

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

THAMYRES VIEIRA DOS SANTOS

**Identificadores persistentes: aplicabilidade na organização e acesso à informação
científica**

São Paulo

2021

THAMYRES VIEIRA DOS SANTOS

Identificadores persistentes: aplicabilidade na organização e acesso à informação científica

Versão Corrigida

(versão original disponível na Biblioteca da ECA/USP)

Dissertação apresentada à Escola de Comunicações e Artes da Universidade de São Paulo para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Cultura e Informação

Orientadora: Profa. Dra. Giovana Deliberali Maimone

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Santos, Thamyres Vieira dos
Identificadores persistentes: aplicabilidade na
organização e acesso à informação científica / Thamyres
Vieira dos Santos; orientadora, Giovana Deliberali
Maimone. - São Paulo, 2021.
175 p.: il.

Dissertação (Mestrado) - Programa de Pós-Graduação em
Ciência da Informação / Escola de Comunicações e Artes /
Universidade de São Paulo.
Bibliografia
Versão corrigida

1. Identificadores persistentes. 2. Informação
científica. 3. Recuperação da informação. I. Deliberali
Maimone, Giovana. II. Título.

CDD 21.ed. - 020

SANTOS, T. V. Identificadores persistentes: aplicabilidade na organização e acesso à informação científica. 2021. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2021.

Aprovado em:

Banca Examinadora

Prof. Dr.: Giovana Deliberali Maimone (Presidente da banca)

Instituição: Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP)

Julgamento: _____

Prof. Dr.: Marivalde Moacir Francelin

Instituição: Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP)

Julgamento: _____

Prof. Dr.: Rogério Aparecido Sá Ramalho

Instituição: Universidade Federal de São Carlos (UFSCar)

Julgamento: _____

Prof. Dr.: Edberto Ferneda

Instituição: Universidade Estadual Paulista (UNESP)

Julgamento: _____

*Aos pesquisadores que discutem alternativas
para otimizar o tempo de pesquisa de seus pares.*

AGRADECIMENTOS

À Deus, em primeiro lugar, por me guiar em todos os desafios aos quais me condiciono, porque Dele e para Ele são todas as coisas. Agradeço também à minha família, por ser suporte ao longo do caminho.

À minha orientadora, Profa. Dra. Giovana Deliberali Maimone, por confiar e acreditar em mim desde o princípio. Sua paciência e generosidade são admiráveis. Agradeço por me incentivar de tantas formas e por possibilitar ricas experiências em parceria, desde publicações até o estágio em docência. Iniciar a trajetória acadêmica em sua companhia foi uma grata surpresa e, sem dúvida, uma grande inspiração para mim.

Ao Prof. Dr. Marivalde Moacir Francelin, por ser uma referência para mim desde a graduação. Agradeço a receptividade e incentivo desde o início da pós-graduação, quando ainda era aluna especial, e pelas contribuições durante a banca de qualificação. Sua trajetória como pesquisador e docente são inspiradoras para mim, que estou no início deste caminho.

Ao Prof. Dr. Rogério Aparecido Sá Ramalho, pelas contribuições durante a banca de qualificação e pela generosidade em apontar valiosos caminhos para a pesquisa, durante o mestrado e para além dele.

Aos docentes do Programa de Pós-graduação em Ciência da Informação, por me possibilitarem tantos aprendizados ao longo dessa trajetória. Em especial, ao Prof. Dr. Francisco Carlos Paletta, pelas palavras de incentivo e confiança que fazem a diferença nessa jornada, e ao Prof. Dr. Rogério Mugnaini, pelo valioso aprendizado proporcionado por suas disciplinas.

Às bibliotecárias da ECA, Lilian Viana e Marina Macambyra, pela disponibilidade e ajuda em conseguir literatura científica. Mesmo durante uma pandemia e com as bibliotecas fechadas, agradeço por sempre me oferecerem alternativas para seguir com a pesquisa.

À Edna Terezinha Rother, por ser minha grande mentora no âmbito profissional e acadêmico, possibilitando que essas duas vertentes da minha vida caminhassem juntas neste momento. Sua trajetória é admirável e sem sua confiança e apoio, este trabalho não teria acontecido. Todos os meus interesses de pesquisa são frutos do seu incentivo e conhecimento transmitido. Sou grata por me lançar em tantos desafios e propiciar que eu encontrasse meu caminho na Biblioteconomia e Ciência da Informação.

À Kioko Kusuki de Oliveira, pelo constante incentivo e pela paciência com os ajustes nos meus horários malucos de trabalho.

À Sheila Maria da Silveira, pela força e confiança de sempre, além da disponibilidade em revisar este trabalho. Agradeço também por me ajudar a segurar profissionalmente alguns

pratinhos, quando eles já estão quase caindo. Conversar e trocar figurinhas com você me traz um enorme aprendizado, sempre.

Ao amigo Leonardo Gonçalves Silva, pela parceria e força ao longo deste e de todos os outros caminhos da vida. Agradeço a disponibilidade em revisar este trabalho. Que sigamos juntos em tantos outros desafios, um empurrando ao outro, sempre.

À amiga Regiane Alves de Assis, por ser sempre um dos maiores apoios em todas as decisões da vida, incluindo as acadêmicas. Sou grata pela força e pela alegria que sempre transmite. Agradeço também por ler este trabalho e pelas palavras de confiança.

Ao Ms. José Mário de Oliveira Mendes, pela confiança e por acreditar, desde o princípio, que eu poderia seguir no caminho da pesquisa. Foi um privilégio ter sido sua aluna na especialização e aprender tanto com sua generosidade em sala de aula.

Enfim, agradeço a todos que contribuíram com o desenvolvimento desta pesquisa, direta ou indiretamente.

Permanecemos a maior parte do tempo no mundo prático de nossos códigos restritos. Se nos afastássemos dele o tempo todo, tornar-nos-íamos literalmente loucos (FOUREZ, 1995, p.22).

RESUMO

SANTOS, Thamyres Vieira dos. **Identificadores persistentes: aplicabilidade na organização e acesso à informação científica**. 2021. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2021.

Discorre sobre o uso dos identificadores persistentes na organização e acesso à informação científica em ambiente digital. Enquanto principal insumo para desenvolvimento da ciência, a informação científica deve ser alvo de processos de representação e organização que propiciem uma recuperação adequada em fontes de informação, garantindo a visibilidade da pesquisa, a de seus autores e suas instituições. Dessa forma, o objetivo é analisar os identificadores persistentes, que são códigos alfanuméricos, como estratégia de representação, organização e acesso à informação científica em ambientes digitais. Por meio de embasamento teórico, realiza-se uma discussão da aplicabilidade dos identificadores persistentes baseada em suas principais características, evidenciando seu papel na otimização do processo de recuperação informacional, embora um único padrão para autores ou instituições ainda não seja amplamente utilizado e difundido na comunidade acadêmica. A análise da utilização de seis identificadores, entre autorais e organizacionais, em 39 bases de dados ou plataformas de conteúdos científicos, em três repositórios de dados de pesquisa e oito repositórios de *preprints* multidisciplinares, mostra que as iniciativas ainda não são largamente adotadas e a maioria das fontes não apresentam opções de busca por identificadores. Sendo assim, é fundamental a realização de pesquisas que discutam a utilização dos identificadores persistentes e evidenciem a necessidade de fortalecimento de iniciativas únicas, sejam elas autorais ou organizacionais, propiciando uma ampla compreensão e utilização por parte de pesquisadores e instituições, consolidando-se como facilitadores no acesso à informação científica, sobretudo em ambiente digital.

Palavras-chave: Identificadores persistentes. Informação científica. Recuperação da informação.

ABSTRACT

SANTOS, Thamyres Vieira dos. **Identificadores persistentes: aplicabilidade na organização e acesso à informação científica**. 2021. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2021.

Discusses the use of persistent identifiers in the organization and access to scientific information in a digital environment. As the main input for the development of science, scientific information must be the target of processes of representation and organization that provide an adequate retrieval at information sources, guaranteeing the visibility of the research, as authors and institutions. In this sense, we intend to analyze persistent identifiers, which are alphanumeric codes, as a strategy for organizing and representing information in digital environments for accessing information. Using a theoretical basis, there is a discussion of the applicability of persistent identifiers based on their main characteristics, highlighting their role in optimizing the information retrieval process, although a single standard for authors or institutions is not yet used and disseminated in the academic community. The analysis of the use of six identifiers, between author and institutional identifier, in 39 databases or scientific content platforms, in three research data repositories and eight multidisciplinary preprint repositories, shows that initiatives are still not widely adopted and most sources do not present options for search for identifiers. Therefore, it is essential to carry out research that discusses the use of persistent identifiers and highlights the need to strengthen unique initiatives, for authors or organizations, providing a broad understanding and use by researchers and institutions, consolidating themselves as facilitators of access to scientific information in a digital environment.

Keywords: Persistent identifiers. Scientific information. Information retrieval.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplos de identificadores persistentes e sua utilização prática.....	86
Figura 2 – Conteúdo informacional de dados e metadados.....	90
Figura 3 – Anatomia de um identificador baseado na web	93
Figura 4 – Proliferação de "padrões"	96
Figura 5 – Citação imaginária para uma publicação fazendo o uso completo de identificadores persistentes.....	97
Figura 6 – Estrutura do ISBN	99
Figura 7 – Itens que compõem o DOI	100
Figura 8 – Identificação de autoria com nomes iguais	102
Figura 9 – Base de busca de registros no ISNI.....	115
Figura 10 – Busca realizada na base do ISNI.....	116
Figura 11 – Resultado de busca na base do ISNI	117
Figura 12 – Identificador ORCID atrelado ao registro ISNI.....	118
Figura 13 – Exemplo do conjunto de informações de um registro Ringgold.....	120
Figura 14 – Busca por instituição na base GRID	122
Figura 15 – Perfil da Universidade de São Paulo na base GRID	122
Figura 16 – Visualização de listagem de instituições da GRID no Excel	123
Figura 17 – Busca por instituições no ROR	124
Figura 18 – Exemplo de perfil do pesquisador no ORCID	127
Figura 19 – Integrações do ORCID com redes, bases de dados e plataformas de comunicação científica.....	129
Figura 20 – Menu dropdown com organizações no ORCID.....	130
Figura 21 – Exemplo de exibição do Ringgold ID e ISNI no registro ORCID	131
Figura 22 – Autorização de acesso à plataforma externa no ORCID.....	132
Figura 23 – Exemplo imaginário de publicação científica com autores que possuem perfil ORCID.....	132
Figura 24 – Exemplo de perfil do pesquisador na base de dados Web of Science	135
Figura 25 – Exemplo de perfil do pesquisador no Publons.....	136
Figura 26 – Exemplo de perfil do Scopus ID do pesquisador	138

LISTA DE TABELAS

Tabela 1 – Quantidade de bases de dados com opção de busca por identificadores	140
Tabela 2 – Quantidade de repositórios de dados de pesquisa com opção de busca por identificadores	142
Tabela 3 – Quantidade de repositórios de preprints com opção de busca por identificadores	143

LISTA DE QUADROS

Quadro 1 – Orientações das normas sobre a inserção dos autores do documento	56
Quadro 2 – Opções para identificar e avaliar autores científicos.....	82
Quadro 3 – Exemplos de identificadores utilizados na comunidade científica.....	86
Quadro 4 – Panorama do uso de identificadores para diferentes recursos de pesquisa	89
Quadro 5 – Características desejáveis para identificadores	93
Quadro 6 – Características dos identificadores persistentes para sua análise de aplicabilidade na organização e acesso à informação científica	112
Quadro 7 – Vantagens e limitações do Open Researcher and Contributor ID (ORCID).....	133

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Problemas de pesquisa	16
1.2	Hipóteses de pesquisa.....	16
1.3	Objetivos	17
1.3.1	Objetivo geral	17
1.3.2	Objetivos específicos	17
1.4	Metodologia de pesquisa.....	17
2	ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO CIENTÍFICA EM AMBIENTE DIGITAL.....	19
2.1	Fontes de informação científica	25
2.1.1	Repositórios	28
2.2	Objetos digitais.....	38
3	RECUPERAÇÃO E VISIBILIDADE DA PESQUISA DO PESQUISADOR E DAS INSTITUIÇÕES NA COMUNIDADE CIENTÍFICA.....	42
3.1	Controle de autoridade e atribuição autoral	46
3.2	Normas bibliográficas e manuais metodológicos	52
3.3	Uso da tecnologia para desambiguar autorias e instituições.....	61
3.4	Impacto nos indicadores métricos científicos	69
3.5	Redes acadêmicas em ambiente digital: espaço para manutenção da visibilidade .	74
3.6	Autenticidade e autoridade de conteúdos em ambiente digital: desafios da comunidade científica	78
4	IDENTIFICADORES PERSISTENTES	84
4.1	Identificadores bibliográficos.....	98
4.2	Identificadores autorais.....	101
4.3	Identificadores organizacionais	107

5	ANÁLISE E APLICABILIDADE DE IDENTIFICADORES ORGANIZACIONAIS E AUTORAIS	111
5.1	Iniciativas	112
5.1.1	International Standard Name Identifier (ISNI)	113
5.1.2	Ringgold Identifier	119
5.1.3	Global Research Identifier Database (GRID)	121
5.1.4	ORCID	126
5.1.5	ResearcherID	134
5.1.6	ScopusID	137
5.2	Opções de busca em fontes de informação científica	139
5.2.1	Bases de dados	140
5.2.2	Repositórios de dados de pesquisa	141
5.2.3	Repositórios de <i>preprints</i>	142
5.3	Discussão geral da análise realizada	143
6	CONSIDERAÇÕES FINAIS	146
	REFERÊNCIAS	152
	APÊNDICE A – Análise de bases de dados	170
	APÊNDICE B – Análise de repositórios de dados de pesquisa	174
	APÊNDICE C – Análise de repositórios de <i>preprints</i>	175

1 INTRODUÇÃO

A informação é o principal ativo das organizações contemporâneas, sendo utilizada como insumo para tomada de decisão e planejamento de ações que resultem em otimização de suas práticas e melhor posicionamento em seu segmento de atuação. O desenvolvimento das organizações por meio de práticas de gestão da informação e conhecimento é pauta recorrente entre os especialistas de mercado e está presente nas discussões acadêmicas. Assim como a informação se mostra fundamental nas corporações, a informação científica é o principal insumo para desenvolvimento da ciência, a qual é referência para as discussões tanto no meio organizacional quanto fora dele.

