IFLA LRM), which improved the content approached in its preliminary version. Likewise, ontologies must increasingly be included in the scope of library activities, for the fortune of the scientific community.

10.FINAL CONSIDERATIONS

Information science has been carrying out discussions on the ways of representing information (ALVARENGA, 2003; BAPTISTA, 2007; CAPLAN, 2003; HOFFMAN, 2009). And this is comprehensible and necessary, since the world of information is getting more and more huge and complex. In this scenario, it is becoming difficult to individualize an FRBR item through a simple description, like has always been done in the early days of information representation. Consequently, the most specific is an information, the most complex is its description, and the most difficult is its retrieval, specially to people who are lay at research techniques.

This is an actual theme both for Computer Science and for Information Science, which consolidates the interdisciplinarity between these two subjects. And indeed, there will be necessary efforts from both areas for reaching a concrete solution to these metadata and standards and retrieving tools methods.

This study raises questions about whether the current citing and referencing practices meet user's needs. Indeed, this was one of the main reasons which substantiated the first steps on the revision of representative description, which culminated in the development of FRBR. Our study provides a first insight on these matters across several disciplines, but its outcomes should not be considered definitive, and more in-depth discussions on such matters should be carried out.

Our findings suggest that reference styles do not fully accomplish with their role of guiding authors and publishers on providing concise and well-structured bibliographic metadata within bibliographic references to allow the easy and accurate identification of the referenced works, especially when types of publications not usually cited are being considered, in addition to articles and books.

In response to RQ 1 concerning the claims pointed by Sweetland (1989), our study confirmed that even with the use of bibliographic tools like reference styles and reference manager softwares, several of the issues highlighted by Sweetland (1989), introduced by Chart 1, still hold today. For instance, the variety of reference styles and the errors which are not corrected before publication support the speculation that journal editors did not seem to spend a huge effort on revising bibliographic metadata within their articles.

As an answer to RQ 2, our data showed that the errors pointed out by Sweetland (1989) still hold today, plus other issues identified along the study. The way information is produced, stored, retrieved, used, and represented has changed since the time in which Sweetland developed his study. However, it seems that the way information is approached by descriptive representation from citing and referencing perspective has remained the same since then. Besides, nowadays, authors can count on technological resources to help dealing with citing and referencing metadata, i.e., the reference managers. However, consciously, or not, the increasing amount of reference styles contributes to multiple (and sometimes totally different) interpretations of similar guidelines, which ends up acting as a barrier to standardization of bibliographic metadata concerning citing and referencing matters. We also detected possible causes for the lack of standardization on citing and referencing, other than those claimed by Sweetland (1989), such as the use of shortcomings, customizations, and omissions of bibliographic elements within reference styles. In some cases, the importance of having clear citation apparatus seemed not being a primary concern by both the authors and the publishers. Besides, the training of authors in bibliographic reference writing and in-text mentions and quotations of passages of a cited works seemed poor. Since failures pointed by Sweetland (1989) were not overcome, but incremented, as highlighted in our study, we suggested that the revision by which descriptive representation is going through should also be extended to comprise the citing and referencing normalization domains.

In response to RQ 3, which refers to the impacts that may be expected by readers on retrieving information from citing and referencing metadata considering such a revision, a trend towards an asymmetry on the way information may be represented within bibliographic references and bibliographic catalog was detected. In particular, the adoption of FRBR principles in bibliographic catalogs may affect the way bibliographic records are stored since each catalog could follow its own approaches to describe a certain document (e.g., only the FRBR Manifestation level of the first edition) which may differ from the others.

Reference styles do not provide broad and clear coverage of all aspects concerning bibliographic metadata description in the form of bibliographic references. Consequently, a lack of standardization is evinced in several aspects of the scientific universe. In addition, not following any reference style or not following the adopted reference style properly, suggests problems in the editorial process quality. Issues regarding the presentation of bibliographic references head the list of the most frequent problems found in the scientific literature (GALVAN, et al. 2017, apud CARRI-ÓN *et al.*, 2017). Authors and the journals where they publish are the most affected entities by the errors in citing and referencing (RUIZ PÉREZ; DELGADO LÓPEZ-CÓZAR; JIMÉNEZ CONTRERAS, 2006) and this make such subject an object of investigations (OSCA-LLUCH; CIVERA MOLLÁ; PEÑARANDA ORTEGA, 2009).

Concerning RQ 4, we can consider that Sweetland's claims not only remains updated but also were incremented by other factors, some of them briefly introduced in the answer for RQ 2. As approached along the discussions supporting this study, changings towards the improvement on the way which bibliographic metadata is provided within bibliographic references depends on a joint effort between the Scientific community (here understood as the authors and the publishers – including all the agents involved with the editorial process), besides the Information Science and Computer Science community.

In response to the RQ 5, we noticed that the problems with citing and referencing metadata have increased since Sweetland's time. Publishers seem to consider mentions, quotations, and in-text reference pointers as separate elements when they are part of the same whole that complement each other. For instance, the adoption of technological resources, i.e., the reference managers, which intended to assist authors in dealing with such metadata, in fact, contribute to the non-accomplishment of standardization purposes, by considering different versions of the same reference style and by not clearly providing such information.

Concerning RQ 6, referring to the possible impact of the FRBR regarding descriptive representation on information retrieval from citing and referencing metadata perspective, the data we gathered suggest that the application of the FRBR and, possibly, the IFLA LRM principles and concepts within bibliographic catalogs may establish greater distancing between descriptive representation's facets, i.e. cataloging, citing and referencing. Such an outcome, supported by the other findings of our study, suggests and evinces the need of a revision of citing and referencing matters starting from the FRBR concepts. Unfortunately, citation errors continued to appear, as did an increasing number of complaints about them (SWEETLAND, 1989, p. 293). This seems to confirm the necessity, as Tillet (2009, p. 203) anticipated, to “have some adjustments to make in bibliographic descriptions to be more precise about what we are describing–what FRBR entity-level and what relationships to other works, expressions, manifestations, and items. These are things that could be built into the creation of descriptive metadata in future cataloging systems”. In addition, Tillet (2009, p. 203) adds, there is value in having “standardized citations for works and expressions (uniform titles for works and expressions), and we can see the potential for decreasing cataloging work and costs by linking the appropriate subject headings and classification numbers to those citations that, in turn, can be linked to records for the manifestations that embody those works and expressions”.

On this matter, Masic (2013, p. 150) suggests that the “basic rule when listing the sources used is that references must be accurate, complete and should be consistently applied”. However, since we do not know the way readers will seek the information, we need more and more elaborated and practical systems and, above all, with the most international recognition (PATINO-DIAZ, 2005, p. 16). Citing and referencing should, therefore, be thought and redesigned jointly with the descriptive representation revision, to assure a unique metadata language among the various instruments and contexts in which information is represented.

One last point to consider is the role of publishers in this scenery. Publishers represent one of the most important means for the scientific communications, based on a two-way relationship with authors. Publishers depend, at first, on the authors scientific production to have what to publish, while authors depend (not exclusively) on publishers to give publicity and visibility to their writings. This suggests that publishing is not only a way of making money, but it is also a mechanism to enhance the scientific communication. Starting from this, some features attributed to the scientific text can be considered as ways to boost and facilitate the flow of scientific communication, as it favors both the connection between citing and cited works (and, consequently, the citations network) and between readers/researchers and works of interest, as Ranganathan's laws presuppose.

As answer to RQ 7, we noticed that for articles adopting author-date sequence, in-text reference pointers usually provide the cited author’s surname followed by the year of publication of the cited work. As for articles adopting citation-sequence system, in-text reference pointers are usually composed by numbers (whose formatting showed great variation), that correspond to the numerical ascending ordination under which the bibliographic references list are usually assorted. We also noticed that in-text reference pointers referring to quotations usually provide the pagination from where the quoted passages were extracted in the cited works but, it is suggested that a consensus should be established concerning the purposes of providing such data within in-text reference pointers, since the analysis found that pagination is not provided in all cases it was expected to be and, the reverse way is also true. Despite these caveats, we found that yes, there is a basic set of descriptive elements provided by in-text reference pointers regarding to mentions, which are the author's surname followed by the year of publication of the cited works, for articles adopting author-date system, as for in-text reference pointers referring to quotations considered in articles also adopting the author-date system such metadata are added to the pagination of the quoted passage in the cited works. Considering articles adopting citation-sequence system, we can say that both mentions, and quotations are usually marked up in the text bodies using numbers, although the normalization on the ways such numbers are provided is another issue to be considered.

As answer to RQ 8, we mapped the descriptive elements provided by the 34140 bibliographic references considered the bibliographic references list of the 729 articles composing our sample. Such mapping evinced the most cited types of publications in each discipline and showed that articles and books led the rankings. For instance, such types of publications were among the three most cited ones in 96% of the observed subject areas. The exceptions were Computer Science and Medicine. The analysis also supported the identification of the standard set of metadata provided in bibliographic references across disciplines, considering the different types of cited works identified in the research sample, as detailed approached in Chart 15. However, it cannot be dismissed that such standard metadata set, that we called "starred metadata set", by itself, provides a precarious description of the cited works and this requires some efforts to improve the quality of bibliographic metadata provided in bibliographic references across the disciplines and the scientific communication channels, including the publisher's webpages which are the platforms where the publications are first made publicly available.

As answer to RQ 9, the analysis showed that FaBiO Ontology can be considered as an instrument for improving bibliographic descriptions, especially considering the facilities potentially provided by the technology on writing and managing bibliographic metadata. Particularly considering FaBiO Ontology, the results achieved in the analyzes demonstrate that it complies with the “starred bibliographic metadata set” and goes beyond, being compatible also for the description of most of the other entities and bibliographic elements identified in the bibliographic references composing the sample. However, some improvements are suggested to FaBiO Ontology. For instance, we did not identify any FaBiO’s classes suitable for the description of some bibliographic elements like the date and the place where a Conference was held or the responsible for a particular translation of a work. Since such types of metadata can be relevant, depending on the context they are being considered, it is recommended that FaBiO Ontology include appropriate classes for the description of such type of metadata and others, as pointed in Chart 16. Surely, FaBiO Ontology has a lot to contribute to the bibliographic universe as it is however, some improving could make it even better.

It is true that FRBR (in which FaBiO is based on) is intended to support the description of bibliographic objects in a broader perspective in comparison with bibliographic references. However, any instrument intended to describe bibliographic references based on the FRBR model, like FaBiO, is expected to cover all the bibliographic elements liable to be applicable in such ways of representing information and informational objects. So, adding terms for describing some specific characteristics of publications would be appreciated and make such ontology vocabulary richer and more comprehensive. For instance, we did not identify terms to properly describe some FRBR entities within FaBiO’s terms, like illustrators and translators, which although are not usually considered as main access points to a publication, such metadata can be important in a particular context. Inclusively, such approaches apply beyond the scope of our analysis: considering the description of a painting (as a type of art) for example, we did not notice terms for identifying different techniques of painting, like watercolor or oil painting, which may be an important information for an art gallery, for example.

In short, we noticed that issues involving standardization issues concerning citing and referencing metadata are long standing and were aggravated by the expressive increase in diversity in which information can be recorded and represented. The distancing of the forms of manipulation and processing of bibliographic metadata within the scope of descriptive representation, whose landmark was the publication of the FRBR, signals the need for a joint review of the facets that comprise it, including the registration of citing and referencing metadata. Technology has a lot to contribute to this process, considering the possibility of using a single intercommunication protocol to make the metadata machine readable, even though they are structured according to different guidelines. In this sense, even though there are thousands of bibliographic guidelines, as well as the different styles of references, the unification of the interpretation of bibliographic languages can result in a rationalization of the time dedicated to writing, revising, and normalizing bibliographic metadata and, in this sense, FaBiO Ontology is suggested as an alternative for the semantic enrichment of metadata.

Bibliographic universe and the issues that permeate metadata management matters are very complex and demands not only the experts' efforts but also the multidisciplinary involvement. More and more Librarianship and Computer Science approach to each other and such partnership tends to result in notable benefits to the citing network, to the information retrieval, to the bibliographic metadata management and to the information and for Scientific universe not to mention the researcher, directly. In this context, this study is only the tip of the iceberg in a vast field to be explored. From this perspective, further in-depth discussions about the applicability of semantic tools to the bibliographic universe are strongly encouraged. As a suggestion, we point the analysis of other semantic models existing online which allow the treatment of bibliographic metadata and the coverage of the informational needs of publishers, authors and readers. Also, it is suggested that SPAR Ontologies may support open citations model and, therefore, studies mapping and expanding the discussions first introduced by this study are also encouraged. A micro perspective of the issues approached in this study are also welcomed.

It is also pertinent to say that this subject represents such a huge question, with so many paradigms and concepts (some correct, other not so much) and traditions, that a single thesis would never be enough even to propose a solution to all the questions pointed in these discussions. But certainly, it represents a step towards the excellence in scientific production normalization.

The findings of this study demonstrate the need to care for the way information is being represented and evinced that Information Science do not domain all the requisites to properly deal with this scenery, especially considering the multiple and multidisciplinary approaches that can possibly be attributed to information in many aspects. Facing this, it is not a surprise that Information Science is no longer self-sufficient in the treatment of information, mainly considering the changes being carried not only in the information representation field but, also in the different embodiments and formats information may assume nowadays, which are even more and more diverse. In this context, we retake the premises of this study, which were all confirmed by the analysis:

a) The most common errors detected in the articles analyzed by Sweetland in a study carried out in 1989, still can be observed in the articles published nowadays. This suggests that something is going wrong (or unclear) between the elaboration and the application of reference styles and standards for writing bibliographic references;
b) The FaBiO Ontology, although might be improved to better meet the bibliographic description needs, is a suitable tool for describing the bibliographic metadata composing the bibliographic reference, which reaffirms and evinces the contribution that Computer Science might make to the Information Science and the need that both disciplines work together on Information management.

The confirmation of such premises can be assumed as a contribution of this study to Information Science, as the starting point towards the improvement of the bibliographic metadata management, considering the current information context and the trends where the information and its representation are heading. The main beneficiary of the interdisciplinary aspects between Information Science and Computer Science, beyond themselves, will be the scientific community, since the attribution of semantic aspects to the bibliographic metadata might simplify the processes of writing and normalizing bibliographic references, mentions and quotations, and, more than this, might strengthen the citation networks through the semantic relations among works related to each other. It is still a long way to go towards that but, a first step was taken here.

REFERENCES

ABNT. **ABNT NBR 6023 - informação e documentação - referências**: elaboração. Rio de Janeiro: ABNT, 2018.

ALMEIDA, M.; SOUZA, R.; FONSECA, F. Semantics in the Semantic Web: a critical evaluation. **Knowledge Organization**, Germany, v. 38, n. 3, 2011. Available at: http://mba.eci.ufmg.br/downloads/SemanticsSemWebko_38_2011.pdf. Accessed in: May 11 2021.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli:** Revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 8, n. 15, p. 18-40, 2003. Available at: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p18/5233. Accessed in: May 11. 2021.

ÁLVAREZ DE TOLEDO, L. **Principales estilos de citas bibliográficas**. [*S.l.*], June 19 2012. [message in blog]. Infobiblio. Available at: http://infobiblio.es/principales-estilos-de-citas-bibliograficas. Accessed in: Dec. 28 2012.

ANGELONI, M. T. Elementos intervenientes na tomada de decisão. **Ciência da Informação,** Brasília, DF., v. 32, n. 1., p. 17-22, jan./abr. 2003. Available at: http://revista.ibict.br/ciinf/article/view/1015/1070. Accessed in: May 11 2021.

ANGLO-AMERICAN CATALOGING RULES. 2. ed. Chicago: ALA, 2002.

BACCA, M. ed. Introduction to metadata. 2nd. ed. Los Angeles: The Getty Research Institute, 2008. Available in: https://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf. Accessed in: May 11 2021.

BAPTISTA, D. M. O impacto dos metadados na representação descritiva. **Revista ACB: Biblioteconomia em Santa Catarina,** Florianópolis, v. 12, n. 2, p. 177-190, jul./dez., 2007. Available at: https://revista.acbsc.org.br/racb/article/download/529/663. Accessed in: May 10 2021.

BARBEAU, E. **Why thousands of citation styles exist**. [*S. l.: s. n.*], 2018. Available at: https://www.citethisforme.com/blog/2018/08/17/thousands-citation-styles-exist. Accessed in: Aug.17 2020.

BERESTOVA, T. F. The ontological unity of bibliographic information and metadata. **Scientific and Technical Information Processing,** New York, v. 44, n. 1, p. 27-37, 2017. Available at: https://link.springer.com/content/pdf/10.3103/S0147688217010087.pdf. Accessed in: May 10 2021.

BERNERS-LEE, T. **Answers for young people.** [*S.l.*, 199?]. Available in: https://www.w3.org/People/Berners-Lee/Kids.html. Accessed in: May 10 2021.

BIBME. Why are there so many citation styles?. [*S. l.: s. n.*], 2017. Available at: https://www.bibme.org/blog/updates/many-citation-styles/. Accessed in: Aug. 17 2020.

BLACKBURN, S. **The Oxford dictionary of Philosophy**. Oxford: Oxford University Press, 1996.

BONACICH, P. Power and centrality: a family of measures. **American Journal of Sociology**, Chicago, v. 92, n. 5, p. 1170-1182. Available at: https://doi.org/10.1086/228631. Accessed in: May 11 2021.

BREITMAN, K. **Web semântica:** a internet do futuro. São Paulo: Gen, 2007.

BROOKES, B. C. Information technology and the science of information. *In*: ODDY, R. N. (ed.). *et al*. **Information retrieval research.** London: Butterworths, 1981. p. 1-8.

CALLAWAY, E. Publishing elite turns against impact factor: senior staff at societies and leading journals want to end inappropriate use of the measure. **Nature**, London, v. 535, n. 7611, p. 210-211, Jul. 2016. Available at: https://www.nature.com/news/polopoly_fs/1.20224!/menu/main/topColumns/topLeftColumn/pdf/nature.2016.20224.pdf. Accessed in: May 11 2021.

CAPLAN, P. **Metadata fundamentals for all librarians.** Chicago: ALA, 2003.

CAPURRO R.; HJØRLAND, B. O conceito de informação. Rio de Janeiro, **Perspectivas em Ciência da Informação,** Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007. Available at: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/download/54/47. Accessed in: May 10 2021.

CARRIÓN, N. T. *et al*. Normas para referenciar la bibliografía consultada en los trabajos de investigación. **European Journal of Podiatry / Revista Europea de Podología**, [Corunha], v. 3, n. 1, p. 26-33, Jan. 2017. Available at: https://revistas.udc.es/index.php/EJP/article/view/ejpod.2017.3.1.1711. Accessed in: May 11 2021.

CHAUMIER, J. **Le traitement linguistique de L'information**. 3. ed. Paris: Éditions ESF-Entreprise Moderne d'Édition: Libraries techniques, 1988.

CHEN, P. **Modelagem de dados**: a abordagem entidade-relacionamento para projeto lógico. São Paulo: Mcgraw Hill, 1990.

CLARIVATE ANALYTICS. **InCites Journal Citation Reports Help**. [*S.l*]: Clarivate Analytics, [2019?]. Available at:

http://help.incites.clarivate.com/incitesLiveJCR/glossaryAZgroup/g8/4346-TRS.html. Accessed in: May 11 2021.
COYLE, K. **FRBR before and after**: a look a tour bibliographic models. Chicago: ALA editions, 2016. Available at: http://kcoyle.net/beforeAndAfter/978-0-8389-1364-2.pdf. Accessed in: May 11 2021.

CRONIN, B. Norms and functions in citation - the view of journal editors and referees in psychology. **Social Science Information Studies**, Sevenoaks, v. 2, n. 2, p. 65-77. Available at: https://doi.org/10.1016/0143-6236(82)90001-1. Accessed in: May 11 2021.

CRUNCHBASE. Eric Smith. c2021. Available at: https://www.crunchbase.com/person/eric-schmidt. Accessed in: May 12 2021.

CSL. **CSL Styles: a guide for publishers**. [S.l.]: CSL, [2020?]. Available at: https://citationstyles.org/publishers/. Accessed in: May 11 2021.

CUNHA, M. B. da.; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia.** Brasília, DF: Briquet de Lemos, 2008.

CUTTER, C. A. **Rules for a dictionary catalog.** 4th ed. Rewritten. Washington: Government Printing Office, 1904. Available at: https://digital.library.unt.edu/ark:/67531/metadc1048/m1/6/. Accessed in: May 11 2021.

DAVENPORT, T. H. **A ecologia da informação.** São Paulo: Futura, 1998.

DAVIS, I. **VANN**: A vocabulary for annotating vocabulary descriptions. [*S.l.*], 2005. Available at: https://vocab.org/vann/#preferredNamespacePrefix. Accessed in: May 11 2021.