Além de enfrentar o desafio de gerenciar suas informações, a sociedade e as organizações, científicas ou não, lidam com o contexto amplamente digital no qual todas as suas atividades, insumos e dados estão inseridos. As informações científicas são frutos dos esforços de organizações científicas, universidades e centros de pesquisa, e podem ser localizadas via bases de dados bibliográficas, catálogos de serviços de informação ou acervos físicos e estão voltadas sobretudo para a comunidade científica, embora também existam iniciativas de divulgação científica que visam tornar o conteúdo produzido acessível à população, que também deve conhecer os meios para acessar informações confiáveis e certificadas pela academia e delas se apropriar.

As fontes desse tipo de informação são conhecidas e acessadas pela comunidade científica mas, ainda assim, nem todos os conteúdos são recuperados com eficácia e assertividade, não garantindo que os acadêmicos se sintam seguros quanto à qualidade de suas buscas.

A recuperação da informação científica perpassa diversas discussões e nuances, sejam elas derivadas da variação da indexação das bases de dados até o próprio manuseio das ferramentas de busca. Diante de tantas possibilidades advindas do desenvolvimento das tecnologias de informação e comunicação (TICs), é essencial que os conteúdos científicos, sua organização, representação, recuperação e utilização também sejam contemplados em estudos que tenham como objetivo otimizar sua disseminação, tendo em vista o desenvolvimento científico. É fundamental reconhecer que a disseminação de conteúdos científicos, sobretudo por intermédio de revistas científicas eletrônicas disponibilizadas por bases de dados, se torna mais dinâmica, e essa característica contribui consideravelmente para o acesso e desenvolvimento da ciência.

Diante do grande e desordenado volume de informações disponíveis em meio digital, é fundamental a reflexão sobre técnicas e ferramentas que se proponham a organizar e facilitar o acesso a esses conteúdos, minimizando a propagação de informações falsas, não certificadas e descontextualizadas. A *web* facilita o acesso à informação, que se torna mais dinâmica e com amplo alcance, características fundamentais no cenário científico, considerando que a informação mais recente em tempo adequado impacta diretamente no desenvolvimento de pesquisas, tecnologias e políticas públicas. Contudo, o acesso descomplicado e rápido também resulta na disponibilidade de milhares de conteúdos que acabam sendo disponibilizados em rede de forma desordenada ou não padronizada.

Considerando o cenário científico, as plataformas, bases de dados, repositórios e outros ambientes com a proposta de reunir essas informações, se mostram adeptos de técnicas e tecnologias que buscam otimizar a recuperação desses conteúdos pelo pesquisador, contribuindo para que a informação relevante chegue até o usuário de forma assertiva.

Diversas bases de dados de diferentes áreas do conhecimento adotam vocabulários controlados com o objetivo de sistematizar a organização de seus conteúdos e contribuir para o fácil acesso do pesquisador, desenvolvendo tecnologias para reunir publicações de um mesmo autor ou até oferecem aos seus usuários a oportunidade de solicitar correções quando essas ferramentas falham por algum motivo.

A forma de organização da informação científica em ambiente digital também remete aos profissionais da informação e aos membros da comunidade científica à problemas que já eram relativamente comuns nos suportes físicos da informação como, por exemplo, a dificuldade em localizar publicações por nomes de autor ou instituição devido à inexistência ou existência ineficaz de um controle de autoridade, ou seja, diferenciar pesquisadores e instituições homônimas, questão amplamente discutida também no contexto da indexação em catálogos de bibliotecas, arquivos ou centros de documentação.

No caso das publicações em ambiente digital, a quantidade de conteúdo disponível aumenta consideravelmente, seja na forma de artigos ou de periódicos, tornando a tarefa de buscar determinadas publicações mais árdua, sendo necessário haver um conjunto de informações que possibilitem ao pesquisador a certeza de que encontrou o artigo que buscava como por exemplo, título, autores, periódico, volume, número e paginação.

Em primeiro lugar, pode-se inferir que o controle de autoridade realizado em suportes físicos de informação era de responsabilidade do serviço em questão, seja ele uma biblioteca, serviço de catalogação na fonte, editoras, entre outros profissionais envolvidos. Contudo, em

meio digital, a informação científica perpassa e é disseminada por diferentes profissionais: indexadores das bases de dados, periódicos científicos e os próprios autores. É essencial considerar uma melhor alternativa para a consolidação e recuperação das afiliações desses autores e instituições de forma eficaz, evitando desperdício de tempo do usuário ao tentar sanar suas necessidades informacionais por meio da utilização de recursos no ambiente tecnológico.

Nesse contexto, a utilização de identificadores persistentes se apresenta como uma proposta que visa otimizar a recuperação de informações científica em ambiente digital. Os identificadores são códigos alfanuméricos que podem ser utilizados para diferenciar publicações, autores e instituições, contribuindo para a recuperação eficaz de resultados de pesquisa sem que sejam afetados pelas variações de nomenclatura. Existem várias iniciativas advindas de organizações com e sem fins lucrativos, vinculadas ou não a grandes grupos comerciais.

1.1 Problemas de pesquisa

- a) Como e com qual objetivo os identificadores persistentes podem ser utilizados em prol da recuperação e acesso da informação científica em ambientes digitais?
- b) Os identificadores persistentes possuem características e requisitos técnicos que impactam diretamente na sua função na comunidade científica?
- c) Os identificadores persistentes otimizam a busca do pesquisador nas plataformas de conteúdos científicos?

1.2 Hipóteses de pesquisa

- a) Os identificadores persistentes facilitam a recuperação e acesso da informação científica em ambiente digital considerando sua assertividade conferida por meio do código alfanumérico;
- b) Os identificadores persistentes devem atender a requisitos técnicos desejáveis para que contribuam de forma assertiva com a interoperabilidade informacional e adesão da comunidade científica;

- c) Os identificadores persistentes otimizam a busca do pesquisador nas plataformas de conteúdos científicos, seguindo iniciativas da comunidade científica em promover a interoperabilidade informacional.

1.3 Objetivos

1.3.1 Objetivo geral

- a) Caracterizar a utilização dos identificadores persistentes como estratégia de organização e representação da informação em ambientes digitais de acesso à informação científica,

1.3.2 Objetivos específicos

- a) Caracterizar e analisar os identificadores persistentes, especificamente três identificadores persistentes organizacionais e três identificadores persistentes autorais, e discutir como proporcionam assertividade na busca, recuperação, acesso e visibilidade aos conteúdos científicos buscados em plataformas que os hospedam;
- b) Verificar quais as fontes de informação, entre bases de dados e repositórios, adotam estratégias como a possibilidade de busca ou filtro por identificadores persistentes em suas ferramentas de busca.

1.4 Metodologia de pesquisa

- a) Revisão bibliográfica:

O levantamento bibliográfico inicial foi realizado nas bases de dados Scopus, Web of Science, PubMed e BRAPCI. No decorrer das leituras, também foram consultadas as referências de cada um dos materiais e mais estudos pertinentes à temática foram encontrados e utilizados no decorrer dos capítulos que precedem à análise.

- b) Análise e aplicabilidade:

Considerando a caracterização e discussão a respeito dos identificadores persistentes organizacionais e autorais, foi realizada uma análise de três identificadores persistentes organizacionais e três identificadores persistentes autorais, por meio de verificação experimental das plataformas, subsidiada também por discussões encontradas na literatura científica. Os identificadores que foram objeto do estudo são: International Standard Name Identifier (ISNI); Ringgold Identifier; Global Research Identifier Database (GRID); ORCID; Researcher ID; e Scopus ID. A escolha por estes identificadores se deu devido à sua maior recorrência de discussão na literatura científica e por serem alvo de maior divulgação e utilização na comunidade científica.

Dessa forma, com a análise detalhada dos seis identificadores, foi realizada uma segunda análise em 39 bases de dados ou plataformas de conteúdos científicos, em 3 repositórios de dados de pesquisa e 8 repositórios de *preprints* multidisciplinares.

A segunda análise tem objetivo de verificar quais dessas fontes possuíam opções de busca, ou realização de filtros durante a pesquisa, por identificadores persistentes, especificamente pelos identificadores organizacionais e autorais.

Nesse sentido, a pesquisa inicia-se com uma revisão de literatura que buscou contextualizar a informação e comunicação científica e a forma pela qual ocorre sua organização na *web*, incluindo a discussão sobre as fontes de informação utilizadas na comunidade científica, incluindo o conceito de objeto digital.

Após essa breve contextualização, foi necessário discutir, partindo dos conceitos inicialmente levantados, como a recuperação da informação científica em ambiente digital impacta diretamente na visibilidade da pesquisa, do pesquisador e de suas instituições.

Dessa forma, foi possível caracterizar os identificadores persistentes, com principal foco nos organizacionais e autorais e discutir sua utilização antes de analisar especificamente seis deles quanto à sua utilização e presença em fontes informacionais.

2 ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO CIENTÍFICA EM AMBIENTE DIGITAL

No contexto acadêmico-científico o compartilhamento de informações em ambiente digital propicia um acesso horizontalizado aos pesquisadores do que é produzido em sua linha e área de pesquisa, ou seja, um acesso plural e com menos barreiras, beneficiando o desenvolvimento de novos estudos e contribuindo para o desenvolvimento da ciência.

Contudo, assim como na sociedade e nas organizações de forma geral, na comunicação científica o cenário digital impõe desafios ao mesmo tempo em que se apresenta como seu principal facilitador. De um lado tem-se a facilidade de transmitir conteúdo a pesquisadores com uma velocidade que fora do contexto digital seria inviável, enquanto do outro temos o volume crescente de informações que se não estiverem devidamente organizadas e representadas, não serão recuperadas pelo pesquisador.

A informação científica é insumo vital produzido pelas universidades, centros de pesquisa, organizações e indústrias para o desenvolvimento de novas técnicas e tecnologias que contribuam para a melhoria de processos e políticas que impactam diretamente na sociedade. Considerando o amplo desenvolvimento das TICs, a informação científica teve seu fluxo e dinâmica alterados, chegando mais rápido ao alcance dos pesquisadores. Para abordar as ferramentas e técnicas utilizadas que tornam mais acessíveis os conteúdos, é fundamental discorrer sobre os conceitos de informação em ciência e refletir a respeito de seu posicionamento e papel dialogando com a literatura da área de Ciência da Informação.

As discussões sobre informação ocorrem em diferentes áreas do conhecimento e o termo é por elas apropriado de acordo com as características intrínsecas a cada uma. A informação é comumente mencionada como um ativo de valor nas organizações, um objeto ao qual a gestão das empresas não se atentava e não conseguia visualizar seu valor enquanto insumo para o desenvolvimento interno. Targino menciona que “[...] em pleno século 21, a cultura do pós-modernismo impera no meio social e o eixo da economia passa a ser a produção, a distribuição e a difusão da informação” (TARGINO, 2007, p. 96). Considerando o posicionamento de destaque assumido pela informação, é natural que existam diversos autores que buscam defini-la.

McGarry (1999) menciona que a raiz do termo vem de *formatio* e *forma*, com a ideia de dar forma a algo. Por sua vez, Le Coadic (2004) caracteriza a informação como um significado transmitido por meio de uma mensagem inscrita em um suporte a um ser consciente.

As definições mencionadas possuem em comum a ideia de informação como algo que possui formato, significado e relevância. Partindo de tal pressuposto, a informação no cenário científico se torna fundamental para o desenvolvimento do conhecimento, intrínseco aos pesquisadores, que já refletiram sob a informação inicialmente recuperada e utilizada. Le Coadic (2004, p. 26) afirma que “[...] a informação é seiva da ciência. Sem informação, a ciência não pode se desenvolver e viver”. Está na informação a base do desenvolvimento científico; por meio dela pode-se verificar a evolução das práticas científicas que nunca se esgotam ou se tornam definitivas. Conforme aponta Targino:

No que se refere à ciência, a compreensão da **informação científica conduz à relevância da ciência como processo social, dinâmico, contínuo e cumulativo**, que busca, essencialmente, desvendar e compreender a natureza e os seus fenômenos, por meio de métodos sistemáticos e seguros. No entanto, por conta da dinamicidade intrínseca à própria natureza, os resultados são sempre provisórios. Isto é, os seus sistemas explicativos não são permanentes, inserem-se num processo ininterrupto de investigação, criando e alterando convicções, modificando hábitos, gerando leis, provocando acontecimentos, ampliando, de forma permanente e progressiva, as fronteiras do conhecimento – o que permite inferir, mesmo sem pretensões históricas, que a ciência influencia a humanidade e é por ela influenciada, o tempo todo (TARGINO, 2007, p. 96, grifo nosso).

O posicionamento central da informação científica no desenvolvimento das atividades de pesquisa e conseqüentemente no desenvolvimento da ciência é evidente. Com a disponibilidade desse conteúdo em ambiente digital nota-se sua potencialização em termos de velocidade de disseminação e alcance, considerando que para o desenvolvimento científico é fundamental que as informações sejam compartilhadas rapidamente e sem fronteiras, alcançando o pesquisador a informação desejada e em tempo assertivo.