DCMI USAGE BOARD. **DCMI Metadata Terms**. United States, 2020. Available at: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/. Accessed in: May 11 2021.

DRUCKER, P. The coming of the new organization. **Harvard Business Review,** Harvard, Jan./Feb., p. 45-53, 1998.

EASE. **EASE statement on inappropriate use of impact factors**. [*S.l.*]: EASE, 2007. Available at: https://ease.org.uk/wp-content/uploads/ease_statement_ifs_final.pdf. Accessed in: May 11 2021.

EIS, D. Linked data, JSON-LD e web semântica. *In*: BRAZILJS CONF 2016, 6.,2016, Porto Alegre. Available at: https://www.youtube.com/watch?v=vAjBRcg3ydM. Accessed in: May 10 2021.

ELLIS, D. **Progress and problems in information retrieval.** 2nd ed. London: Library Association Publishing, 1996.

FERNANDES, M. V. Sprite - Comercial Óleo de Fígado de Bacalhau 1996. [*S.l., s.n.*], 2013. Available at: https://www.youtube.com/watch?v=z_rSS21MCjU. Accessed in: May 12 2021.

FONDAZIONE BRUNO KESSLER. **Journal citation ranking and quartile scores**. Trento, Fondazione Bruno Kessler, c2012. Available at: https://researchassessment.fbk.eu/quartile_score. Accessed in: May 11 2021.

FONSECA, F. The Double Role of Ontologies in Information Science Research. **Journal of the American Society for Information Science and Technology**, United States, v. 58, n. 6, p. 786-793, Apr. 2007. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20565. Accessed in: May 11 2021.

FREI, H. P. **Information retrieval:** from academic research to practical applications, 1996. Available at: http://www.ubilab.org/publications/print_versions/pdf/fre96.pdf. Accessed in: May 10. 2021.

FUSCO, E. **Modelos conceituais de dados como parte do processo da catalogação:** perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais. 2010. 249 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Available at: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/fusco_e_do_mar.pdf. Accessed in: May 10 2021.

GALVÃO, M. C. B. Construção de conceitos no campo da ciência da informação, **Ciência da Informação**, Brasília, DF, v. 27, n. 1, p. 46-52, 1998. Available at: https://doi.org/10.1590/S0100-19651998000100006. Accessed in: May 11 2021.

GEEK and poke. [*S.l.: s.n.,* c2010?]. Available at: http://geekandpoke.typepad.com/geekandpoke/2010/04/meta.html. Accessed in: May 12 2021.

GILL, T. Metadata and the Web. *In*: BACA, M. (ed.). **Introduction to metadata**. 3rd ed. Los Angeles: The Getty Institute, c2016. Available at: http://www.getty.edu/publications/intrometadata/metadata-and-the-web/. Accessed in: May 11 2021.

GILLILAND, A. J. Setting the stage. *In*: BACCA, M. (ed.). **Metadata**. 2nd ed. Los Angeles: The Getty Research Institute, 2008. p. 1-19.

GLEICK, J. **A informação:** uma história, uma teoria, uma enxurrada. São Paulo: Companhia das Letras, 2013.

GRATZ, A. Why are there so many different citation styles?. [*S. l.: s. n.*], 2016. Available at: https://libraries.mercer.edu/research-tools-help/citation-tools-help/DifferentCitationStylesHandout.pdf. Accessed in: May 11 2021.

GRUBER, T. R. A Translation approach to portable ontology specifications. **Knowledge Acquisition**, United Kingdom, v. 5, n. 2, p. 199-220, June 1993. Available at: https://doi.org/10.1006/knac.1993.1008. Accessed in: May 11 2021.

GUARINO, N. Formal ontology and information systems. *In*: GUARINO, N. (coord.). **Formal ontology in information systems**. Amsterdam: IOS Press, 1998. p. 3-15.

GUERRERO-BOTE, V. P.; MOYA-ANEGÓN F. A further step forward in measuring journals' scientific prestige: the SJR2 indicator. **Journal of Informetrics**, Netherlands, v. 6, n. 4, p. 674-688, Oct. 2012. Available at: https://doi.org/10.1016/j.joi.2012.07.001. Accessed in: May 11 2021.

GUERNSEY, L. Cyberspace citations: scholars debate how best to cite research conducted on the Internet. **The Chronicle of Higher Education**, Washington, DC, v. 42, n. 18, p. A18-A21, Jan. 1996.

HART, H. Rules for compositors and readers at the University Press. Oxford: Oxford University Press, 1893.

HJØRLAND, B. Book review: the future of classification, edited by Rita Marcella and Arthur Maltby. **Journal of the American Society for Information Science and Technology**, United States, v. 53, n. 1, p. 57, 2002. Available at: https://doi.org/10.1002/asi.10008. Accessed in: May 11 2021.

HOFFMAN, G. L. Meeting users' needs in cataloging: what is the right thing to do?. **Cataloging & Classification Quarterly**, United States, v. 47, n 7, p. 631-641, 2009. Available at: https://doi.org/10.1080/01639370903111999. Accessed in: May 11 2021.

HOUAISS, A.; SALLES, V. M. de. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 1999.

ICMJE. **Preparing a manuscript for submission to a medical journal**. [*S. l.*]: ICMJE, c2020. Available at: http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#g. Accessed in: May 11 2021.

IFLA. **Best Practice for National Bibliographic Agencies in a Digital Age**: Functional requirements for bibliographic records (FRBR). 2017. Available in: https://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8915. Access in: Nov. 13 2019.

IFLA. **Digital libraries**: metadata resources. Den Haag: IFLA, 2005. Available at: https://www.ifla.org/node/9337. Accessed in: May 10 2021.

IFLA. **FRBR-Library reference model**. Den Haag: IFLA, 2016. Available at: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf. Accessed in: May 11 2021.

IFLA. **Statement of International Cataloguing Principles**. [*S.l.*]: IFLA, 2009. Available at: https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf. Accessed in: May 11 2021.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. **Functional requirements. for bibliographic records:** Final Report. [Washington, DC ?]: IFLA, 2009. Available at: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf. Accessed in: May 11 2021.

ISO. **ISO 690 - information and documentation - guidelines for bibliographic references and citations to information resources**. Geneva: ISO, 2010.

JOUDREY, D. N. *et al.* **Introduction to cataloging and classification**. 11th ed. Santa Barbara, CA: ABC-CLIO, 2015.

KARCHER, S.; ZUMSTEIN, P. **Citation styles: history, practice, and future**. Hoboken: Athorea, [ca. 2018]. Available at: https://www.authorea.com/users/102264/articles/124920-citation-styles-history-practice-and-future/_show_article. Accessed in: May 10 2021.

KHOO, C. S. G.; NA, J. Semantic relations in information science. **Annual Review of Information Science and Technology**, United States, v. 40, n. 1, p. 157-228, 2006. Available at: https://doi.org/10.1002/aris.1440400112. Accessed in: May 11 2021.

KURTH, C. Being realistic about motivation. **Phylosophy Studies**, Netherlands, v. 176, p. 2751-2765, 2019. Available at: https://doi.org/10.1007/s11098-018-1149-9. Accessed in: May 11 2021.

LANCASTER, F. W. **Indexação e resumos**. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LANNING, S. A modern, simplified citation style and student response. **Reference Services Review**, London, v. 44, n. 1, p. 21-37, 2016.

LASSILA, O.; MCGUINESS, D. L. **The role of frame-based representation on the Semantic Web**. Troy, 2001. Available at: https://www.researchgate.net/publication/229101030_The_role_of_frame-based_representation_on_the_Semantic_Web. Accessed in: May 11 2021.

LE COADIC, Y. **A Ciência da Informação**. 2. ed. rev. atual. Rio de Janeiro: Briquet de Lemos, 2004.

LIMA, G. A. de. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.25, número especial, p. 57-97, fev. 2020. Available at: https://periodicos.ufmg.br/index.php/pci/article/view/22283. Accessed in: May 11 2021.

LOGAN, R. K. **Que é informação?:** a propagação da organização na biosfera, na simbolosfera, na tecnosfera e na econosfera. Tradução de Adriana Braga. Rio de Janeiro: Contraponto: PUC-Rio, 2012.

LÓPEZ CARREÑO, R.; MARTÍNEZ MÉNDEZ, F. J. Tendencias de la descripción bibliográfica en revistas españolas de acceso abierto. **Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información**, Ciudad de México, v. 29, n. 65, p. 15-38, enero/abr. 2015. Available at: https://doi.org/10.1016/j.ibbai.2016.02.013. Accessed in: May 11 2021.

LOURENÇO, C. A. Metadados: o grande desafio da organização da informação na Web. **Informação & Sociedade**, João Pessoa, v. 17, n. 1, p. 71-80, jan./abr. 2007. Available from: https://periodicos.ufpb.br/ojs/index.php/ies/article/view/466/1466. Accessed in: May 11 2021.

MAIMONE, G. D. *et al.* Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade – Estudos**, João Pessoa, v. 21, n. 1, p. 27-35, jan./abr. 2011. Available at: https://brapci.inf.br/index.php/res/v/91937. Accessed in: May 11 2021.

MARTINS, N. D. **Oncologia pediátrica.org:** contribuições do design multimédia para o apoio a familiares, amigos e doentes na luta contra o Cancro Pediátrico. 2007. Dissertation (Master in Multimedia Art) – Faculdade de Belas Artes, Universidade do Porto, Porto, 2007. Available at: https://repositorio-aberto.up.pt/bitstream/10216/7306/2/DissOncologiaPed.pdf. Accessed in: May 11 2021.

MASIC, I. The importance of proper citation of reference in biomedical articles. **Acta Inform. Med.**, Sarajevo, v. 21, n. 3, p. 148-155, 2013. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3804522/pdf/AIM-21-3-148.pdf. Accessed in: May 11 2021.

MÉNDEZ RODRÍGUEZ, E. M. **Metadatos y recuperación de información:** estándares, problemas y aplicabilidad em bibliotecas digitales. Gijón: Trea, 2002.

MORENO, F. P. **Requisitos Funcionais para Registros Bibliográficos - FRBR: um estudo no catálogo da Rede Bibliodata**. 2006. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência, Universidade de Brasília, Brasília, DF, 2006. Available at: http://eprints.rclis.org/7593/1/DISSERTA%C3%87%C3%83O_FERNANDA_MORENO_-_UnB.pdf. Accessed in: May 11 2021.