Considerando que a informação possui papel central no desenvolvimento científico, sendo utilizada para produção de conhecimento pelos pesquisadores, pode-se afirmar que “[...] todo conhecimento começa por algum tipo de informação e se constitui em informação” de forma que “[...] o estoque de conhecimentos é alterado com o input de novas informações, através de adições, reestruturações ou mudanças” (CINTRA et al., 1994, p. 13).

A legitimação da informação científica ocorre em sua disseminação e conseqüente utilização por seus pares, ou seja, demais pesquisadores em suas respectivas áreas de atuação que utilizam aquele conteúdo para avançar ou iniciar pesquisas, as quais impactarão diretamente na sociedade; afinal, “[...] a ciência estimula e orienta a evolução humana, interfere na identidade dos povos e das nações, estabelece as verdades fundamentais de cada época”

(TARGINO, 2000, p. 2). Neste caso, quando a troca de conteúdos fica somente entre membros da própria comunidade científica, o processo é chamado de comunicação científica. Targino (2000, p. 10) afirma que “a comunicação científica é indispensável à atividade científica, pois permite somar os esforços individuais dos membros das comunidades científicas”.

A informação científica como insumo para produção de conhecimento é disponibilizada por meio de iniciativas de comunicação científica, voltadas para a comunidade acadêmica. Atualmente tais iniciativas consistem em bases de dados, plataformas, servidores, blogs, sites, redes sociais acadêmico-científicas ou também redes sociais de uso do público geral. Contudo, nem sempre todas as iniciativas de comunicação em ciência podem ser assertivas quanto à garantia de recuperação da informação baseada nas necessidades do pesquisador, pois o grande volume de informação disponível já se consolida como problema. Embora o contexto atual seja o da informação em ambiente digital, a preocupação com a sua disponibilidade e acesso nasce antes mesmo da consolidação e popularização dos meios digitais.

Após a Segunda Guerra Mundial, as atividades de pesquisa e ciência despertaram grande interesse e ocasionaram um aumento crescente de registros de conhecimentos em ciência e tecnologia, ocasionando o fenômeno conhecido como explosão informacional (OLIVEIRA, 2011, p. 11-12). Já as revistas científicas, consideradas como o meio mais ágil para veiculação de conhecimento científico, ainda que algumas áreas ainda possuam o livro como suporte de destaque, surgiram em 1665 em Paris e Londres (POBLACIÓN, 2011, p. 9; WITTER, 2011, p. 11) e desde então, surgiram muitas outras revistas na comunidade científica mundial em todas as áreas do conhecimento, sendo atualmente aproximadamente 40 mil (VEIGA, 2019).

É pertinente que o conhecimento científico seja produzido com afincamento e disseminado através de meios de divulgação reconhecidos na academia, como os periódicos científicos, para que as descobertas e reflexões documentadas possam contribuir para o desenvolvimento de novos conhecimentos. Mas a eficácia deste cenário se dá somente por meio da acessibilidade e recuperação eficaz desse conteúdo, utilizando técnicas que sejam capazes de otimizá-lo.

Mas, para que o conhecimento da sociedade não se perca e possa ser compartilhado, ele é registrado num dado suporte: livro, imagem, foto, disco, etc., passando a se constituir num documento. O desenvolvimento científico e tecnológico tem proporcionado à sociedade uma massa enorme de conhecimentos, portanto de documentos, que precisam ser tratados adequadamente, para que haja não só a sua divulgação, como também a criação de novos conhecimentos, cumprindo assim a rotina natural da própria ciência. Daí o papel fundamental da área de documentação, responsável pela triagem, organização, conservação da informação, bem como pela viabilização a seu acesso (CINTRA et al., 1994, p. 14-15).

Pauta-se a discussão sobre a organização e acesso à informação científica no conceito de “informação como coisa”, conforme sugere Buckland (1991) ao discutir a informação como evidência, mediante a qual é possível se informar, ou seja, pode ser um objeto como dado ou documento capaz de instruir de alguma forma. Desse modo, na construção do texto utiliza-se o termo “informação científica” enquanto dado ou documento que pode ser recuperado e utilizado pelo pesquisador de forma que lhe seja pertinente para o desenvolvimento de pesquisa. Ainda nesse sentido, entende-se que “a recuperação da informação é reconhecida como a recuperação de referências de documentos em resposta às solicitações (demandas expressas por informação)” (ARAÚJO JÚNIOR, 2007, p. 72).

Ao considerar a informação científica como indispensável para o desenvolvimento científico, é fundamental que os pesquisadores consigam acessar esse volumoso conjunto de dados e informações que indiquem o que já foi desenvolvido nas mais diferentes áreas de pesquisa, para que então se consiga avançar partindo de evidências já deixadas por seus pares e registradas em sua produção intelectual, esteja elas nas revistas ou em outro meio de comunicação científica. Assim, não é aceitável que o conhecimento científico não esteja acessível aos pesquisadores de forma que exista o risco de se repetir um experimento, técnica ou hipótese que demande tempo e investimentos de pesquisa, sendo que o resultado já foi alcançado por outro pesquisador, exceto os casos em que seja de fato necessária sua repetição por motivos relevantes ao desenvolvimento da pesquisa, pois “acreditam alguns que este fenômeno da ‘redescoberta’ possa se tornar um dos principais fatores limitadores da taxa de avanço da ciência na sociedade contemporânea” (CINTRA; TÁLAMO; LARA; KOBASHI, 1994, p. 16).

Em ambiente digital é perfeitamente possível que um pesquisador não consiga localizar alguma informação que lhe pouparia tempo e investimento de pesquisa. Berry (2012) discute que a forma de fazer pesquisa atualmente tem sofrido mudanças significativas, dentre elas está a mediação feita pela tecnologia, uma vez que dificilmente se encontrará um pesquisador que não utilize alguma ferramenta digital em suas atividades; o autor ressalta que, logicamente, a aplicabilidade e adoção dessa nova metodologia é particular de cada área ou subárea e essa mediação da tecnologia é objeto de estudo das chamadas Humanidades Digitais. Essa área, conforme afirma Kaplan (2015), propicia conexões entre as tecnologias e as humanidades, construindo diálogos entre a Ciência da Informação e as demais áreas das Ciências Humanas,

inclusive no contexto do *big data*, considerando a explosão informacional de conteúdos disponibilizados em ambientes digitais.

Hoje a busca de uma informação que seja uma resposta pertinente e relevante às nossas perguntas requer a reconstrução de um complexo cenário onde sejam agregadas as populações de fontes e canais de informação, de modo a permitir processos seletivos, organizados e econômicos de busca e recuperação (GONZÁLEZ DE GÓMEZ, 1993, p. 217).

Para que a informação científica esteja disponível em expressivo volume em ambiente digital, presume-se que sua produção ocorra também em intensa velocidade; e dessa forma, “a velocidade de produção de informação tem como consequência quase imediata a obsolescência de conhecimentos” (CINTRA et al., 1994, p. 16). É pouco provável que todos os pesquisadores consigam se manter atualizados diante de tantas informações novas em sua área de atuação e essa ampla disponibilidade se torna uma barreira ao pesquisador ao mesmo tempo em que brilha como a melhor das soluções.

Assim, claro está que ninguém pode, nem mesmo numa área de especialidade, aventurar-se a “conhecer” tudo o que se publica. Mas também é claro que uma pessoa pode conseguir informações parciais em níveis satisfatórios, graças aos meios desenvolvidos para guarda e recuperação da informação (CINTRA et al., 1994, p. 17, grifo do autor).

Meadows (1999, p. 211) discute que o cientista se depara o tempo todo com diversas informações, mas uma parcela considerável delas passa sem ser absorvida, uma vez que precisa de informação de fato “pertinente”, termo utilizado pelo autor entre aspas pois determinar pertinência pode ser muito amplo, uma vez que depende do contexto; “o requisito básico é que a informação seja fornecida no momento e na forma que torne mais provável sua absorção pelo pesquisador” (MEADOWS, 1999, p. 211). Moreira González (2011, p. 13) também afirma que “[...] os documentos nos proporcionam informação. Mas não geram conhecimento até que sua mensagem seja alcançada e assimilada pelo usuário”. A informação pertinente precisa ser acessível ao pesquisador no momento exato de sua busca, entendendo que aquele seja o momento em que sua necessidade informacional deve ser sanada para assimilação e geração de conhecimento, sendo necessário para que isto ocorra um trabalho sistemático por parte de profissionais responsáveis por gerenciar e organizar esses conteúdos.

A organização da informação é o processo geral que se dedica a fornecer acesso à informação de forma assertiva por meio da adequada elaboração e aplicabilidade de instrumentos

para seu tratamento. Lima e Alvares (2012, p. 35) afirmam que “o principal objetivo da organização da informação é recuperar objetos informacionais, que são as informações registradas nos mais variados suportes, [...] fornecer acesso à informação, por meio da estruturação dos elementos de organização do conhecimento”.

Enquanto o processo de representação da informação ocorre quando profissionais da informação extraem informações documentárias com as quais podem trabalhar para identificar tal documento, esses registros de identificação podem ser feitos por meio de linguagem natural como, por exemplo, resumo, autores, local e data, ou por meio de linguagem documentária, padronizada e que busca um padrão para recuperação mais eficaz dos documentos através da representação genérica de seu assunto ou temática. Conforme afirmam Lima e Alvares (2012, p. 36), “a representação não substitui o documento, mas possibilita identificar seus atributos fundamentais, os quais orientam o usuário para uma tomada de decisão sobre a necessidade de consulta do documento que originou a representação”. Diz ainda Alvarenga que:

A organização da informação ou conhecimento, que como já dito compreenderia um processo de representação, destina-se prioritariamente à recuperação eficaz por parte dos usuários. Para que tal ocorra torna-se necessário que profissionais da informação desenvolvam e implementem sistemas representacionais que estabeleçam a confluência entre a organização cognitiva imposta ao conhecimento pelo seu produtor (representação primária) e a organização conceitual imposta ao documento pelo especialista da informação (representação secundária) (ALVARENGA, 2003, p. 29).

A qualidade da representação da informação é imprescindível para uma recuperação assertiva de documentos científicos pelo pesquisador. Trata-se de uma atividade que demanda um processo muito bem alinhado de curadoria dos dados disponíveis sobre o documento, tanto em relação aos seus atributos físicos quanto em relação à representação de seu conteúdo, partindo do que foi exposto pelos autores do documento até o processo de padronização dos termos que serão oferecidos para busca pelo usuário.

A representação da informação científica não substitui o documento, mas auxilia o pesquisador no processo de escolha do *paper* que pode contribuir de forma efetiva com o desenvolvimento de sua pesquisa. A preocupação com a adoção e manutenção de técnicas para otimizar esses processos de representação é anterior à popularização das TICs, porém em ambiente digital é notável o quanto o volume de dados disponíveis impõe desafios ainda maiores a essa área, sobretudo porque os suportes e formas de se buscar, e inclusive de receber informações, científicas ou não, também se modificaram e trouxeram consigo a missão de

identificar de fato quais fontes, conteúdos e formas de buscá-los que se mostram mais eficientes.

2.1 Fontes de informação científica

São muitas as fontes que atualmente podem ser utilizadas para a pesquisa científica e o ideal é que o objetivo de todas elas seja contribuir para que a comunidade científica possa encontrar conteúdo relevante e pertinente para desenvolver suas pesquisas. Considerando a máxima *publish or perish*, na qual os pesquisadores são comumente bombardeados por intensas metas de publicação, então não surpreende que a comunidade esteja envolta por bases de dados comerciais que possuem diversos interesses além do desenvolvimento da pesquisa científica.

Algumas iniciativas como repositórios institucionais ou repositórios de dados e publicações *pre-prints*, buscam armazenar e garantir o direito de disponibilizar tudo o que foi produzido por pesquisadores de determinadas instituições ou artigos que ainda não foram oficialmente publicados, mas que já podem receber contribuições de outros membros da comunidade científica para sua otimização e que também já podem ser utilizados para o desenvolvimento de novas pesquisas, sem que seja necessário aguardar todo o processo editorial de submissão até o aceite de um artigo em periódicos científicos.

Sendo assim, todas essas fontes de informação científica estão à disposição do pesquisador e em parte delas estão aplicadas técnicas, em algumas mais ou menos assertivas, para garantir que a informação correta chegue ao usuário no momento adequado. É fundamental refletir, de forma geral, a respeito de cada uma dessas aplicações para que então seja possível avançar em novas propostas de organização desses conteúdos, sobretudo em ambiente digital.

É necessário para o pesquisador interessado no progresso científico pautar-se no que já foi desenvolvido anteriormente por seus pares, seja para seguir a partir de determinado ponto deixado ou para criticar a metodologia e pesquisa realizados, e assim estabelecer uma nova hipótese sob a qual se proponha a trabalhar. Para que isso seja possível, durante seu processo de formação, o cientista conhece as principais fontes de informação científica utilizadas em sua área de estudos.

Não surpreende que com o amplo acesso ao ambiente digital, muitos pesquisadores no início da carreira acadêmica adquiram o hábito de fazer pesquisa em fontes de informação não consagradas e reconhecidas pela comunidade científica, que seriam as bases de dados especializadas ou interdisciplinares. Essa prática, do ponto de vista metodológico é inadequada, uma vez que é essencial que o pesquisador consulte e se pautem em pesquisas científicas

realizadas com rigor metodológico e que estejam devidamente registradas em fontes que possuam, por exemplo, um processo formal de avaliação por pares, como é o caso das revistas científicas. Conforme afirma Guedón (2010, p. 25), “em sua área de especialização e em seu alcance geográfico, as revistas podem ser categorizadas de acordo com algum tipo de escala, no qual estão envolvidos prestígio, autoridade e visibilidade”. Sabe-se que existem as chamadas revistas predatórias, utilizadas para publicar artigos científicos a preços altos e que não possuem confiabilidade em seu processo de avaliação, e muitas vezes o pesquisador pode tornar-se usuário de periódicos que disponibilizam informação científica não confiável.

Outra consequência da explosão informacional amplificada em ambiente digital é que se torna cada vez mais difícil as unidades de informação como as bibliotecas, os centros de documentos ou os institutos de pesquisa, conseguir adquirir acesso a todas as bases de dados reconhecidas e utilizadas pelo público de sua instituição de pesquisa.

Diante desse potencial obstáculo, ganham força iniciativas à pesquisa, como as de acesso aberto (AA), também conhecida como *open access*. Denúncias também são realizadas diante de questões como pesquisas científicas financiadas por agências de fomento estaduais ou federais e que têm seus resultados publicados em periódicos científicos de acesso restrito.

O movimento AA foi uma iniciativa internacional que surgiu com a declaração *Budapest Open Access Initiative* em 2002. No presente trabalho será adotado o entendimento de Bailey Junior (2005) sobre a utilização do AA, no qual o “aberto” remonta a utilização de conteúdos sem desrespeitar os direitos autorais.

[...] o OA, oficializado em 2003, fortalece o controle e o gerenciamento da produção científica e articula a força política necessária para a ruptura dos paradigmas científicos vigentes e pautados nas práticas das grandes editoras científicas internacionais de revistas. O surgimento de publicações geradas no âmbito do Acesso Aberto se expande, sobretudo, no ambiente acadêmico. Ao provocar a reconstrução de processos da comunicação científica, tais publicações promovem, potencialmente, condições de visibilidade e acessibilidade que extrapolam as práticas tradicionais. Desse modo, ambos os movimentos estimulam **significativas discussões sobre o papel fundamental da comunicação científica, em especial quanto à responsabilidade social da ciência e da tecnologia e da pesquisa científica, com ênfase para as financiadas com dinheiro público** (FERREIRA; MARCHIORI; CRISTOFOLI, 2010, p. 82, grifo nosso)

Além da dificuldade de aquisição ou assinatura de todas as fontes de informação científica reconhecidas, o pesquisador também se depara com a velocidade na produção de conteúdos e sua conseqüente obsolescência. É fundamental que o pesquisador não somente

conheça as fontes especializadas e certificadas, mas também seja capaz de escolher os melhores recursos e ferramentas disponibilizadas para facilitar seu trabalho de curadoria, ou seja, de escolha de conteúdos pertinentes à sua necessidade, trabalho este com a qual muito contribuem as técnicas de organização e representação da informação, conforme abordado anteriormente.

A velocidade com que o conhecimento é renovado, tornando ultrapassada a literatura ainda recente – especialmente em algumas áreas do conhecimento – acarreta problemas para todos os interessados: é difícil para o cientista manter-se informado ou atualizado e é também difícil e caro para os centros de informação manter suas coleções atualizadas, pois o número de fontes aumenta com igual velocidade (MUELLER, 2000, p. 25).

Contudo, mesmo diante da velocidade de produção e disseminação da informação científica, é certo que em ambiente digital o amplo acesso às fontes de informação facilita o trabalho do pesquisador e possibilita maiores condições de desenvolver seu trabalho. Cada vez mais as iniciativas como o AA reiteram discussões na literatura na área da Ciência da Informação como a feita por Le Coadic (2004, p. 26): “a informação é a seiva da ciência [...] Fluído precioso, continuamente produzido e renovado, a informação só interessa se circula e, sobretudo, se circula livremente”.

Na *web*, as bases de dados e sistemas de recuperação da informação científica oferecem múltiplas alternativas de busca para seus usuários em termos de filtros, interface de busca, *layout*, dentre outros; porém é fundamental questionar se as alternativas oferecidas também contribuem para uma recuperação efetiva dos conteúdos, uma vez que em ambiente *web* a velocidade e volume das informações se potencializam, sendo essencial o engajamento no desenvolvimento de alternativas que solucionem ou ao menos minimizem a questão.

[...] possibilitam o planejamento de estratégias de busca com maior nível de complexidade envolvendo vários contextos na mesma estratégia; permitem a utilização de busca de palavras apenas dos títulos e resumos dos documentos, isto é, termos da linguagem natural; buscam os termos específicos de linguagens controladas, nos campos de descritor; buscam por autores; por ano de publicação; por títulos de periódicos; por classificação; permitem, também, a busca de conceitos compostos ou simples e a possibilidade de truncagem de raízes de palavras e de substituição de caracteres no meio dos termos, dentre outros recursos de recuperação. (LOPES, 2002, p. 60).

Todas as opções e campos de busca oferecidos pela fonte de informação devem buscar padrões a serem seguidos, para que possam ser bem utilizados pelos pesquisadores com necessidades informacionais específicas a serem sanadas.

A discussão sobre *big data*, ou seja, sobre um volume informacional intenso propiciado pela dinâmica digital em que há a possibilidade de compartilhamento e acesso de forma mais prática se comparada aquela que era feita com os suportes físicos, é aplicável ao contexto de recuperação da informação científica, pois o fluxo de novos artigos científicos, livros, patentes, etc., em qualquer área do conhecimento é intenso, e atualmente as bases de dados indexam diversos conjuntos de periódicos com esses conteúdos.

O acesso aos grandes sistemas de recuperação de informação, também denominados de bancos de dados e, conseqüentemente, às suas bases de dados veio ampliar significativamente a qualidade das buscas bibliográficas, visto que essas bases proporcionam diversificados pontos de acesso à informação. (LOPES, 2002, p. 60).

Considerando o contexto da disponibilidade da produção científica para fomentar o desenvolvimento em ciência, sempre aliado com a preocupação de organização e representação adequada, algumas iniciativas como os repositórios digitais se posicionam de forma diferente às consagradas bases de dados científicas e aliados a elas podem subsidiar e desenvolver estratégias mais eficientes de recuperação da informação pelo pesquisador ou por instituições que tenham como objetivo analisar essa produção científica enquanto insumo estratégico.

2.1.1 Repositórios

Ao considerar o contexto do AA na comunidade científica, são imprescindíveis as iniciativas que possuem como principal objetivo colaborar para que a ciência se beneficie de conteúdos abertos realizados pelas universidades ou instituições de pesquisa, de forma que os pesquisadores utilizem as informações disseminadas por seus pares para construir novos conhecimentos. Nesse sentido e considerando também a organização da informação científica em ambiente *web*, tem-se os denominados repositórios digitais, como um dos intermediários entre a informação e seu potencial usuário.

O primeiro repositório digital foi criado por Paul Gispang, o *LANL preprint archive*, nos Estados Unidos da América, como um arquivo de *preprints* de Física no início da década de 1990, momento em que a quantidade de informações produzidas estava em crescimento; em 1999 esse repositório passou a ser chamado de arXiv (GONÇALEZ et al., 2018, p. 5237).

Para Cerrao e Castro (2018, p. 95), “os repositórios digitais são ambientes de tecnologia aberta, provedores de acesso livre a documentos eletrônicos, que gerenciam, armazenam e

disseminam a informação digital”. Os repositórios digitais também podem ser definidos como “[...] ambientes informacionais que viabilizam os processos de acesso, preservação da informação e da interoperabilidade entre os recursos informacionais” (GONÇALEZ et al., 2018, p. 5237).

Os repositórios surgem em um contexto de crescimento informacional desordenado e possuem como objetivo serem suportes ou ferramentas que visam proporcionar em ambiente digital um acesso específico e organizado aos conteúdos, sejam eles por áreas ou instituições específicas, de modo que usuários em potencial consigam sanar suas necessidades informacionais de forma mais adequada possível.

Os repositórios digitais seguem estruturas e padrões específicos que os possibilitam depositar, armazenar a disponibilizar informações de forma que elas sejam interoperáveis. Cerrao e Castro (2018, p. 95) complementam que “a padronização da representação e da descrição dos recursos informacionais é essencial para o pleno funcionamento de repositórios digitais” e a Ciência da Informação possui um papel ativo na idealização de padrões para permitir qualidade e confiabilidade na recuperação e disseminação da informação. Para garantir acessibilidade e recuperação adequadas em ambiente *web*, é fundamental que a interoperabilidade de repositórios digitais seja uma preocupação primordial em seu desenvolvimento, sendo também um procedimento preconizado pela *Open Archives Initiative* (OAI). Nesse sentido, a implementação de padrões adequados perpassa pela estruturação dos metadados.

Os metadados podem ser definidos como “[...] os dados estruturados que descrevem, explicam, identificam, localizam e representam um recurso [...] os metadados permitem que os dados possam ser encontrados e interpretados, por meio de identificadores únicos e persistentes” (SIMIONATO, 2017, p. 15). Eles são padronizados com o objetivo de garantir a recuperação eficaz da informação e sua consequente interoperabilidade entre sistemas, uma vez que é possível que determinado sistema saiba exatamente o caminho e o tipo de informação que irá encontrar para conseguir disponibilizá-la adequadamente. Cerrao e Castro afirmam:

Os padrões de metadados são estruturas organizadas e padronizadas contendo um conjunto de metadados, com finalidade de melhor representação e recuperação da informação, de acordo com as práticas e as necessidades de uso adequado de metadados em diferentes ambientes e situações ao decorrer do tempo (CERRAO; CASTRO, 2018, p. 96).

O principal objetivo dos repositórios digitais, inseridos no contexto do AA, é fornecer acesso à informação, o que contribui diretamente com o desenvolvimento de novas práticas em ciência. Para que esse objetivo seja atingido é preciso que esses repositórios atendam os padrões e estruturas adequadas para possibilitar interoperabilidade sem deixar de garantir a confiabilidade dos conteúdos por eles disponibilizados.

Dessa forma, os repositórios digitais não são simplesmente ferramentas que possuem o objetivo de reunir e disponibilizar produção científica, mas também possuem a preocupação de tornar esses dados interoperáveis seguindo padrões estabelecidos para que estejam alinhados às possibilidades de compartilhamento disponíveis em ambiente *web*.

2.1.1.1 Repositórios de dados de pesquisa

No contexto acadêmico-científico, iniciativas como o AA se mostram eficazes na reafirmação da posição da ciência como intrínseca à sociedade, de maneira que ela seja vista como essencialmente feita por e para o meio social, sendo beneficiada pelo cidadão ao mesmo tempo em que o beneficia. Nesse sentido, pode-se refletir sobre a importância da abertura não somente das publicações enquanto resultado da pesquisa, mas também dos dados relativos a todo o seu processo, ainda que seja fundamental a reflexão do pesquisador a respeito dos interesses e fatores envolvidos nesse processo.

O termo “dado de pesquisa” tem definições diferentes de acordo com a área de pesquisa, se transformando de acordo com os objetos e metodologias, e pode ser o resultado de um experimento de laboratório, estudo ou observação de fenômenos (SAYÃO; SALLES, 2016). A Organização de Cooperação e de Desenvolvimento Econômico (2007) define dados de pesquisa como registros numéricos, textuais, iconográficos ou sonoros utilizados como fonte primária para a pesquisa científica e como instrumento para validação e reprodução da pesquisa para sua validação.

Embora a publicação científica traga detalhado o percurso metodológico de pesquisa, é pertinente que mais detalhes a respeito de sua elaboração sejam conhecidos por outros pesquisadores. A pesquisa científica em diferentes áreas do conhecimento é um processo envolto em dados, sejam eles, por exemplo, dados de entrevistas realizadas em vídeo, áudio ou escrito ou inclusive amostras coletadas e trabalhadas em laboratório experimental. Todos os insumos utilizados no processo podem ser considerados como dados relevantes de pesquisa, inclusive anotações de experimentos que em um primeiro olhar podem se mostrar irrelevantes,

mas que descrevem detalhes a respeito de como determinada situação aconteceu no decorrer da pesquisa.

De acordo com a *National Science Board* (2005, p. 19), os dados de pesquisa devem ser categorizados, para que então sua guarda e preservação sejam garantidos, como observacionais, computacionais ou experimentais: observacionais como resultado de fenômenos em determinadas condições, espaço e tempo, os quais não podem ser coletados novamente; computacionais como resultado da execução ou simulação de um programa; e experimentais como mais complexos, uma vez que são resultados de experimentos realizados mediante supervisão em laboratório.

A conduta ética do pesquisador de descrever todo o caminho percorrido para se chegar a um determinado resultado ou considerações a respeito da hipótese de pesquisa é válido e indispensável, contudo, a publicação científica em si que será acessada por outros pesquisadores da área não vai detalhar todos esses dados enquanto insumos para sua realização.

Compartilhar os dados de pesquisas é uma necessidade. Hoje, a sociedade enfrenta desafios científicos muito mais complexos que pressupõem a interdisciplinaridade, colaborações entre laboratórios de pesquisa, permuta de informações e de competências. Com isso, novas competências associadas ao trabalho com os dados de pesquisa estão surgindo para corroborar com o descobrimento de novos conhecimentos a partir desses volumes de dados (AVENTURIER; ALENCAR, 2016, p. 3).

Para o pesquisador que, ao verificar determinado resultado exposto em um trabalho científico, pretende seguir metodologia semelhante para conseguir comprovar sua própria hipótese em sua área ou linha de pesquisa, a disponibilidade dos dados de pesquisa otimizaria o seu investimento de tempo e recursos para coletar novos dados, sendo que aqueles já poderiam ser utilizados em alguma medida, seja como contribuição ao trabalho ou evidência para descartar aquele percurso de pesquisa.

As vantagens dessa disponibilização de dados de pesquisa não se dão somente para o pesquisador que irá se beneficiar para otimizar seus recursos, mas também para os pesquisadores, instituições e laboratórios experimentais que tiveram o reconhecimento por parte da comunidade científica em ter seus conteúdos utilizados e devidamente referenciados, respeitando os devidos direitos de utilização.