NAUMIS PEÑA, C. Registro bibliográfico y referencia bibliográfica: una revisión conceptual. **Revista Interamericana de Bibliotecología**, Medelín, v. 31, n. 1, p. 227-245, ene./jun. 2008. Available at: https://aprendeenlinea.udea.edu.co/revistas/index.php/RIB/article/download/1925/1586. Accessed in: Nov. 11, 2019.

NORDQUIST, R. **How to use block quotations in writing**. [*S.l.*], 2019. Available at: https://www.thoughtco.com/what-is-block-quotation-1689173. Accessed in: May 11 2021.

OCLC. [*S.l.: s.n.,* c2018]. Available at: https://www.oclc.org/en/home.html. Accessed in: Feb. 15 2018.

OCLC. **OCLC Research activities and IFLA's Functional Requirements for Bibliographic Records**. [*S.l.*]: OCLC, c2020. Available at: https://www.oclc.org/research/activities/frbr.html. Accessed in: Aug. 11 2021.

O'CONNOR, M. Style manuals and guides for authors and editors: prescriptive or descriptive?. *In*: BALABAN, M. (coord.). **Scientific information transfer: the editor's role: proceedings of the first International Conference of Scientific Editors, April 24-29, 1977, Jerusalem**. Dordrecht: D. Reidel Publishing Company, 1977. p 287-290.

OSCA-LLUCH, J.; CIVERA MOLLÁ, C.; PEÑARANDA ORTEGA, M. Consecuencias de los errores em lãs referencias bibliográficas: el caso de la revista Psicothema. **Psicothema,** Oviedo, v. 21, n. 2, p. 300-303, 2009. Available at: http://www.psicothema.com/pdf/3630.pdf. Accessed in: May 12 2021.

PATIÑO DÍAZ, G. **Citas y referencias bibliográficas**. Bogotá: Pontificia Universidad Javeriana, 2007.

PEARCE-MOSES, R. **A glossary of archival and records terminology**. Chicago: The Society of American Archivists, 2005. (Archival fundamentals series. II). Available at: http://files.archivists.org/pubs/free/SAA-Glossary-2005.pdf. Accessed in: May 11 2021.

PERONI, S. **SPAROntologies/fabio. [*S.l.*], 2018. Available at:** https://github.com/SPAROntologies/fabio/blob/master/docs/2018-05-02/fabio.html**. Accessed in: May 11 2021.**

PERONI, S. *et al*. Setting our bibliographic references free: towards open citation data. **Journal of Documentation**, London, v. 71, n. 2, p. 253-277, 2015. Available at: https://doi.org/10.1108/JD-12-2013-0166**.** Accessed in: May 11 2010.

PERONI, S.; SHOTTON, D. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. **Journal of Web Semantics**: Science, services and agents on the World Wide Web, Netherlands, v. 17, p. 33-43, Dec. 2012a. Available at: https://doi.org/10.1016/j.websem.2012.08.001. Accessed in: May 11 2021.

PERONI, S.; SHOTTON, D. **FaBiO, the FRBR-aligned Bibliographic Ontology**. **[*S.l.*], 2019. Available at:** https://sparontologies.github.io/fabio/current/fabio.html**. Accessed in: May 11 2021.**

PERONI, S.; SHOTTON, D. **FRBR-aligned Bibliographic Ontology (FaBiO)**. **[*S.l.*], 2012**b**. Available at:** http://www.sparontologies.net/ontologies/fabio. Accessed in: May 11 2021.

PERONI, S.; SHOTTON, D. **SPAR Ontologies**. **[*S.l.*, c2021?].** Available at: http://www.sparontologies.net/#. Accessed in: May 11 2021.

PERONI, S.; SHOTTON, D. The SPAR Ontologies. In: VRANDEČIĆ D. *et al*. (cords.). **The Semantic Web – ISWC 2018**: ISWC 2018: lecture notes in Computer Science. Cham: Springer, 2018. p. 119-136. v. 11137. Available at: https://doi.org/10.1007/978-3-030-00668-6_8. Accessed in: May 11 2021.

PESSOA, C. **Informação não é conhecimento.** 2014. Available at: http://www.carlospessoa.com.br/artigo/informacao-nao-e-conhecimento-2/. Accessed in: Ago.18 2017.

PIWOWAR, H. *et al.* The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. **PeerJ**, United States, v. 6, e4375, 2018. Available at: https://doi.org/10.7717/peerj.4375. Accessed in: May 11 2021.

POLLOCK, J. T. **Web semântica para leigos**. Rio de Janeiro: Alta books, 2010.

QUEM disse. [*S.l.: s.n.*], c2021. Available at: https://quemdisse.com.br/frase/imagem-nao-e-nada-sede-e-tudo-obedeca-sua-sede/100831/. Accessed in: May 12 2021.

RDA TOOLKIT **Carrier type**. United States, c2021a. Available at: https://access.rdatoolkit.org/Content/Index?externalId=en-US_ala-0dd63c0d-066f-3ddc-885a-ff83c25cf752&highlight=carrier&highlight=type. Accessed in: May 11 2021.

RDA TOOLKIT **Content type**. United States, c2021b. Available at: https://access.rdatoolkit.org/Content/Index?externalId=en-US_ala-1a6a5219-2cd5-3762-87e8-6ee28e327b8b&highlight=content&highlight=type. Accessed in: May 11 2021.

RDA Toolkit Introduction to RDA. (c2021c). Available at: https://access.rdatoolkit.org/Guidance/Index?externalId=en-US_ala-d3b460ae-1818-3459-8c9c-9dd5fda8edf8. Accessed in: May 12 2021.

RIVA, P. Introducing the Functi onal Requirements for Bibliographic Records and related IFLA developments. **Bulletin of the American Society for Information Science and Technology**, Maryland, v. 36, n. 6, p. 7-11, 2007. Available at:

https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/bult.2007.1720330604. Accessed in: May 11 2021.

RIVA, P.; Le BŒUF, P.; ŽUMER, M. **IFLA Library Reference Model**: a conceptual model for bibliographic information. Den Haag, IFLA, 2017. Available at: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf. Accessed in: May 11 2021.

RODRIGUES, M. E. F.; LIMA, M. H. T. de; GARCIA, M. J. de O. A normalização no contexto da comunicação científica. **Perspectivas em Ciência da Informação,** Belo Horizonte, v. 3, n. 2, p. 147-156, jul./dez. 1998. Available at: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/download/603/372. Accessed in: 10 mai. 2021.

ROWLEY, J. **A biblioteca eletrônica.** Brasília, DF: Briquet de Lemos, 2008.

RUIZ-PÉREZ, R.; DELGADO LÓPEZ-CÓZAR, E.; JIMÉNEZ-CONTRERAS, E. Criterios del Institute for Scientific Information para la selección de revistas científicas: su aplicación a las revistas españolas: metodología e indicadores. **International Journal of Clinical and Health Psychology**, Granada, v. 6, n. 2, p. 401-424, mayo 2006. Available at: https://www.redalyc.org/articulo.oa?id=33760211. Accessed in: May 11 2021.

RUSSO, M. **Fundamentos de Bibliotecomonia e Ciência da Informação**. Rio de Janeiro: e-papers, 2010. (Coleção Biblioteconomia e Gestão de unidades de informação. Série didáticos, n. 1).

SANTOS, E. A. **Raw and aggregated data for the study introduced in the doctoral thesis "FRBRizing bibliographic references structures: an approach from in-text references pointers and SPAR Ontologies for describing bibliographic metadata**. Version 1.0. [S.l.]: Zenodo 2021. Available at: https://zenodo.org/record/4759403#.YJ6dsqhKjzY. Accessed in: May 14 2021.

SANTOS, E. A. dos; PERONI, S.; MUCHERONI, M. L. Citing and referencing habits in medicine and social sciences journals in 2019. **Journal of Documentation**, United Kingdom, v. ahead of print, n. ahead of print, 2021. Available at: https://doi.org/10.1108/JD-08-2020-0144. Accessed in: May 11 2021.

SANTOS, E. A. dos; PERONI, S.; MUCHERONI, M. L. **Glossary of the article "Citing and referencing habits in Medicine and Social Sciences journals in 2019**. Version 1.0. [S.l.]: Zenodo 2020a. Available at: https://doi.org/10.5281/zenodo.3996578. Accessed in: May 11 2021.

SANTOS, E. A. dos; PERONI, S.; MUCHERONI, M. L. Glossary of the article " **Raw and aggregated data for the study introduced in the article "Citing and referencing habits in Medicine and Social Sciences journals in 2019**. Version 1.0. [*S.l.*]: Zenodo 2020b. Available at: https://doi.org/10.5281/zenodo.3996582. Accessed in: May 11 2021.

SANTOS, E. A. dos; PERONI, S.; MUCHERONI, M. L. **Workflow for retrieving all the data of the analysis introduced in the article "Citing and referencing habits in Medicine and Social Sciences journals in 2019**. Version 1.0. [*S.l.*]: Zenodo 2020c. Available at: https://dx.doi.org/10.17504/protocols.io.bbifikbn. Accessed in: May 11 2021.

SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos:** Arquivística, Biblioteconomia, Documentação, Informática. 2. ed. rev. e ampl. Campinas, SP: Átomo, 2012.

SANTOS, J. C.; MARTINS, F. V.; KARL, H. B. Different perspectives on internationalization research: a bibliometric review. **Revista Ibero-Americana de Estratégia**, São Paulo, v. 14, n. 4, p. 93-118, out./dez. 2015. Available at: https://www.researchgate.net/publication/304420569_Different_Perspectives_on_Internationalization_Research_A_Bibliometric_Review. Accessed in: May 11 2021.

SCIMAGO. SJR - SCImago Journal & Country Rank. [*S.l.*], c2020. Available at: http://www.scimagojr.com. Accessed in: May 11 2021.

SEARLE, J. **Minds, brains and science:** the 1984 Reith Lectures. New York: Penguin Books, 1991.

SETZER, V. W. **Dado, informação, conhecimento e competência.** [São Paulo], 2015. Available at: https://www.ime.usp.br/~vwsetzer/dado-info.html. Accessed in: Abr. 18 2017.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal,** New York, v. 27, n. 3, p. 379-423, Jul. 1948. Available at: http://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf. Accessed in: May 10 2021.

SHOTTON, D.; PERONI, S. **DoCo, the document components ontology**. [*S.l.*: *s.n*.], 2015. Available at: https://sparontologies.github.io/doco/current/doco.html#d4e382. Accessed in: May 11 2021.

SHOTTON, D. Funders should mandate open citations. **Nature**, London, v. 553, p. 129, Jan. 2018. Available at: https://www.nature.com/magazine-assets/d41586-018-00104-7/d41586-018-00104-7.pdf. Accessed in: May 10 2021.