A gestão de dados científicos abertos pode propiciar vários benefícios no campo acadêmico, dentre eles, permitir o reuso/compartilhamento de dados, reduzir custos e retrabalho, dar credibilidade as pesquisas na medida em que

permite testar os resultados, ampliar a visibilidade dos autores e possibilitar novos estudos a partir de abordagens ou novas interações entre estes dados (LECARDELLI; MONTEIRO; LUCAS, 2019, p. 11-12).

Tendo em vista que a disponibilidade dos dados de pesquisa beneficia tanto os pesquisadores quanto as instituições e seus respectivos laboratórios, é fundamental que a comunidade científica também se atente para os suportes e estratégias que são utilizadas atualmente para a disponibilização desses dados, se são ou não eficazes, e quais são as oportunidades de melhoria que podem ser implementadas, considerando que se trata de um tipo de dado sensível e de alto impacto, ou seja, seu fácil acesso condiciona a otimização no desenvolvimento científico das mais diversas áreas e linhas de pesquisa.

Nesse contexto, diante da necessidade de compartilhamento de dados, surgem os denominados **repositórios de dados de pesquisa**, que segundo Torres-Salinas, Martín-Martín e Fuente-Gutiérrez (2014) estão dedicados a armazenar, descrever e disseminar dados científicos.

Para Monteiro e Sant’Ana, esses repositórios “são mantidos por conjuntos de ações que viabilizam o armazenamento de dados coletados visando a otimização da recuperação, o que amplia as potencialidades de reuso destes dados entre os pesquisadores [...]” (MONTEIRO; SANT’ANA, 2016, p. 653).

Nesse sentido, podemos compreender que a missão dos repositórios de dados de pesquisa está intrinsecamente alinhada ao propósito de disseminação dos próprios dados, ou seja, conferir acesso e visibilidade à pesquisa e às instituições de forma que esses conteúdos tenham condições de ser acessados via ambiente próprio, caracterizado enquanto repositório de dados.

A disponibilização dos dados das pesquisas utilizados nos artigos em repositórios de acesso aberto, seguindo padrões de registro que assegurem a autoria, o uso e citação dos dados assim como do artigo correspondente, é recomendável pois contribui para a replicabilidade das pesquisas, aumenta a visibilidade e as citações das pesquisas e dos periódicos (SCIELO, 2017, p. 20).

As iniciativas nacionais e internacionais estão alinhadas para facilitar o acesso aos dados de pesquisa de forma a colaborar diretamente com o desenvolvimento de novos trabalhos e práticas científicas pelos pesquisadores de diferentes áreas do conhecimento.

A comunidade científica nacional e internacional tem se empenhado em desenvolver políticas ou, ao menos, iniciativas, para conscientizar pesquisadores a depositarem seus dados de pesquisa por meio da elucidação de seu papel para o desenvolvimento da ciência.

Uma fonte que pode ser consultada pela comunidade científica a respeito de repositórios de dados é o re3data¹, que fornece uma série de informações, desde 2012, a respeito de mais de 2000 repositórios de diferentes áreas de pesquisa.

Destacam Sayão e Sales (2015), que a gestão de dados científicos oferece condições para que eles sejam acessados e compreendidos por outros pesquisadores ao longo do tempo. Para isso, é crucial que a descrição dessas informações nos repositórios seja feita de forma adequada.

Inclusive Tartarotti, Dal'Evedove e Fujita (2019) reiteram a importância de uma descrição adequada sobretudo em relação aos metadados, que sejam bem estruturados e tornem possível a documentação das informações sobre o contexto de criação daqueles dados, possibilitando, por exemplo, a relação entre os dados e o artigo de pesquisa; além de destacar a importância de identificadores persistentes como o *Digital Object Identifier* (DOI), para que seja possível recuperar o objeto digital sem se tornar refém de mecanismos como endereços de sites que venham a ser modificados e se tornarem indisponíveis ao longo do tempo.

É importante que se pense nessa conexão entre os dados de pesquisa, o artigo científico gerado a partir deles e também os autores e instituições envolvidas, para que a disponibilidade dos dados não seja somente um protocolo a ser cumprido a favor das práticas de ciência aberta, sem que seja uma atitude consciente e devidamente estruturada, para que então consiga atingir seu verdadeiro propósito.

Dessa forma, fica evidente o quanto gerenciar dados de pesquisa vai muito além da inicial conscientização dos pesquisadores e demais atores na comunidade científica, é necessário também planejar a estrutura em que se encontram esses dados e garantir que eles sejam coletados, armazenados, recuperados e descartados, se essa for a conduta adequada, para que então eles possam de fato ser utilizados para a construção de informação e do conhecimento científico.

Sendo assim, as discussões a respeito de padrões, ferramentas, condutas e políticas utilizadas nos dados de pesquisa precisam continuar e se intensificar, inclusive de maneira a

¹ <https://www.re3data.org/>

considerar o tratamento e a disponibilidade de materiais cada vez mais diversos, uma vez que possuem papel essencial no desenvolvimento da ciência.

2.1.1.2 Repositórios de *e-prints*

Um das estratégias para a aceitação, incentivo e uso do modelo de AA a conteúdos científicos pelos pesquisadores são a criação e a divulgação de repositórios digitais como instrumentos que viabilizem o depósito e o acesso de publicações.

Os repositórios temáticos ou institucionais disponibilizam publicações científicas que passaram pelo processo de avaliação por pares e já se encontram publicados oficialmente em um periódico científico, contudo existem também os repositórios de *e-prints* que também podem ser considerados repositórios digitais, porém estão envoltos em polêmicas considerando que armazenam material que pode ou não ter sido avaliado por pares.

Os denominados repositórios de *e-prints* armazenam produção científica, mas com algumas particularidades em sua definição e com uma proposta levemente diferenciada se comparados com os repositórios temáticos e institucionais. Para Weitzel (2006, p. 128), os repositórios de *e-print* são considerados um tipo de repositório digital (temático ou institucional) que utiliza o *software E-prints* da Universidade de Southampton, reunindo textos já publicados ou não e que são passíveis de comentários por parte de outros pesquisadores; função que não excluiria a avaliação por pares, mas ofereceria uma nova forma de avaliação através desses comentários, resultando em versões atualizadas da publicação com base nas sugestões e julgamentos feitos pela comunidade científica que a acessou.

Após a definição de repositório de *e-print* é fundamental esclarecer que esse repositório possui a versão digital de um artigo científico que pode não ter sido ainda publicado, ou seja, ser um *pre-print*, ou pode também conter a versão digital de um artigo que já foi submetido ao processo de avaliação por pares, ou seja, ser um *post-print*.

Normalmente a versão *post-print*, também conhecida como o manuscrito ou versão aprovada, pode ser depositada em um repositório depois um período de embargo que pode chegar até 36 meses após a data de publicação oficial do artigo no periódico (AGUIA, 2020).

Benkler (2010) cita o exemplo do ArXiv e Cogprints, como uma criação de arquivo aberto, que teve início como um repositório de artigos de Física e posteriormente de Ciência da Computação e Matemática; nele os pesquisadores postam suas publicações, recebem

comentários e publicam versões revistas com base nas sugestões antes de enviá-los ao sistema formal de revisão por pares, caso sejam de fato publicados.

Os repositórios de *e-prints* estão envoltos em polêmicas e desconfianças como “a baixa percepção de que esse modelo otimiza o fluxo da ICT” (WEITZEL, 2006, p. 7), se referindo a Informação em Ciência e Tecnologia (ICT), pois muitos não conferem a eles legitimidade baseados no pressuposto de que os conteúdos disponibilizados não possuem confiabilidade garantida da mesma forma que possuem via processos de comunicação científica tradicionais, como através dos periódicos científicos.

Assim como ocorre nos repositórios de dados de pesquisa, nos repositórios de *e-prints* os próprios pesquisadores/autores realizam o depósito da versão de seu artigo de pesquisa. Esses repositórios se assemelham à construção de uma rede colaborativa, onde outros pesquisadores, que também contribuem com o depósito de publicações, são os próprios avaliadores, ou seja, a avaliação também é feita por especialistas.

Esse modelo faz uma analogia à produção *peer to peer* (P2P), uma produção inserida no contexto da economia colaborativa e economia política da produção entre pares. Para Bauwens (2010, p. 1), o P2P aborda os processos que tem como objetivo aumentar a participação generalizada de participantes equipotenciais. Além disso:

[...] produzem valor de uso através da cooperação livre entre produtores que têm acesso a capital distribuído: este é o modo de produção P2P, um 'terceiro modo de produção', diferente da produção com fins lucrativos e da produção pública efectuada por companhias detidas pelo estado. O seu produto não reside num valor de troca destinado ao mercado mas num valor de uso dirigido a uma comunidade de utilizadores (BAUWENS, 2010, p. 1).

De forma semelhante aos processos P2P, os repositórios de *e-prints* se mantêm através da livre cooperação entre usuários, além de todos os esforços terem como objetivo atingir cada vez mais colaboradores, aumentando conseqüentemente a quantidade de produção científica disponível e o alcance desses conteúdos na rede. Damasio (2018, p. 160) discute que os *pre-prints* são como os trabalhos apresentados em evento, passíveis de comentários e sugestões naquele momento, mas que se encontram em formato eletrônico armazenado em um repositório ou servidor e que é acessado em ambiente digital; esses servidores suportam arquivos manipuláveis e não somente estáticos, como é o caso do formato de documento portátil *Portable Document Format* (PDF), ou seja, possibilitam o arquivamento de imagens, relatórios, textos e dados.

Esse armazenamento de dados é comum aos repositórios de dados de pesquisa, que também abarcam uma quantidade maior de informações a respeito da pesquisa e diversos materiais que fizeram parte dela, a diferença para os repositórios de *e-prints* ou especificamente de *pre-prints* se daria na possibilidade de contribuição ou melhoria daquele conteúdo.

Conforme afirmam Silva e Lazzarin (2019, p. 157), “[...] o periódico não é o único caminho para a comunicação científica” e iniciativas como os repositórios de *e-prints* podem acelerar a comunicação de descobertas e inovações entre os membros da comunidade científica. Os repositórios de *e-prints* atuam como uma forma paralela e ágil de acesso ao que se tem pesquisado mais recentemente em determinada área, uma vez que os conteúdos não passaram por longos processos de avaliação por pares como na submissão em periódicos científicos; em algumas áreas, como a das Ciências da Saúde, esse último aspecto se acentua, pois a agilidade da divulgação de um resultado de pesquisa impacta diretamente na melhoria dos processos em saúde pública.

Nesse sentido, pode-se refletir acerca da pandemia de COVID-19, durante a qual a comunidade científica mundial se mobiliza para construir caminhos que a possibilitem compreender os reais efeitos da doença nas mais distintas faixas etárias e populações, de forma a consolidar um possível tratamento ou vacina eficaz. Para que se consiga atingir esse objetivo, os pesquisadores precisam consigam acessar literatura recente sobre a temática o mais rápido possível, afinal são muitas vidas perdidas enquanto não se encontra uma forma de combater a doença, e isso implica em uma maior agilidade nos processos de revisão por pares pelos periódicos em que os artigos foram submetidos, para que eles fiquem acessíveis o mais rápido possível. Também implica na disponibilidade desses artigos, considerando que nem todos os pesquisadores e instituições de pesquisa do mundo possuem assinaturas de grandes editoras e revistas científicas a custos elevados. Diante do caos mundial, editoras internacionais disponibilizam parte de suas coleções gratuitamente para que os pesquisadores encontrem as melhores evidências para continuar suas pesquisas (CAPES, 2020).

Os repositórios digitais fazem parte de uma tentativa de saída do processo de comunicação científica tradicional, contribuindo para um acesso à informação mais horizontalizado, alinhado à proposta da OAI. A iniciativa dos repositórios tem como principal objetivo facilitar o acesso dos conteúdos científicos aos pesquisadores de forma mais ágil, mas sem que se perca a qualidade do que é produzido.

Existem algumas redes sociais científicas com propostas semelhantes aos repositórios digitais, especificamente, aos de *e-prints*, que possibilitam ao pesquisador divulgar sua

produção científica via mídias digitais. As redes Academia.edu e ResearchGate funcionam como repositórios de conteúdo científico onde os pesquisadores, além de disponibilizarem suas publicações, acessam outras áreas de interesse, mas ambas permitem anexos de produção científica que não passaram por nenhum processo formal de avaliação por pares. Mais detalhes sobre as redes serão abordados em tópicos mais adiante neste trabalho.

Seja nos repositórios de *e-prints* ou redes sociais científicas é perceptível quão importantes são as iniciativas alinhadas ao OAI no movimento de tornar o acesso a ciência mais horizontalizado, pois possuem cada vez mais destaque e mostram potencial de serem amplamente adotadas na comunidade científica. Contudo, é relevante que essas iniciativas estejam alinhadas também com a qualidade da disposição dos conteúdos nessas plataformas, buscando estratégias para organizar e representar essas informações adequadamente para que cheguem ao pesquisador.

Damasio (2018) mostra que os *pre-prints* foram adotados para utilização da Crossref, a agência reguladora que atribui o identificador DOI, de forma que o número atribuído ao *pre-print* não seja o mesmo número DOI do artigo final publicado, uma vez que compete ao editor do periódico atribuir esse novo código. A Crossref (2018) afirma que o *pre-print* e o artigo final publicado pelo periódico são documentos distintos, mas que logicamente possuem uma relação, que deve ser realizada e ficar clara por meio da marcação e preenchimento correto de seus metadados.

A preocupação com a identificação correta dos documentos, sejam eles nas versões *pre-print* ou nas versões finais, deve se estender de igual forma para a identificação correta dos autores dessas publicações e das instituições nelas envolvidas, de forma que seja possível identificar esses dados independente do ambiente em que esteja hospedado.

É bom para as instituições de pesquisa conseguirem recuperar a produção científica publicada em ambientes como os servidores de *e-prints*, inclusive para analisar e cogitar possíveis parcerias entre autores, considerando seu nível de colaboração para com aqueles trabalhos que foram depositados nos servidores.

Uma iniciativa brasileira especificamente em relação aos *pre-prints* se dá com a *Scientific Electronic Library Online* (SciELO) que anunciou em 2018 o servidor SciELO Preprints. Lançado em 2020, servidor pode armazenar trabalhos que ainda não foram submetidos a nenhum periódico e são depositados pelos próprios autores, ou podem ser depositados por periódicos indexados no SciELO caso sugerido pela própria revista aos autores no momento de submissão (antes de iniciar o processo de revisão por pares) ou após o aceite

do trabalho; no caso dos trabalhos submetidos ao servidor após o aceite pela revista, eles receberão o mesmo número DOI do artigo final; caso não sejam manuscritos aceitos previamente por periódicos, recebem um número DOI exclusivo do servidor *pre-print* (SCIELO, 2020).

O modelo tradicional de comunicação científica, via periódicos científicos, se consolidou e se tornou hegemônico, entretanto, com o advento das mídias digitais, as TICs e iniciativas de AA à informação, como a OAI, surgiram contextualizadas à nova realidade. Os repositórios de *e-prints*, com sua proposta levemente diferenciada dos repositórios digitais tradicionais, surgem exatamente com uma forma alinhada à essa nova realidade que conecta informação com tecnologias de livre acesso. Sua contribuição é fundamental para o progresso da ciência, uma vez que as informações e conteúdos utilizados nas pesquisas podem ser disponibilizados por meio de uma única ferramenta. A oportunidade de ter o trabalho avaliado por outros especialistas possibilita correções, aperfeiçoamentos e sugestões de melhorias antes mesmo de ser enviado para um processo de avaliação via periódico que, talvez, pode até aumentar as possibilidades de ser publicado, pois se configura como um meio de divulgação colaborativo, embora seja fundamental o constante olhar crítico do pesquisador diante da atuação e papel dessas plataformas.

Esse modelo de repositório possui um impacto direto na comunicação científica em meios digitais. Sua real proposta normalmente não é conhecida por pesquisadores, gerando desconfiança e abalando a sua expansão. É fundamental que a comunidade científica conheça essa ferramenta alternativa para que contribua para seu desenvolvimento de maneira que se molde conforme as necessidades apontadas pelos usuários.

2.2 Objetos digitais

O amplo acesso à informação científica em ambiente digital proporciona recebimento de informações atualizadas a respeito de temas de pesquisa pelos pesquisadores em um curto espaço de tempo, o que não ocorreria sem a rapidez proporcionada pelas tecnologias inseridas na comunicação científica.

Além de proporcionar esse intercâmbio de informações de forma mais ágil, a tecnologia também contribui ativamente para otimizar processos que fora do contexto digital demandam tempo e esforços maiores. Nesse sentido, é possível caracterizar a informação científica em ambiente digital como um tipo de objeto digital, considerando que sua descrição, representação

e identificação foram disponibilizadas na *web*.

O termo objeto digital é comumente utilizado no contexto da preservação digital e pode ser definido como ambíguo pois se refere aos aspectos conceituais e técnicos, ou seja, o conteúdo intelectual, a formatação e sua estrutura digital (HOFMAN, 2002 apud YAMAOKA; GAUTHIER, 2013, p. 79).

Neste trabalho será considerada a definição genérica de Thibodeau (2002) para objeto digital como objeto de informação, de qualquer tipo e em qualquer formato, que esteja disponível de maneira digital e o complemento realizado por Santarém Segundo (2010), destacando que os objetos digitais são compostos por informações, metadados e identificadores.

Todos os objetos digitais são entidades com herança múltipla; ou seja, as propriedades de qualquer objeto digital são herdadas de três classes. Todo objeto digital é um objeto físico, um objeto lógico e um objeto conceitual, e suas propriedades em cada um desses níveis podem ser significativamente diferentes. Um objeto **físico** é simplesmente uma inscrição de sinais em algum meio físico. Um objeto **lógico** é um objeto que é reconhecido e processado por software. O objeto **conceitual** é o objeto tal como é reconhecido e compreendido por uma pessoa ou, em alguns casos, reconhecido e processado por um aplicativo de computador capaz de executar transações de negócios (THIBODEAU, 2002, p. 3, tradução nossa, grifos do autor).

A informação científica que é normalmente acessada pelos pesquisadores inseridos na comunidade científica, enquanto um objeto digital disponível na *web*, pode ser inserido em meio físico, processado por *softwares* e reconhecido por pessoas. Suas características enquanto objeto digital, possibilitam amplo alcance, visibilidade e compartilhamento. Nesse contexto, falar sobre os metadados é primordial, pois são normalmente definidos como dados sobre dados, mas que, cada vez mais por causa do ambiente digital, o conceito tem sido empregado em diversos contextos com diferentes propósitos e tecnologias (CAMPOS, 2007).

Com o objetivo de mencionar uma definição mais assertiva ao contexto do trabalho e conseqüentemente menos generalista, foi adotada a seguinte:

[...] dado que descreve atributos de um recurso, caracteriza suas relações, apoia sua descoberta e uso efetivo, e existe em um ambiente eletrônico. Usualmente consiste em um conjunto de elementos, cada qual descrevendo um atributo do recurso, seu gerenciamento, ou uso (VELLUCCI, 1998, p. 192 apud CAMPOS, 2007, p. 19).

Normalmente a comunidade científica acessa diretamente o objeto digital, ou a informação científica, mas poucos se atentam aos metadados da publicação, que inclusive são

os responsáveis por sua recuperação quando, por exemplo, é realizada uma estratégia de busca em determinada fonte de informação. Mey e Silveira (2009) destacam algumas vantagens dos metadados:

a) a rapidez do registro bibliográfico, concomitante à rapidez da própria mídia do conhecimento; b) a normalização requerida a um registro bibliográfico – não encontrada nos grandes e populares serviços de busca, que demonstram excessiva revocação e baixíssima relevância – e esta normalização possibilita um alto índice de relevância; c) a possibilidade de busca por campo, como nos catálogos, o que também influi na relevância dos recursos de acesso remoto recuperados (MEY; SILVEIRA, 2009, p. 133).

Os metadados são o principal insumo utilizado pelas fontes de informação científica, e sua responsabilidade está sob os autores das publicações científicas, considerando sua atenção às informações fornecidas durante o processo de submissão dos periódicos científicos e durante o processo de submissão e diagramação e das próprias bases de dados, considerando suas tecnologias e ferramentas para *upload* e disponibilização dos dados. Nesse sentido, Habermann (2020) afirma que criar metadados que estejam em conformidade aos padrões pode ser um esforço com impacto direto ao compartilhamento e reuso dos dados.

A normalização que os metadados requerem é indispensável não só à recuperação, mas também à preservação e interoperabilidade por eles proporcionada.

E o papel essencial dos metadados na recuperação das publicações científicas nas fontes de informação em ambiente digital tem lançado desafios ao seu próprio desenvolvimento, sobretudo no que diz respeito ao controle de autorias dos artigos científicos.

Assim os recursos são codificados em vários esquemas de metadados, bibliotecas digitais crescem, e a internet e os metadados codificados se movem em direção à interoperabilidade, os problemas de desambiguação de nomes e identidades representam problemas no desenvolvimento dos metadados. Bases de dados e recursos para pesquisa devem ser capazes de determinar se a pessoa que escreveu o artigo A também escreveu o artigo B. Pesquisas podem querer recuperar todos os itens escritos ou criados por uma pessoa específica. Pesquisadores podem precisar determinar exatamente quem escreveu um artigo para entrar em contato com o autor propondo futura colaboração ou fazer perguntas sobre os dados. A maioria das práticas com metadados são suportam facilmente a desambiguação de nome e o problema cresceu assim como o número de recursos e variedades de metadados (ELLIOTT, 2010, p. 1, tradução nossa).

A rápida recuperação da informação científica possibilitada pela *web*, por meio dos objetos digitais descritos com os metadados, também abre oportunidades para o

desenvolvimento de ferramentas e tecnologias que possam descrever e identificar esses recursos de maneira automática e que reduza esforços profissionais desnecessários, incluindo o próprio controle de autoridades, temática que será abordada adiante, que é primordial para a recuperação informacional e que demanda atenção às regras e padrões adotados, consolidando o objetivo dos processos de organização e representação da informação.

3 RECUPERAÇÃO E VISIBILIDADE DA PESQUISA DO PESQUISADOR E DAS INSTITUIÇÕES NA COMUNIDADE CIENTÍFICA

Ao considerar o grande volume de informação científica em meio digital e discutir a recuperação de literatura científica pela comunidade acadêmica, deve-se atentar para a forma como autores e instituições são representados nas fontes de informação e como isso impacta diretamente na sua visibilidade e conseqüentemente na de suas pesquisas.

Normalmente, quando se fala em pesquisa científica comumente se pensa nas próprias publicações e na produtividade da ciência de uma forma geral. Contudo, quando as publicações são o único foco, outros dois importantes componentes do universo da pesquisa são negligenciados: autores e instituições (HAAK; MEADOWS; BROWN, 2018). Estes dois atores precisam ser atentamente considerados em sua representação na literatura científica, considerando sua implicação direta na identificação e recuperação em fontes de informação.

Gasparyan et al. (2016) afirmam que os problemas com a autoria não são novos no contexto editorial e provavelmente existiam antes mesmo da publicação do primeiro periódico científico, o *Philosophical Transactions of the Royal Society* em 1665. Por décadas, o uso incorreto de abreviaturas, nomes escritos incorretamente, nomes e sobrenomes mal posicionados afetaram pesquisadores em todo o mundo (GASPARYAN et al., 2016). Bourne (1977) realizou um estudo em relação à recuperação da informação científica em bases de dados e evidenciou os altos índices de erros de ortografia e como esses impactam nos resultados de busca.

Fox (2016) comenta que saber o nome de alguém não é suficiente para conseguir conectá-lo de fato a uma pessoa real. É consideravelmente comum que pesquisadores tenham homônimos, pessoas com o mesmo nome e que também se dedicam à pesquisa científica, e aqueles que não os têm, poderão tê-los futuramente. Considerando essa hipótese, ao buscar por um nome em uma fonte de informação bibliográfica serão recuperadas as publicações de todos os autores com aquele nome. A situação também é delicada caso pesquisadores tenham nomes completos distintos, mas sobrenomes iguais e iniciais iguais, por exemplo, João da Silva e José da Silva: ao buscarem por seus nomes nas bases de dados como “Silva, J”, suas publicações serão recuperadas juntas.

Há também pesquisadores que optam por adotar o nome de cônjuge e seus artigos começam a ser publicados a partir do novo sobrenome, gerando diferentes entradas para o nome enquanto solteiro e casado, somadas às variações ao longo da trajetória acadêmica ocasionada

por divórcios e demais casamentos, se houver.

Os nomes possuem características próprias de acordo com a cultura e idioma (STROTMANN; ZHAO, 2012). Nesse sentido, existem as discussões a respeito de autores asiáticos. Qiu (2008) menciona o caso de três estudantes chineses chamados Wang Xiao-yan, Wang Xiao-rong e Wang Xiao-xue e todos publicam com o nome abreviado “Wang, X.”, fato que torna difícil que a recuperação das publicações dos três alunos seja feita separadamente. Gasparian et al. (2016) também comentam a respeito das particularidades e desafios em relação à padronização de nomes gregos, russos e do sul da Índia.

É comum que em normas bibliográficas de citações e referências se encontrem orientações sobre a citação ser realizada apenas pelo sobrenome do autor e nas referências este também conste atrelado apenas à primeira letra do nome. Contudo, este uso é particularmente sério ao considerar nomes japoneses e coreanos (STROTMANN; ZHAO, 2012). Eugene Garfield, o célebre pesquisador que é considerado o fundador da bibliometria e cientometria e responsável pela criação do *Institute for Scientific Information* (ISI), já alertava na década de 1960 a respeito do uso de apenas parte do nome pelos pesquisadores:

A prática americana mais comum de largar ou não ter um nome do meio (segundo) ou um terceiro nome é uma fonte constante de dificuldade para bibliotecários, indexadores, editores e todos aqueles que devem usar a literatura. Enquanto A. Kantrowitz é "obviamente" um especialista em coração (para o cardiologista) e, obviamente, um físico para os físicos, não é óbvio para muitos outros interessados em seus trabalhos, dos quais um é autor de um artigo publicado em um jornal multidisciplinar (GARFIELD, 1969, p. 763, tradução nossa).

Os pesquisadores transgêneros, além de não encontrarem acolhimento por parte de editoras e periódicos científicos ao solicitar a mudança de seus nomes em publicações anteriores, vivendo diante da escolha de perderem todo o seu histórico de publicações ou terem que conviver com esse amontoado de publicações que exibem uma identidade que pertence ao seu passado (TANENBAUM, 2020), podem também não ter suas pesquisas recuperadas nas fontes de informação bibliográficas.

A formação do jovem pesquisador, desde os cursos de graduação, deveria prever a orientação para que ao iniciar a trajetória acadêmica, o aluno preferencialmente adote uma única forma de seu nome a ser utilizada nas atuais e futuras publicações científicas, com o objetivo de reduzir impactos negativos por mudanças no nome publicado e a sua consequente recuperação. Além da adoção e divulgação de orientações bem claras ou nominas institucionais,

por parte das universidades e instituições de pesquisa, com o objetivo de orientar claramente a forma correta pela qual seus estudantes, pesquisadores e docentes devem mencionar o nome da instituição em suas afiliações.