SIEGLER, M. G. **Eric Schmidt**: every 2 days we create as much information as we did up to 2003. [*S.l.*]: Verizon Media, c2021. Available at: https://techcrunch.com/2010/08/04/schmidt-data/. Accessed in: May 12 2021.

SINGHI, A. **Dear Sprite, why so green?**. [India]: The Coca Cola Company, 2017. Available at: https://www.coca-colaindia.com/choices/dear-sprite-why-so-green. Accessed in: May 12 2021.

SMITH, B.; WELTY, C. Ontology: Towards a New Synthesis. *In*: WELTY, C; SMITH, B. (coord.). **Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS '01)**. Ongunquit, ME: ACM Press, 2001. p. iii-x.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**, United States, v. 50, n. 12, p. 1119-1120, 1999. Special issue. Available at: https://doi.org/10.1002/(SICI)1097-4571(1999)50:12%3C1119::AID-ASI12%3E3.0.CO;2-I. Accessed in: May 11 2021.

SWEETLAND, J. H. Errors in bibliographic citations: a continuing problem. **The library quarterly**, Chicago, v. 59, n. 4, p. 291-304, Oct. 1989. Available at: https://www.journals.uchicago.edu/doi/10.1086/602160. Accessed in: May 11 2021.

TAYLOR A. G., **Introduction to cataloging and classification**. 10th ed. Westport: CT:Libraries Unlimited, 2006.

TENNANT, R. Cross-database search: one-stop shopping. **Library Journal,** New York, 2001. Available at: http://roytennant.com/column/?fetch=data/69.xml. Accessed in: May 11 2021.

TILLET, B. FRBR and cataloging for the future. **Cataloging & Classification Quarterly**, United States, v. 39, n. 3-4, p. 197-205, 2009. Available at: https://doi.org/10.1300/J104v39n03_12. Accessed in: May 11 2021.

TILLET, B. **The FRBR Model (Functional Requirements for Bibliographic Records)**. 2003a. paper presented at the ALCTS Institute on Metadata and AACR2, 4-5 April, San Jose, CA. Available at: https://www.loc.gov/catdir/cpso/frbreng.pdf. Accessed in: May 11 2021.

TILLET, B. **What is FRBR? a conceptual model for the bibliographic universe**. Library of Congress Cataloging Distribution Service, [Washington, DC], 2003b. Available at: https://www.loc.gov/cds/downloads/FRBR.PDF. Accessed in: May 11 2021.

TIME to remodel the journal impact factor. **Nature**, London, v. 535, n. 7613, p. 466, July 2016. Available in: https://www.nature.com/articles/535466a.pdf. Accessed in: May 11 2021.

UCM WRITING CENTER. Why are there so many citation styles?. [*S.l.*]: UCM Writing Center, 2016. Available at: http://ucmwriting.weebly.com/blog/why-are-there-so-many-citation-styles#. Accessed in: May 11 2021.

UZUNALLI, G. *et al.* Dynamic transition of the blood-brain barrier in the development of non-small cell lung cancer brain metastases. **Oncotarget**, United States, v. 10, n. 59, p. 6634-6348, 2019. Available at: https://doi.org/10.18632/oncotarget.27274. Accessed in: May 11 2021.

VICKERY, B. C. Ontologies. **Journal of Information Science**, United States, v. 23, n. 4, p. 277-286, 1997.

W3C. **Metadata and resource description**. 2001. Available from: https://www.w3.org/Metadata/. Accessed in: May 11 2021.

WEB OF SCIENCE GROUP. **Backward and forward citations**. [*S.l.*]: Web of Science Group, 2019. Available at: https://clarivate.libguides.com/woscc/citationnetwork. Accessed in: May 11 2021.

WIKIPEDIA. **Impact factor**. [*S.l.*]: Wikipedia, 2019. Available at: https://en.wikipedia.org/wiki/Impact_factor#cite_note-garfield-4. Accessed in: May 11 2021.

WURMAN, R. S. **Ansiedade de informação.** 5. ed. São Paulo: Cultura, 1995.

ZABEL, D.; RIMLAND, E. Ranghanathan's relevant rules. **Reference & User Services Quarterly,** [Chicago?], v. 46, n. 4, p. 24-26, 2007. Available at: https://doi.org/10.5860/rusq.46n4.24. Accessed in: May 10 2021.

ZHUGE, H. General Citation. *In*: ZHUGE, H. **Multi-dimensional summarization in cyber-physical society**. Amsterdam: Elsevier, 2016. Chapter 8. p. 107-114. (Computer Science Reviews and Trends).

ZIMAN, J. **Conhecimento público.** Belo Horizonte: Itatiaia; São Paulo: Edusp, 1979.

GLOSSARY

One of the findings of this research in the very beginning of its development was the divergences on terminological understandings between Information Science and Computer Science. Citations, bibliographic references, references, among others are some examples of shared terms between these two areas. However, some professionals may have different views on the same term and that situation might configure misunderstandings within the scope of this research considering its multidisciplinary approach. Facing this, it was identified the convenience of describing some technical terms to clarify its meanings in the ambit of this study.

It is also important to point that some terms may have more than one meaning (and some of them actually do have many), including within other subject areas. For these cases, please keep in mind that this glossary was written considering a librarian's overview, with some influence of Computer Science.

Article

The smallest part of a periodic publication that, although is itself a distinct entire publication, generally integrates a journal issue. Study of a particular subject, with proper methods and discussion that generally is submitted to an evaluation process and approved by an editorial board prior to publication.

Bibliography

The term bibliography is used for designating a systematic list of sources (e.g., books, articles, and any other press publication) used to write a work (e.g., an article). It usually includes the indication of the sources (publications) consulted and used by an author to formulate and support his own ideas about a topic and, the ones considered as a convenient (important, useful) background reading for its readers, like a reading suggestion list, (CUNHA; CAVALCANTI, 2008).

It is also used to designate a systematic list that gather a compilation of descriptive elements of a group of publications regarding to specific common characteristics shared among them, also in the format of bibliographic references. Bibliographies may be general or specialized and be organized and presented under several approaches: by subject, by chronological order (current, prospective, retrospective), by region, by authorship, or even as a descriptive list of publications targeted to a specific category of readers (CUNHA; CAVALCANTI, 2008).

Bibliography is also used to refer to a set of descriptive metadata, gathered to describe a single source of information (e.g., a book, a book chapter, an article etc.), similar than a bibliographic reference. A bibliography and a bibliographic reference do not differ in a matter of appearance, but in a matter of content and context. Bibliographies and bibliographic references look very similar, but bibliographies refer to the description of a source of information that may be indicated as a reading recommendation or simply an informative note, while bibliographic references refer to the description of a source of information that has been cited in a work.

The term bibliography is ambiguous since it was first used, due to the semantic polyvalence that has been assigned to the term (NAUMIS PEÑA, 2008) and this is not different considering the list of information sources approach, because the term bibliography denotes a pressed source of information, since the radical –bibion refers to books and the prefix –graphy to the writing. These ambiguities are strongly present in the scientific universe and it is common to find the term bibliography used instead of the term bibliographic references, particularly as bibliographic references lists' titles.

For the purposes of this glossary, the term bibliography must be considered as a list of descriptive elements, gathered in bibliographic references format, which stands for a compilation of works suggested as a complimentary content to the content of a work, or a compilation of bibliographic references which indicate the publications consulted by an author for grounding his ideas but not properly cited in the body of the text.

Bibliographic reference

> The textual entity within a citing work that identifies a cited work. A bibliographic reference contains some of the elements of the full bibliographic record for the cited work, arranged in a specific format determined by the house style of the citing publication. Some journal house styles require omission of particular elements of the bibliographic record regarded as essential by others, such as the names of all the authors, the title of the article, and the DOI. For a journal article, the bibliographic reference minimally comprises: first author's surname and initials, publication year, abbreviated journal name, volume number, and first and last page numbers. Typically,

> when the citing work is a scientific journal article, each bibliographic reference is complete in itself and forms a reference list item in the “References” section at the end of the article. In other types of publication, particularly in the humanities, bibliographic references may be contained within footnotes, may be mixed with comments, and may contain pointers, such as “ibid.” (abbreviation of the Latin ibidem, meaning “the same place”) and “op. cit.” (abbreviation of the Latin phrase opere citato, meaning “in the work cited”), that refer the reader to a previous bibliographic reference from which information needs to be extracted and duplicated to complete the current incomplete bibliographic reference. For this reason, automated parsing of bibliographic references within humanities publications is particularly difficult. Because errors can be introduced when an author creates a bibliographic reference, a published bibliographic reference should not be trusted to be a fully accurate expression of the information contained within the authoritative bibliographic record for that cited work (PERONI et al., 2015, p. 256-257).

Some publishers use both the terms “references”, “bibliography”, and also “bibliographic references” indistinctly to refer to different things. It suggests a terminological understanding issue that must be clarified and maybe the necessity of adopting a new terminology, once the term “bibliographic” may not be suitable nowadays (see complimentary information at bibliography term definition, above).

In the ambit of this glossary, the term “references” will never be used to refer to bibliographic references (see proper definition for references term below) and must be understood exclusively as a specific kind of reference that refers the reader to an original source of information cited by an author in the text body of a work, and contains the indication of the minimum and indispensable standardized set of precise and detailed descriptive bibliographic metadata for the identification of an information, an entire publication or a specific part of it, a speech, or anything else that may be citable, in order to enable it to be located and retrieved (ABNT, 2018; CUNHA; CAVALCANTI, 2008; ISO, 2010).

Bibliographic references list

The complete descriptive list of bibliographic references of works effectively cited in the main body of a text (work), ordered alphabetically, by the surnames of the cited authors or, numbered consecutively according to the first mention of each publication within the work (the ordination (logical sequence) of bibliographic references in a bibliographic references lists depends on the guidelines of the bibliographic style adopted for writing each work). Formally, the bibliographic references lists should not include bibliographic references of works for which there are any references (mentions or quotations) in the main body of the work.

Bibliographic style

A set of rules, recommendations or guidelines for presenting and formatting academic and scientific works, including bibliographic references. Some bibliographic styles may also contain guidelines on presenting mentions and quotations.

Bibliographic record

A data record containing metadata that fully describes a particular publication and is held in some authoritative information system or library catalogue. A bibliographic record comprises a set of entities defined by the publisher, although the bibliographer or the cataloguer, who usually create bibliographic records, can include some descriptive information they consider important about the publication being described. A bibliographic record of an article, for example, include, but are not restricted to the names of all authors, the title of the article, the journal title, the volume number, the issue number, the first and last page numbers, the full publication date, the publisher's name, the copyright information, the peer-reviewed status, the open access status, the digital object identifier (DOI) for the article and the International Standard Serial Number (ISSN) for the journal (PERONI et al., 2015).

An example of a bibliographic record is available within the definition of the term "descriptive element".

Block quotation

Block quotation (also called long quotation, extract, set-off quotation, or display quotation (NORDQUIST, 2019)) is a quotation (see the definition for quotation below) that exceed a specific length for which there is no single rule to define the extent limit from which a block quote should be considered as so. Regarding to formatting and presentation rules, block quotations differ from run-in quotations but, once more, there is no a single rule for indicating these differences. Block quotations are typically distinguished visually in the main body of the text (SHOTTON; PERONI, 2015), generally by a left margin indent and a smaller font size than the main text body but, each bibliographic style usually defines its own guidelines for quotations presentation, block quotations included.

Citation

Citation is one of the basic elements of a scientific paper that involves reading, representation, selection, language use, and language understanding behaviors (ZHUGE, 2016), which corresponds to an conceptual link established when a citing work mentions a cited work, or part of it, by including one or more mentions of the cited work content within the body text of that citing work, denoting the bibliographic reference of the cited work at appropriate points (according to the bibliographic style used) through the inclusion of one or more in-text reference pointers and consequently by including the bibliographic reference of the cited work in the bibliographic references list of the citing work. (CUNHA; CAVALCANTI, 2008; PERONI et al., 2015; SHOTTON, 2018).

Citations may be created in the text body of a citing work by the literally reproduction of an extract of the cited work (quotations) or by the basement of a discourse supported (or inspired) on the main idea of the content of the cited work. So, it is proper to consider that mentions are the representation of external information within the text body, and that those mentions (associated to an in-text reference pointer and a corresponding bibliographic reference) constitute the basis that establishes links among works, that is, citations, and supports the citation network.

When indicating a citation, authors generally intend to clarify, base or illustrate a specific approach of a subject; to confirm, deny or make a counterpoint (questioning) to a theory; to highlight similar or opposite expert’s opinions about a subject; to reinforce or compare an argument; or to avoid plagiarism and give authorship credits to a statement. Guernsey (1996) complements that the citation is the “mark of honest scholarship”. Lanning (2016, p. 22) points that “authors cite sources to show the pedigree of their thinking” (this shows authority), not to plagiarize (that shows accountability), and to allow the reader to find what the author has cited (that shows discoverability). It is also useful to say that citations are used as inputs for bibliometric studies.

Although citations correspond to an abstract concept, they are always identifiable within scientific text bodies. Zhuge (2016) points out that “explicit citations are the components of scientific papers and books” (see the definition for citation components below).

Citation networks

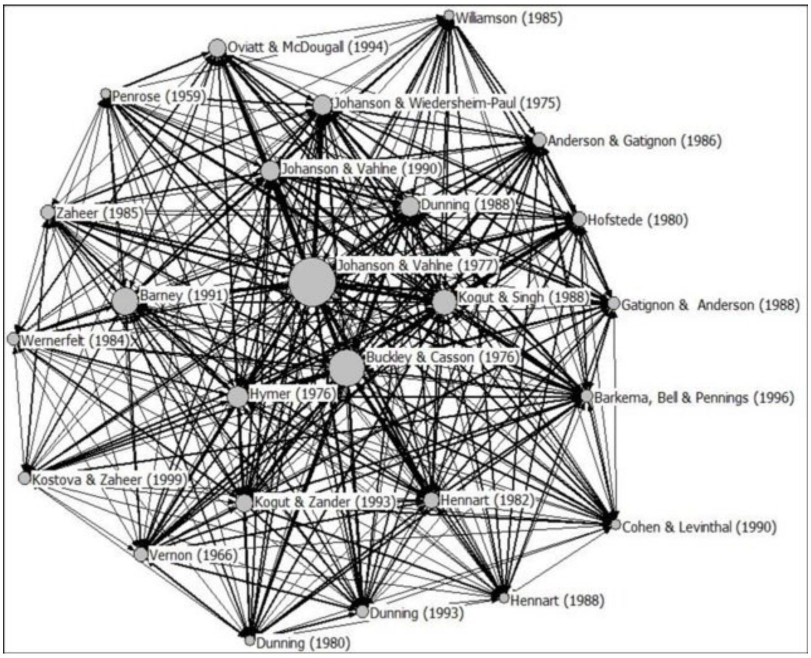
Citations networks are formed by the conceptual links that connect different works through citations (citations connections) and makes them string attached among themselves in the backstage. “Citations knit together the whole world of scholarship into a gigantic citation network into a global endeavor, a directed graph with publications as the nodes and citations as the links between them, assigning credit to other researchers” (PERONI et al., p. 255; SHOTTON, 2018).

The connection between one work and another is explicit through citations. It not just about a matter of connections among works, but a matter of connections of the ideas and the contents these works stand for. That is to say that there is an implicit subject relationship established between the cited work and the citing work (WEB OF SCIENCE GROUP, 2019).

Citation networks provide data about the development of a scientific area, the role and level of scientific contribution for researchers, institutions, and countries, and also, supports the estimating of the development of the areas, the decision on the possible directions a research might take and also, the impact assessment of a research (ZHUGE, 2016).

> Different from static text, the citation network dynamically renders the source, the formation and evolution of a study, the backbone, the impact of researchers and institutions, potential knowledge flows through citation links [135], and the networks of cooperation between researchers and between institutions with the evolution of the area. Summaries of different scales can be obtained through zoom-in-and-zoom-out on the citation network. It is feasible to transform a citation network into a text by using some language patterns (for example, “the idea of A was extended by B,” “the idea of A was used by B,” and “the idea of A inspired B”) to represent different citations, main roles, relations, and development track (ZHUGE, 2016).

Figure 1 – Example of a citation network



Source: Santos; Martins; Karl, (2015).

Like other bibliographic analysis tools, citations networks may be constructed under different perspectives: by article, by author, by publisher and so on, according to the necessity of the researcher. That is to say that there is infinity citations networks within the citation network that comprises the whole citations existing in the world.

Citation style

See bibliographic style

Citing work

The work that contains a bibliographic reference to another work (PERONI et al., 2015, p. 257).

Cited work

The work that is being referred to by such a bibliographic reference (PERONI et al., 2015, p. 257).

Descriptive element

A word, phrase, or group of characters representing a distinct unit of information that forms part of an area of formal description (PEARCE-MOSES, 2005). Describing implies in the representation in written or in spoken language, of the object that is being described. Basically, this representation intends to detail every single characteristic of the described object. In a bibliographic context, a descriptive element corresponds to the smallest unit of information that comprises the description of a work or a publication be in a bibliographic record, be in a bibliographic reference (e.g., the author, the publication title, the editor, the year of publication, etc.).

Below follows the descriptive representations for the work "Alice in wonderland". The first one corresponds to a part of a bibliographic registry in Machine Readable Cataloging (MARC) format and the second, a bibliographic reference format. Each red box content represents a single descriptive element. A set of descriptive elements usually corresponds to the descriptive representation of the information. A descriptive element might be used in any information description tool, including bibliographic registers and bibliographic references.

Examples:

a) Bibliographic registry

00000970cam a22002771a 4500
0017805326
00520190731112101.0
008850926s1928 ilua 000 1 eng
1001_ |a Carroll, Lewis, |d 1832-1898.
24510 |a Alice's adventures in Wonderland / |c by Lewis Carroll ; original Tenniel illustrations.
260__ |a Chicago : |b A. Whitman, |c c1928.
300__ |a 160 p. : |b ill. ; |c 21 cm.
4900_ |a Just right book
530__ |a Also available in digital form.
540__ |a No known restrictions; no copyright renewal found, Jul 24 2019
85641 | http://hdl.loc.gov/loc.gdc/scd0001.00025713291

b) bibliographic reference

LEWIS, C. **Alice's adventures in Wonderland.** Chicago: A. Whitman, c1928.

Document

Any kind of publication which is citable and also may be cited. This is not necessarily exclusively about works released by publishing companies (e.g., books, journals, dictionaries, encyclopedias, etc.) but also includes materials published independently or non-officially published by an editorial group (e.g. preprints, datasets, academic works, blog posts, free softwares, etc.).

Editorial board

Group of experts in a journal's field which is responsible for determining the journal guidelines and for leading peer-review processes for the works submitted to the journal.

FRBR Expression

A conceptual definition that corresponds to "the intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms" (IFLA, 2009, p. 19). A work may take the form of text, sound, image, object, movement, etc., or any combination of such forms. Each form the work takes, independent of which one it is, corresponds to a single expression of a work. That is, a work may be realized through one or more than one expression, but an expression corresponds to only one work (IFLA, 2009, 2017) e.g., considering "Alice in Wonderland" as a work, each version of this is an expression: written (book), image (movie), sound (audiobook), object (sculpture), movement (play), etc.

The expression concept is integrant of the WEMI-Model (Work, Expression, Manifestation and Item model – see definition also in this glossary), which tries to identify the core aspects of publications and is the foundation of the FRBR family (IFLA, 2017).

FRBR Entity

The entities represent the key objects of interest to users of a bibliographic data (IFLA, 2009, p. 13) and are defined as a “thing” or an “object” in the real world which may be uniquely identified in relation to all other objects and may be concrete or immaterial. The entities are complemented by the attributes, which indicates the characteristics of each type of entity, or descriptive properties of each member of a set of entities. An association between two or several entities corresponds to a relationship (CHEN, 1990).

As an entity-related concept model, all the FRBR entities are divided into three groups:

a) First group (the products of intellectual production): Work, expression, manifestation, item.
b) Second group (the entities responsible for the physical production and dissemination, or the custodianship of the entities of the first group): person and corporate body.
c) Third group (an additional set of entities that serve as the subjects of intellectual or artistic endeavor): concept, object, event, and place.

For the purposes of this glossary, only the definitions of the entities of the first group were considered and are strongly recommended to be consulted, although it is not an exhaustive definition. For the second and third groups entities definitions it is recommended the consult to the FRBR final report (IFLA, 2009) once it is not regarded to the scope of this glossary.

Freely available access

Refers to the possibility of consulting a permanently available online content that can be retrieved free of charges and without control, e.g., registrations or subscriptions (that indeed, may require payments). Despite being freely online accessible, this characteristic does not provide the consulter with permission of modifying, selling, or reproducing the content.