Um estudo realizado por Pylarinou e Kapidakis (2017) comparou a performance de busca por nomes de instituições nas bases de dados PubMed, Scopus e Web of Science e na plataforma Google Scholar, por nomes de instituições, especificamente por nomes de hospitais, concluindo que a recuperação é mais efetiva se realizada em todas as fontes, de forma conjunta. Os autores do estudo relatam que os dados inseridos no campo de afiliação são uma reação espontânea ao questionamento realizado pelo periódico, ou seja, os dados inseridos não possuem nenhuma forma de controle e são inseridos em linguagem natural. Pois “as instruções aos autores não mencionam ou perguntam por nenhum tipo de controle de autoridade nem se os autores possuem familiaridade com os problemas que as variações de nome podem causar” (PYLARINO; KAPIDAKIS, 2017, p. 216, tradução nossa).

Os esforços para padronizar as entradas das autoridades em catálogos de serviços de informação são relatados em diversos estudos na área de Ciência da Informação, advindos de uma época em que as tecnologias não ocupavam o papel atual na sociedade e conseqüentemente na comunidade científica. A padronização das entradas tinha como principal objetivo representar um autor de forma uniforme, fazendo com que suas pesquisas sempre fossem recuperadas sem equívocos em relação ao seu nome, que poderia ser comum ou semelhante ao de outro pesquisador. Contudo, se trata de um esforço gerenciado por equipes especializadas dos centros de informação, como aquelas compostas por bibliotecários.

Atualmente, são diversos os atores e fontes responsáveis pelo *upload* de pesquisas científicas e das informações de seus autores em ambiente digital, o que ocasiona dispersão em relação aos padrões no registro deste conteúdo.

Sandberg e Jin (2016) afirmam que, historicamente, os catalogadores não realizaram o controle de autoridade para autores de artigos de periódicos por diversas razões, entre elas a enorme quantidade de artigos que tornaria o processo de controle de autoridade demorado e a falta de notificação por parte das editoras a respeito da entrada ou retirada de artigos dos pacotes por elas oferecidos; além da falta de controle por parte dos catalogadores em relação ao que constava nos próprios materiais, considerando que já apresentavam as descrições advindas das próprias editoras e periódicos. É interessante refletir se o investimento de tempo no controle de autoridade de grandes coleções é viável e se compensa, considerando o tempo gasto para controle em uma enorme quantidade de publicações (BURKE; SHORTEN, 2011).

É o que se observa atualmente por meio do envio dos metadados pelos próprios periódicos diretamente para as bases de dados que os indexam. Cada revista possui seus próprios processos e padrões e se não há a preocupação com o controle de autoridades por parte dessas, quando a informação chega ao usuário da base de dados são apresentadas diversas entradas para um mesmo autor.

Os esforços na criação e manutenção de manuais metodológicos concedem diretrizes e traçam alguns padrões a serem seguidos pela comunidade científica quanto aos melhores usos dos nomes de autores e instituições, contudo se não há controle a respeito das regras a serem seguidas ou se não há clareza ou consenso, há pouco avanço. Fonseca (1973), aborda os problemas de comunicação da informação científica e, embora seja uma obra da década de 1970, antes mesmo do surgimento e consolidação do ambiente digital, sinaliza a importância da padronização e estabelecimento de regras para descrição de informações.

[...] não se deve admitir que os canais da comunicação sejam tumultuados e bloqueados por informações transmitidas de acordo com os caprichos pessoais dos autores, as tradições obsoletas dos editores e até com as limitações dos estabelecimentos gráficos. Assim como os motoristas devem conhecer as leis do trânsito e as viaturas devem possuir certos equipamentos de segurança, os autores e editores têm obrigação de conhecer as normas de comunicação científica e os veículos dessa comunicação devem ser normalizados para que as informações neles contidas cheguem mais rapidamente ao conhecimento dos interessados (FONSECA, 1973, p. 55-56).

A falta de padrão nas autoridades pode impactar diretamente nas avaliações de pesquisa que são comumente realizadas por instituições de pesquisa em relação à sua própria produção científica e a de seus pesquisadores ou docentes. Este tipo de impacto atinge diretamente o estabelecimento de melhores práticas, metas, políticas e financiamento na comunidade científica. Os indicadores científicos, utilizados como insumo nas avaliações de produtividade e qualidade da pesquisa, além de significarem o envolvimento de seus autores e instituições no segmento, também conferem visibilidade a eles, sendo amplificada com o uso de redes acadêmico-científicas e perfis em algumas fontes de informação, contribuindo para a identificação correta dos pesquisadores e suas pesquisas, promovendo também a certificação de sua autoria.

3.1 Controle de autoridade e atribuição autoral

Considerando a existência de pessoas ou instituições, que possuem nomes semelhantes ou iguais aos de outras, sejam elas dedicadas ou não à pesquisa, se torna evidente que em algum momento de suas trajetórias poderão ser confundidas se o nome for o único critério tomado como válido para caracterizá-las.

A problemática da ambiguidade de nomes de autores consiste, de modo geral, de duas situações:

- a) a homonímia, que exige que se tenha que distinguir entre diferentes pessoas com um mesmo nome e;
- b) as diversas variações de nome de uma mesma pessoa (MUGNAINI et al., 2012, p. 265).

Nesse sentido, tem-se o conceito de controle de autoridade como “[...] um processo que se baseia na criação e manutenção de ponto de acesso autorizados, padronizados, utilizados para recuperar materiais bibliográficos [...]” (ROSADO; DIAS, 2019, p. 6), podendo estar disponíveis em fontes como catálogos, repositórios, bibliotecas digitais ou bases de dados; sendo que o ponto de acesso pode ser um nome ou termo pelo qual um registro bibliográfico é identificado. Por sua vez, o registro bibliográfico pode ser definido como um conjunto de dados que descrevem um recurso bibliográfico (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016). Sandberg e Jin (2016) afirmam que os catalogadores criaram registros de autoridade por décadas, resultando em imensas bases de dados com nomes padronizados de autores como a *Library of Congress Authorities*.

A descrição bibliográfica realizada com base em processos padronizados e que tenham o objetivo de garantir a assertividade da recuperação informacional, asseguram resultados muito mais efetivos, considerando que este é uma das principais funções de um serviço de informação, e o controle de autoridade é um desses processos que colaboram para a informação pertinente chegar ao usuário.

Clack (1988, p. 36, tradução nossa) destaca que “antes de 1980 o controle de autoridade foi um aspecto negligenciado da organização bibliográfica”. Ressalta-se ainda que:

No contexto dos atuais ambientes informacionais digitais, os estudos sobre o controle de autoridade adquirem cada vez mais importância. Nesses ambientes, precisão e desambiguação, atributos possibilitados pelo controle de autoridade, são características essenciais (ASSUMPCÃO; SANTOS, 2012, p.12)

A consistência pretendida pelo controle de autoridades colabora para que os recursos bibliográficos sejam corretamente recuperados quando buscados por determinado termo pelo usuário que possui uma necessidade informacional específica. Assumpção, Santos e Zafalon (2017a) mencionam que no contexto dos catálogos digitais, os recursos bibliográficos podem ser recuperados pelos diferentes termos que constam no registro, mas ao longo da história da catalogação, por limitações práticas e tecnológicas, a recuperação era possível apenas pelo título do registro, seus responsáveis ou por assunto. Assim, de forma mais direta, o controle de autoridades pode ser definido como:

[...] o processo de unificar, mediante a utilização de uma forma normalizada, os pontos de acesso dos catálogos automatizados e além de mostrar as relações entre os diferentes pontos de acesso. Ou seja, supõe a normalização dos nomes de pessoas, entidades, títulos uniformes ou matérias, que podem constituir o ponto de acesso principal ou os secundários de um catálogo automatizado. Sua finalidade é facilitar a identificação e a recuperação dos documentos armazenados, evitando as confusões a que se podem prestar os homônimos, sinônimos ou a variedade de nomes com os quais podem ser denominados uma pessoa, entidade, obra, tema ou conceito (PASCHUAL, 1999, p. 121, tradução nossa).

Considerando a abordagem da Norma ANSI/NISO Z.39.19:2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, se existirem muitas variações de nomes próprios (pessoas, instituições, organizações, lugares e títulos) sem nenhum tipo de controle, a recuperação da informação bibliográfica não será eficiente (ROSADO; DIAS, 2020).

Contudo, mesmo com a ampla variedade de opções de busca propiciada pelos catálogos digitais, um dos responsáveis pela obra pode ser representado, por exemplo, por diversas maneiras de seu próprio nome, como nome completo ou abreviado, nome pelo qual é mais conhecido, nome anterior ou posterior ao casamento; ou até mesmo um único nome poderia representar mais um responsável (ASSUMPCÃO; SANTOS, 2017a).

E essas variações remontam às funções de um catálogo, conforme discorre Cutter (1904) em sua obra que, embora tenha mais de 100 anos, é de destaque na história da catalogação por ter criado não apenas um código de catalogação, conforme comentam Mey e Silveira (2009), mas elaborado uma verdadeira declaração de princípios.

1. Permitir que a pessoa encontre um livro pelo qual
 - (A) o autor
 - (B) o título
 - (C) o assunto
 } é conhecido
2. Mostrar o que a biblioteca possui
 - (D) de um determinado autor
 - (E) sobre um determinado assunto
 - (F) em um determinado tipo de literatura
3. Para auxiliar na escolha de um livro
 - (G) quanto a sua edição (bibliograficamente)
 - (H) quanto ao seu caráter (literário ou temático) (CUTTER, 1904, p. 12, tradução nossa).

Considerando que uma das funções do catálogo é a de possibilitar que materiais sejam encontrados também por seus autores, Cutter também destaca a importância da listagem de autores:

A listagem de autores do catalogador, mantida alfabeticamente, evita a duplicidade de trabalho. É o registro da forma de um nome (completo) a qual tem sido adotada, com uma nota das autoridades consultadas e suas variações (CUTTER, 1904, p. 133, tradução nossa).

Foi publicada em 1908 a primeira edição do código de catalogação da *American Library Association* (ALA), intitulado *Catalog Rules: Author and Title Entries*, e em 1949 foi publicada a segunda edição, sem que nenhuma das duas contemplasse a criação de registros de autoridade (ASSUMPÇÃO; SANTOS, 2017b).

Quanto ao *Anglo-American Cataloguing Rules*, 2nd edition, 2002 revision (AACR2r) que, embora tenha sido criado já no contexto da automação de bibliotecas, foi elaborado tendo como base o catálogo em fichas e a sua estrutura reflete isto, conforme discutem Assumpção e Santos (2017b). Os autores ainda citam que isso levou à adoção de algumas convenções dos registros nas fichas:

Não há nenhum método ou formato nacionalmente prescrito para a transcrição de dados de autoridade em fichas, mas várias convenções são amplamente utilizadas. A primeira e mais importante convenção é a localização do cabeçalho autorizado no topo da ficha, na mesma forma que ele aparece nas fichas catalográficas. Abaixo dele estão localizadas as remissivas. As remissivas Ver [*see from*] são indicadas pelo prefixo x e as remissivas Ver também [*see also from*] são indicadas pelo prefixo xx. Abaixo delas são dadas quaisquer notas. Finalmente, são dados o nome ou as iniciais do catalogador, a data em que o registro foi preparado e o número de chamada ou o título da publicação que levou à preparação da ficha de autoridade (BURGUER, 1985, p. 23-24 apud ASSUMPÇÃO; SANTOS, 2017b, p. 94).

Os Princípios Internacionais de Catalogação da *International Federation of Library Associations and Institutions* (IFLA), nos quais o AACR2r se baseia, indica que o nome preferido utilizado como ponto de acesso para a entidade deve ser escolhido de acordo com o nome que a identifique “[...] seja porque se encontre mais frequentemente nas manifestações ou porque seja um nome muito aceito, apropriado para os usuários do catálogo [...]” (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016, p. 9). Outro padrão estabelecido pela IFLA para indicação do nome de pessoas indica que quando o nome possuir diversas palavras, ou seja, nome e sobrenome extensos, a escolha da primeira delas para ser um ponto de acesso autorizado deve seguir as convenções adotadas pelo idioma e país mais associados a esta pessoa (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, c2020).

Seguindo a IFLA, o capítulo 22 do AACR2r determina, dentre diversas regras, que a escolha da base do cabeçalho para pessoa seja feita pelo nome que ela é normalmente conhecida, seja seu nome verdadeiro, pseudônimo, alcunha, iniciais, título de nobreza ou outro apelativo; sendo que essa determinação deve considerar a fonte principal de informação de suas obras, publicadas em sua língua ou país de residência/atução. Para os casos de mudança de nome, o AACR2r determina que seja escolhido o nome mais recente, salvo casos em que se acredite que a forma antiga do nome persistirá como a mais conhecida; e para os nomes idênticos, sugere acrescentar como último elemento do cabeçalho informações como datas de nascimento ou morte (FEDERAÇÃO BRASILEIRA DE ASSOCIAÇÕES DE BIBLIOTECÁRIOS, 2004). Rosado e Dias (2020) comentam que as normas estabelecidas por esse capítulo do AACR2r são bem detalhadas, com considerações a respeito de variações na grafia e idioma. Ele tenta cobrir diversas especificidades para que a escolha dos nomes não seja arbitrária, portanto, se trata de uma obra que deve ser considerada para a construção de um controle de autoridade.