Restricted access

Basically, restricted access publications are those whose access is opposite to open access. Restricted access means the online availability of metadata for academic and / or scientific content, whose full access is granted through a user's counterpart, which can be a registration, an individual or institutional subscription or any other form of payment for access or, access control validated by IP address.

Functional Requirements for Bibliographic Records

See FRBR.

FRBR

FRBR (Functional Requirements for Bibliographic Records) is a complex conceptual entity-relationship model of the bibliographic universe, focused on the end-user, which was published in 1998 by the International Federation of Library Associations and Institutions (IFLA) and corresponds to the development of basic requirements for an international standard bibliographic record (COYLE, 2016; TILLET, 2003). “This functional requirement emphasizes the importance of understanding the function of the data elements being recorded and how these elements each contribute to meeting user needs” (RIVA, 2007, p. 7).

“FRBR uses entity-relationship modeling, which is a standard technique, borrowed from Computer Science, used to analyze the structure of data prior to programming, particularly for database design. The modeling is independent of any specific program, code or standard” (RIVA, 2007, p. 8).

Also, the FRBR entity-relationship model provides a structure within which data requirements can be analyzed in a systematic way and, a framework for analyzing the uses that are made of bibliographic data considering each single entity, its relationships, and attributes separately. This is useful not only for the task being performed by the user, but also to map the user tasks associated with the bibliographic resources described in catalogs, bibliographies, and other bibliographic tools they support, which can be four:

a) to find entities that correspond to the user’s stated search criteria;

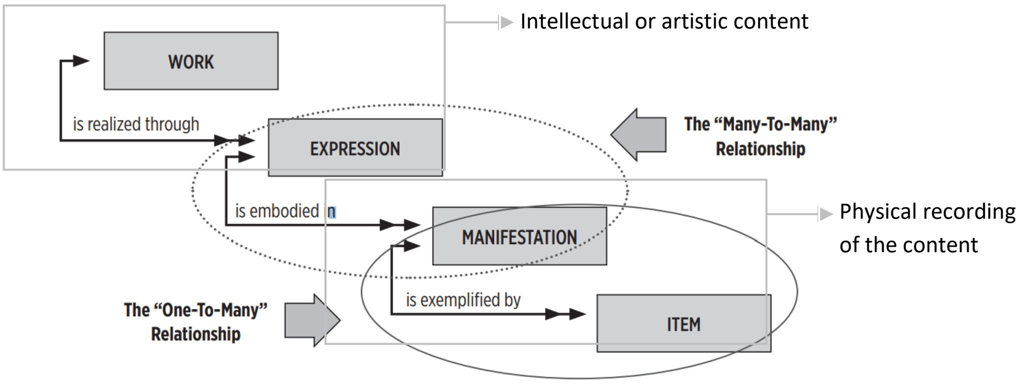
b) to identify an entity;

c) to select an entity;

d) to acquire or obtain access to the entity described (IFLA, 2009, p. 79).

FRBR group 1 entities corresponds to the primary FRBR entities and their relationships which are illustrated in figure 2.

Figure 2 - Many to many and one-to-many relationships in FRBR



Source: adapted from Coyle (2016) and Tillet (2003).

Each box of the figure represents an entity, that together composes the WEMI-Model (work, expression, manifestation, item – see definitions for all these entities also within this glossary). Each black arrow of the figure stands for a relationship between the entities, where one head indicates the possibility of establishing a single relationship, no exceptions admitted, and two heads indicates the possibility of establishing multiple relationships. That is to say that a work may be realized through one or more expressions. An expression may be embodied through one or more manifestations. A manifestation may be exemplified by one or more item.

> Manifestations may contain multiple expressions as indicated by the many-to-many relationship between expressions and manifestations. This is the only many-to-many relationship among the WEMI entities. A manifestation can embody multiple expressions and an expression can be embodied in multiple manifestations. By contrast, an expression can only realize a single work and an item can only exemplify a single manifestation (IFLA, 2016, p. 66).

On the other hand, the one-to-many relationship among manifestation and item entities means that a single manifestation may be exemplified in multiple items.

As already mentioned at the beginning of this definition, FRBR is a complex conceptual entity-relationship model and so, its definition is equally complex and would demand a long discussion that is not intended within the context of this glossary. For further and more detailed information about FRBR it is recommended the consult to the FRBR itself (see IFLA, 2009 bibliographic reference).

Issue

Each of the single edition published under the same volume of the same journal title.

Impact Factor (IF)

See Journal Impact Factor.

In-text reference pointer

> The entity present in the body text of a citing work that denotes a particular bibliographic reference in the reference list or a footnote. In scientific literature, this in-text reference pointer can be presented in different forms – as a square-bracketed or superscripted number (e.g. "[3]" or " 3 "); as a square-bracketed text string comprising the first letter of each author's surname (to a maximum of three) plus the last two digits of the publication year (e.g. "[RDS02]"); or as a parenthesised text string containing, for a single-author publication the author's surname and the publication year (e.g. "(Renear, 2002)"), for a two-author publication both authors' surnames and the publication year (e.g. "(Renear and Jones, 2002)"), or for a multi-author publication the first author's surname followed by "et al." and the publication year (e.g. "(Renear et al., 2002)") (PERONI et al., 2015, p. 257).

In-text citation

The mention of a cited work content which is included within the text body of a citing work and followed by an in-text reference pointer that refers to a correspondent entry in the bibliographic references list. Unlike quotations, in-text citations reproduce fragments of the cited work content according to the interpretation of the author of the citing work, that is, in his own words.

Item

The entity defined as item is a concrete entity. It is in many instances a single physical object (e.g., a copy of a one-volume monograph, a single audio cassette, etc.). There are instances, however, where the entity defined as item comprises more than one physical object (e.g., a monograph issued as two separately bound volumes, a recording issued on three separate compact discs, etc.) (IFLA, 2009, p. 24).

In terms of intellectual content and physical form, an item exemplifying a manifestation is normally the same as the manifestation itself. However, variations may occur from one item to another, even when the items exemplify the same manifestation, where those variations are the result of external actions to the intent of the manifestation producer (e.g., damage occurring after the item was produced, binding performed by a library, etc.) (IFLA, 2009, p. 24).

The item concept is integrant of the WEMI-Model (Work, Expression, Manifestation and Item model – see definition also in this glossary), which tries to identify the core aspects of publications and is the foundation of the FRBR family (IFLA, 2017).

Journal

A journal is a type of a periodical (see definition ahead) that corresponds to the serial and continuous scholarly publication under the same title released in regular time intervals for unlimited period, which contains articles written by researchers, professors and other experts intended for an academic or technical audience. Generally, a journal is identified by an ISSN number and each issue is identified by a consecutive number, and besides, is focused on a specific discipline or subject area (CUNHA, CAVALCANTE, 2008, p. 279).

Journal Impact factor (IF)

The Journal Impact Factor is a scientometric index which was developed in 1961 by Eugene Garfield and is defined as the total number of citations received by a specific journal indexed in the Web of Science Core Collection in the previous two years, divided by the total number of citable works (that comprise articles, reviews, and proceedings papers) published by that journal in the same period. Although the citing works may be articles published in the same evaluated journal, most of them are from different journals, proceedings, or books indexed in Web of Science Core Collection. A Journal Impact Factor of 1.0 means that, on average, the articles published one or two years ago have been cited one time. A Journal Impact Factor of 2.5 means that, on average, the articles published one or two years ago have been cited two and a half times (CLARIVATE ANALYTICS, 2019?; WIKIPEDIA, 2019). Even not being an exact mathematical average, the Journal Impact Factor provides a functional approximation of the mean citation rate per citable item (CLARIVATE ANALYTICS, 2019?) but, however, its use has been criticized and considered as a reductive and dangerous metric. It has also been pointed that “Journals do not calculate their impact factor directly — it is calculated and published by Thomson Reuters” (TIME…, 2016). “Journal Impact Factors were designed to indicate the quality of journals, but researchers often use the metrics to assess the quality of individual papers — and even, in some cases, their authors” (CALLAWAY, 2016). It has also been stated that although impact factor has been developed with the specific purpose to measure the impact of scientific journals it has lately been used for measuring the quality of scientific journals, individual journals and even of individual researchers and to evaluate grant applications and other financial supports to research programs. Because of that, the impact factor has been considered as an unreliable instrument for measuring the quality of journals what may cause unfairness (EASE, 2007).

List of references

See bibliographic references list.

Link rot

Hyperlinks that point to a website, a server or any other source that is no longer available through that link.

Long quotation

See block quotation

FRBR Manifestation

“The physical embodiment of an expression of a work” (IFLA, 2009, p. 21).

The entity defined as manifestation encompasses a wide range of materials, including manuscripts, books, periodicals, maps, posters, sound recordings, films, video recordings, CD-ROMs, multimedia kits, etc. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form (IFLA, 2009, p. 21).

When a work is realized, the resulting expression of the work may be physically embodied on or in a medium such as paper, audio tape, video tape, canvas, plaster, etc. That physical embodiment constitutes a manifestation of the work. In some cases, there may be only a single physical exemplar produced of that manifestation of the work (e.g., an author's manuscript, a tape recorded for an oral history archive, an original oil painting, etc.). In other cases, there are multiple copies produced to facilitate public dissemination or distribution. In those cases, there is normally a more formal production process involved, and a publisher, takes responsibility for the process. In other cases, there may be only a limited number of copies made of an original exemplar for purposes such as private study (e.g., a dubbing of an original recording of a piece of music), or preservation (e.g., a photocopy produced on permanent paper of an author's original typescript) (IFLA, 2009, p. 21-22).

The manifestation concept is integrant of the WEMI-Model (Work, Expression, Manifestation, and Item model – see definition also in this glossary), which tries to identify the core aspects of publications and is the foundation of the FRBR family (IFLA, 2017).

A manifestation represents all the physical objects that bear the same characteristics of intellectual content and physical form. A manifestation is itself an immaterial entity, but describes and represents physical entities, that is all the items that have the same content and carrier. So, writing a bibliographic record of this work (cataloging), is the same as describing the manifestation and this bibliographic record may be shared with other libraries (or information units) which also own a copy of the same manifestation (TILLET, 2003).

Mention

The reproduction of the idea of a cited work under the interpretation and on the own words of the author of a citing work. Although do not require a graphic highlight in the text body like indents or quotes, mentions also must create a citation, considering it demands an indication of an in-text reference pointer and its respective bibliographic reference that must appear in the bibliographic references list.

Original communication

An original communication is a specific kind of article which contains the discussion and the unpublished results of a research.

Periodical

A publication that corresponds to the serial and continuous scholarly publication under the same title released in regular time intervals for unlimited period (CUNHA, CAVALCANTE, 2008)

Publication

A work, which may be edited or not, offered to the public with the consent of the author or of the copyright holder by any way or process (CUNHA; CAVALCANTI, 2008).

Publisher

An individual or organization that produces and markets creative works for distribution (PEARCE-MOSES, 2005). The entity responsible for publishing works. In some cases, the publisher may be a company of the editorial branch, in other cases it may be the author himself.

Quartile

> Based on Impact Factor (IF) data, the Journal Citation Reports published by Thomson Reuters provides yearly rankings of science and social science journals, in the subject categories relevant for the journal (in fact, there may be more than one). Quartile rankings are derived for each journal in each of

> its subject categories according to which quartile of the impact factor distribution the journal occupies for that subject category. Q1 denotes the top 25% of the impact factor distribution, Q2 for middle-high position (between top 50% and top 25%), Q3 middle-low position (top 75% to top 50%), and Q4 the lowest position (bottom 25% of the impact factor distribution)
>
> Unfortunately, papers cannot be easily associated to a single ISI subject category (at least, not always), and one has therefore to consider the full range of quartile rankings of the journal. Following this line, a quartile score (indeed, a discrete distribution) is associated to any paper published in IF-ranked journals by uniformly distributing a unitary mass over the quartile rankings of the journal in which the paper was published (for that year) (FONDAZIONE BRUNO KESSLER, c2012).

Quotation

A quotation results from the act of quoting and corresponds to an exact transcription (copy, reproduction) of part of an external cited work (even published or not) within the text body of the citing work. Any way of expressing the knowledge may be quoted, what means that not only written works may be quoted but also, speeches, music, images, and any other way of representing and expressing information. Quotations are more commonly used in scientific texts for grounding an argumentation, but they also may be useful to present counterpoints, questions, confirmations, agreements, contradictions, and complements. Quotations may also be classified into block quotations and run-in quotations. See definitions separately for more details.

Quotation marks

Quotation marks are punctuation signals (“”) put in the beginning and in the end of run-in quotations to delimit the specific length of the transcription of a cited work.

Quotes

See quotation marks

Quote marks

See quotation marks

Reference

> the word "reference" is colloquially used to mean many things: either the bibliographic reference itself, or the entry in the body text of an article that denotes such a reference, or the act of citing the target publication, or the actual target publication itself (as in "Have you read that reference yet?"). The word also has variety of other meanings, and in particular is widely used in academia to mean a statement about a person's achievements, qualifications, competence and character, supplied, for example, in support of a job application or an academic promotion (PERONI, et al., 2015, p. 255).

The concept of reference is something broad. In the most general context, the most common definition of reference is the act of providing complimentary information by mentioning another instance (someone or something). Generally, a reference occurs by indicating an object, a person, a place, or anything else in a concrete or in an abstract sphere, by means of a sign, a word, a number or any other means of representing the indicated object (e.g., during his speech, the director will refer to different branches of acting. The professor told the students to do the exercises she referred to in the last class).

In a very general scientific approach, the concept of reference regards to a practice to lead the reader to specific complimentary information. In a more accurate view, references may assume various significances, from which two are particularly related to the scientific sphere:

a) Reference as a mention instance: a way to refer the reader to a specific piece of information, which may be within the text itself or in an external source of information. One of the functions of referencing in scientific context is to offer the reader the opportunity to be provided with more accurate, complimentary, or related information about a topic. It is quite useful to elucidate or substantiate a discourse, or for adding detailed data about something. In these cases, the author needs to mention other instances, which may be a publication, a place, another specific part of his own text… etc., in order to include, complement, or simply to registry the existence of an information source, or other specific issue (e.g. Aspirin is a Nonsteroidal Anti-inflammatory Drug indicated to relieve pain, fever, and inflammation (find complimentary information about this and other drugs and chemicals at PubChem database); or: "this issue will be discussed on chapter 5"). It is also true that sometimes a reference demands the indication of a bibliographic reference regarding to it in the bibliographic references list;

b) reference as a pointer instance: reference also may be applied in scientific works to indicate the origin (source) of an information. For example, when an author reproduces someone else's ideas in his own work, the proper way of informing the reader that that specific piece of information is not his authorship is using a reference to the original source of that specific information. This specific single information identifier, which is called in-text reference pointer, is also a reference, since it refers to other information e.g., “Computational ontologies are a means to formally model the structure of a system, i.e., the relevant entities and relations that emerge from its observation, and which are useful to our purposes” (GUARINO; OBERLE; STAAB, 2009, p. 2). It is important to point out that the reference is only the piece of information contained between the parentheses, which indicate the metadata for the reader to find more information about the publication mentioned (cited) at the bibliographic references list at the end of the publication, that is, the author's surnames, the date of the publication and the pages where the cited excerpt may be found.

   Still considering references as a pointer instance, it is proper to say that they can also be used to indicate graphics or figures within the text body, web links, annexes, appendices, glossaries, indexes, tables, a notation for a footnote, and any other element which the author considers may be of interest to the reader (e.g., “the compilation of the data gathered from the analysis may be consulted in annex 2”, or “the graphic 1 shows the distribution of the population, by age”).

At last, considering that the term “reference” has multiple definitions with multiple interpretations and uses, it is useful to remember that a work may comprises lots or references, in-text pointers and bibliographic references included.

For the purposes of this work, the terminological approach that will be considered for the term reference is an indication that forwards the reader to a specific excerpt in the text itself or to other information sources regarding to a subject (CUNHA; CAVALCANTI, 2008).

Reference management software

Dedicated application to automatically organize, manage and format bibliographic references into different bibliographic styles, according to the options selected by the user. Zotero, EndNote (and Endnote Web), Mendeley, Papers and Refworks are some examples of reference management softwares.

Restricted access

Basically, restricted access publications are those whose access is opposite to open access. Restricted access means the online availability of metadata for academic and/or scientific content, whose full access is granted through a user's counterpart, which can be a registration, an individual or institutional subscription or any other form of payment for access or, access control validated by an IP address.

Run-in quotation (simple run-in quotation)

A Run-in quotation, also called simple run-in quotation, is a textual quotation that is included inline and is usually enclosed within quotation marks. (SHOTTON, PERONI, 2015). The length limit for a quotation to be classified as a run-in quotation or as a block quotation is determined by the bibliographic style in use.

Standard

Set of established rules resulted from a particular normalization effort which, after being approved by an established institution, takes the format of a document containing a set of requirements to be attended to establish specific conditions to the realization of an activity as well as provide a background for the establishment of derived standards (CUNHA; CAVALCANTI, 2008).

Style manual

See bibliographic style.

Subject category

A topic subdivision of a general area of Science according to the context of the catalog or database it is applied in (CLARIVATE ANALYTICS, 2019?).

FRBR Work

Work is a conceptual entity that stands for the mental or artistic activity from which results the product of the intellectual or artistic creation. And as an abstract concept, a work is neither physical nor palpable. That is to say that its “formal existence” is necessarily conditioned to the registration in a physical or digital support. Therefore, a work becomes feasible through “individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work” (IFLA, 1997, p. 17). The materialization of a work occurs through its expression that can occur in many ways: written, spoken, sung, painted, or converted into works of art, as in the case of sculptures, and other forms of intellectual production. These expressions must be embodied in one or more manifestation, that is, it must be registered in a physical or digital support to make possible its management and retrieval. Usually, these embodiments also are multiplied in various items, so that it may be distributed in the most convenient form according to the context and other interests, including the editorial ones.

The work concept is integrant of the WEMI-Model (Work, Expression, Manifestation, and Item model – see definition also in this glossary), which tries to identify the core aspects of publications and is the foundation of the FRBR family (IFLA, 2017).

APPENDIX I – Raw and aggregated data of the study

The raw and aggregated data supporting this research are available as a dataset in Zenodo Platform, entitled “Raw and aggregated data for the study introduced in the doctoral thesis "FRBRizing bibliographic references structures: an approach from in-text references pointers and SPAR Ontologies for describing bibliographic metadata", available from Santos (2021).

The dataset is composed of two main Excel files: the first one, containing all the raw and aggregated data supporting this research and, the second, containing exclusively the raw and aggregated data concerning the analysis of the bibliographic elements composing the 53461 bibliographic references included in the bibliographic references lists of the 729 articles composing the sample. The second Excel file, entitled “Chart 14 - Starred metadata set”, in fact, corresponds to the Chart 14 of this research, as the title itself suggests.

The first Excel file, entitled “Thesis_Raw_Data”, contains the completely raw and aggregated data supporting this research. This dataset is composed of 30 sheets, whose contents are explicitly clarified below:

a) **Sheet “Journals Selection”** – contains the list of journals and articles selected for sample composition according to the criteria established in Chapter 6 – The methodology. Strikethrough contents mean the selected journals titles for which we could not obtain access and therefore, were replaced by the next journal in the SCImago rank attending the criteria for journals selection.
b) **Sheet “Articles Data Compilation”** – contains the raw and aggregated data gathered from the analysis of the journals and articles composing the sample.
c) **Sheet “Chart 14 - Starred Metadata Set”** – contains the raw and aggregated data gathered from the analysis of the bibliographic references composing the bibliographic references lists of the articles composing the sample. This analysis supported the establishment of the “starred metadata set” and corresponds to Chart 14 of this study.
d) **All the 27 remaining sheets** – The 27 remaining sheets whose titles correspond to the 27 subject areas considered in this study, contain the raw and aggregated data from the analysis of articles in each particular discipline corresponding to the respective sheet.

The second Excel file, entitled “Chart 14 - Starred Metadata Set” is a copy of the Sheet addressed in the first Excel file, that has the same name. As previously approached, this sheet contains the Chart 14, which was separately considered in this file, since it directly integrates the content of the text body of the Session 7.3.2.

ANNEX I – The FaBiO Ontology

Since the content of the FaBiO Ontology may be updated and, considering its huge extension, the version of the Ontology considered for the analysis supporting this study is available at the dataset Raw and aggregated data for the study introduced in the doctoral thesis "FRBRizing bibliographic references structures: an approach from in-text references pointers and SPAR Ontologies for describing bibliographic metadata", in the file entitled “ANNEX I - FaBiO, the FRBR-aligned Bibliographic Ontology”. The file represents the content of FaBiO Ontology as it was available in Peroni and Shotton (2019), in May 11th, 2021.