Mesmo com os catálogos digitais, que possibilitam a busca e a localização de um maior número de informações, a comunidade científica ainda esbarra nas variações de nomes de autores, seja por desconhecimento da importância dessa padronização por parte do próprio pesquisador ou pela variedade de fontes de informação nas quais o pesquisador pode ser registrado nos metadados de diferentes maneiras. Brandt (1988) comenta a respeito da relação entre catálogos digitais e controle de autoridade em uma época em que seu acesso ainda não era tão difundido como atualmente, mas que revela a preocupação do autor com a gestão dos

dados em ambiente *online*:

À medida que mais e mais bibliotecas migram para catálogos online, o controle de autoridade é visto de perto como uma parte do gerenciamento dos dados que garante consistência e integridade de entradas assim como interrelaciona registros que compartilham um cabeçalho. Quando aplicado a bases de dados bibliográficas, o controle de autoridade é mais que um instrumento utilizado para estabelecer e relacionar cabeçalhos – ele pode de maneira automática e literal “vincular” os registros. Como tal, o controle de autoridade pode ser visto como uma ferramenta essencial para a administração de dados. (BRANDT, 1988, p. 297, tradução nossa).

Considerando o crescente volume informacional no contexto digital, ficou evidente a necessidade de técnicas para otimizar o tempo de tratamento de todos esses conteúdos, incluindo o processo de representação para posterior recuperação. Clack (1988) afirma que quando o computador foi introduzido na realidade das bibliotecas, surgiu uma percepção errada de que todos os problemas de cunho bibliográfico tinham sido solucionados e que garantir a integridade dos registros não era necessário, mas dentro de pouco tempo percebeu-se que os catálogos *online* possuíam falhas na recuperação da informação e isso apontava o controle de autoridade como uma possível solução.

O aumento do uso de arquivos em ambiente digital impulsionou esforços para a criação de padrões, diretrizes e modelos para ampliar o compartilhamento de dados bibliográficos e de autoridades à nível internacional, dentre eles está o modelo *Functional Requirements for Authority Data* (FRAD) (ASSUMPÇÃO; SANTOS; ZAFALON, 2017a). Contudo, antes de falar sobre o FRAD, é fundamental mencionar que em 1998 a IFLA publicou o *Functional Requirements for Bibliographic Records* (FRBR). Sobre eles, pode-se dizer que:

O primeiro objetivo enfocado no desenvolvimento dos FRBR visava proporcionar um quadro estruturado e claramente definido para relacionar os dados contidos nos registros bibliográficos às necessidades dos usuários. Ou seja, os FRBR analisaram os dados necessários à realização da busca bibliográfica pelo usuário, assim como as informações que este esperaria encontrar no registro (MEY; SILVEIRA, 2009, p. 17).

Como o objetivo de expandir esse modelo para que pudesse cobrir também os dados de autoridade, a IFLA criou em 1999 o *Working Group on Functional Requirements and Numbering of Authority Records* (FRANAR) que publicou o FRAD em 2009 (IFLA WORKING GROUP ON FLANAR, 2013). Mey e Silveira apontam que:

Projetou-se esse modelo conceitual visando criar um quadro de referência claramente definido, estruturado, de modo a relacionar os dados dos registros de autoridade às necessidades dos usuários de tais dados; assim como auxiliar na avaliação do potencial de compartilhamento internacional dos dados de autoridade, não restrito ao universo bibliotecário (MEY; SILVEIRA, 2009, p. 38).

Embora a publicação do FRBR tenha suscitado o desenvolvimento de diversos estudos que tentaram utilizar esse modelo conceitual como modelo de dados nos catálogos digitais, as iniciativas tiveram caráter apenas experimental e o FRBR não se consolidou como referência para o desenvolvimento de catálogos (ASSUMPÇÃO; SANTOS; ZAFALON, 2017a).

Também o *Machine Readable Cataloguing* (MARC) possui relação direta com o controle de autoridades. Conforme pontuam Mey e Silveira (2009), o MARC surgiu em 1965/1966, inserido no contexto do avanço computacional da década de 1960, sendo desenvolvido por meio de uma cooperação entre a *Library of Congress* (LC) dos Estados Unidos e a *British Library*. É um formato e padrão para entrada de informações bibliográficas em computadores, ajustando os recursos tecnológicos disponíveis na época à catalogação tradicional, envolvendo a máquina no processo e não automatizando-o.

Em 1969, considerando a experiência com o formato MARC, a LC lançou o MARC II e na década seguinte já era possível encontrar diversas variações do formato em outros países; contudo, em 1976, considerando a preocupação com o controle de autoridade em catálogos digitais, a LC lançou a versão preliminar do documento *Authority work: the creation, use, maintenance, and evaluation of authority records and files*. Em 1981 lançou a primeira edição do documento *Authorities: a MARC Format* e até hoje a LC mantém esse formato atualizado, que hoje se denomina *MARC 21 Format for Authority Data* (ASSUMPÇÃO; ZAFALON; SANTOS, 2017b).

E em 2004, de acordo com Mey e Silveira (2009), iniciou-se o processo de revisão do AACR2r que indicava diversas modificações para uma terceira versão do código, o que resultou no desenvolvimento de um sucessor a ele no ano seguinte, o *Resource Description and Access* (RDA). Apesar de possuir uma forte relação com o AACR2r, o RDA foi projetado para o ambiente digital e possui um escopo mais abrangente (OLIVER, 2011).

São diversas as iniciativas e normas que discorrem sobre a importância do controle de autoridades e o quanto a falta de padronização impacta diretamente na recuperação da informação por parte dos usuários da informação bibliográfica. O mesmo se aplica na descrição e consequente recuperação da informação científica por parte da comunidade científica. A

literatura da área de Ciência da Informação, como visto em trechos anteriores, se preocupa com a consolidação dos nomes de autoridades nos catálogos há algum tempo, contudo com a consolidação dos catálogos digitais e informação disponível de forma mais dinâmica e ampla devido à utilização de fontes informacionais digitais, as soluções propostas pareceram não acompanhar as necessidades cada dia mais emergentes. Algumas soluções aliadas aos mais recentes processos tecnológicos, embora não diretamente envolvidas com a Ciência da Informação, se mostram aliadas ao processo de controle de autoridades em ambiente digital.

3.2 Normas bibliográficas e manuais metodológicos

O aumento exponencial da quantidade de informação disponível em ambientes digitais aliado à ampliação da voz e acesso a uma parcela da população às mídias, conferiu a possibilidade de criar e disponibilizar conteúdos com uma intensidade e velocidade sem precedentes. Quando se fala em conteúdos digitais, não são apenas aqueles produzidos pelas mídias sociais ou na *internet* de forma geral, mas também os conteúdos acadêmico-científicos e industriais, por exemplo, que passam a ser amplamente disseminados e consultados na *web*. Embora sendo realidade do ambiente digital, a preocupação com a falta de padrões e normalização antecede a ele impactando na disseminação das informações. Aponta Dias que:

A normalização é uma característica essencial da atividade humana desde os primórdios da civilização, tendo sido essencial no desenvolvimento da linguagem falada e escrita. Outros aspectos de normalização da atividade humana em épocas passadas podem ser observados quando o comércio entre os povos primitivos exigiu o estabelecimento de medidas padronizadas de peso, dimensão e formas de pagamento (DIAS, 2000, p. 137).

A normalização da atividade humana é intrínseca à organização da sociedade e seus relacionamentos. Encontrar padrões e princípios facilita a troca de informações e conseqüentemente impacta na forma de solucionar e lidar com questões comuns ao cotidiano também em relação aos aspectos corporativos, econômicos ou sociais. As normas são capazes de garantir o orquestrado funcionamento de determinadas atividades e situações.

Desse modo, percebe-se que o conceito de norma apresenta, por um lado, um construto articulador para avaliação de algum fenômeno, e, por outro lado, como artefato institucional para um princípio ou regra. A origem do termo norma está ligada à filosofia grega, e o termo é entendido como motivos ou razões prescritivas (entendendo como prescritiva uma ordenação explícita ou

precisa) referentes a uma determinada realidade do mundo (SILVA, 2013, p. 6).

A atividade científica também segue princípios e regras para se desenvolver de modo a contribuir com resultados de pesquisa que sejam relevantes e que sejam utilizados como insumo por outros pesquisadores. Para isso, é fundamental que a pesquisa científica siga padrões em relação, por exemplo, à estruturação de sua metodologia, à redação de argumentos e discussão e ao embasamento teórico. De acordo com Garvey (1979), que possui uma obra relevante e que embasa discussões sobre comunicação científica, este processo de comunicação em ciência se dá desde o momento de concepção da ideia de pesquisa até o momento em que seus resultados são divulgados e aceitos pela comunidade científica. Neste contexto é interessante o trabalho de Bourdieu (1983), o qual fala sobre o campo científico como aquele que é cenário de uma disputa científica e concorrência, incluindo o processo de revisão por pares.

O processo de aceitação pela comunidade científica perpassa justamente a avaliação por pares, sob a qual as publicações são submetidas antes de serem agregadas a um periódico científico, como um reconhecimento de qualidade; e pode envolver a aceitação no que diz respeito ao nível de utilização da pesquisa para desenvolvimento de outras, por meio da análise de citações recebidas.

De acordo com Rodrigues, Lima e Garcia (1998) o trabalho científico pode ser analisado pela perspectiva de seu conteúdo e de sua forma: seu conteúdo como a pertinência da pesquisa científica para a área de conhecimento em questão, que é avaliado pelos pares; e sua forma como a parte “instrumental”, no sentido de posicionamento nos ritos acadêmicos usuais, como o desenvolvimento inicial do pesquisador e suas habilidades em manusear dados e referencial teórico.

Inserida no contexto de análise da pesquisa científica em seu aspecto formal, pode-se dizer que também inclui a avaliação em relação à adequação do trabalho às normas. A normalização, sobretudo em relação à atividade científica, está intrinsecamente relacionada com a atividade de profissionais bibliotecários, como aqueles que oferecem suporte e guiam os pesquisadores neste processo. Neste sentido, Silva (2013, p. 7) menciona que o conceito de norma utilizado no contexto da Biblioteconomia contempla princípios para organização de acervos embasados nos “[...] estudos de representação da informação, especialmente pelos vieses da catalogação, classificação e indexação, além de outros elementos que promovam o acesso à informação”.

A pesquisa científica precisa da normalização para garantir sua efetiva recuperação pela

comunidade acadêmica. Meadows (1999, p. 120) comenta que a estrutura das publicações científicas melhorou muito ao longo dos anos e que uma parcela considerável dos trabalhos publicados não apresentava referências e quando o faziam não eram inseridos muitos detalhes a seu respeito, o que implicava em sua não localização. E para isso, são utilizadas as normas técnicas que podem ser definidas como aquelas que tem como principal objetivo normalizar e padronizar produtos e serviços visando sua qualidade, sendo consideradas como resultado de um processo que indica características de produtos e serviços aferindo-os qualidade e confiabilidade (CRESPO; RODRIGUES, 2011).

O movimento no sentido de promover a normalização da apresentação, como se assinalou antes, reflete as pressões crescentes exercidas sobre a comunicação científica, mais especificamente como resultado de sua rápida expansão, que dificultou ainda mais para os leitores a localização de informações relevantes. A normalização ajuda nesse sentido (MEADOWS, 1999, p. 120).

A preocupação com a normalização não é recente. Cunha (1973) menciona que a velocidade e eficiência com a qual se transfere a informação produzida depende de diversos fatores e entre eles está sua estrutura.

Essa difusão de informações precisa ser processada de alguma forma e através de sistema que permita pronta assimilação. Isso levanta a questão de padrões ou normas reconhecíveis em bases nacionais e, especialmente, internacionais. Em todas as fases do processo informativo – começando com a expressão do pensamento pelo autor de um trabalho intelectual atinente a qualquer campo de especialização e concluindo com os serviços que esse trabalho venha a prestar aos seus possíveis leitores – a padronização ou normalização é altamente valiosa (CUNHA, 1973, p. 59).

A normalização da informação científica produzida e publicada depende de subsídio técnico de normas e padrões estabelecidos por organizações voltadas para este fim. De acordo com Rodrigues, Lima e Garcia (1998), desde a década de 1960, organizações internacionais que atuam na área de documentação científica como a Organização das Nações Unidas (UNESCO), Federação Internacional de Informação e Documentação (FID), IFLA, *International Organization for Standardization* (ISO) e *International Science Council* (ICSU) tentam prescrever normas.

No Brasil, a Associação Brasileira de Normas Técnicas (ABNT), criada em 1940, é a organização voltada à normalização, pautada nas normas da ISO. As normas ABNT que são

voltadas para a produção acadêmico-científica são parte do escopo do Comitê Brasileiro de Informação e Documentação (CB 014). Segundo o CB 014, sua atuação está voltada para a “normalização no campo da informação e documentação compreendendo as práticas relativas a bibliotecas, centros de documentação e informação, serviços de indexação, resumos, arquivos, ciência da informação e publicação” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, c2014).

Algumas das normas ABNT que possuem o CB 014 como responsável pelas publicações:

- a) ABNT/NBR 10518 – Informação e documentação – Guias de unidades informacionais;
- b) ABNT/NBR 10519 – Informação e documentação – Critérios de avaliação de documentos de arquivo;
- c) ABNT/NBR 10520 – Informação e documentação – Citações em documentos – Apresentação;
- d) ABNT/NBR 10525 – Informação e documentação – Número padrão internacional para publicação seriada – ISSN;
- e) ABNT/NBR 10719 – Informação e documentação – Apresentação de relatórios técnico-científicos;
- f) ABNT/NBR 12225 – Informação e documentação – Lombada – Apresentação;
- g) ABNT/NBR 12676 – Informação e documentação – Métodos para análise de documentos – Determinação de seus assuntos e seleção de termos de indexação – Procedimento;
- h) ABNT/NBR 14724 – Informação e documentação – Trabalhos acadêmicos – Apresentação;
- i) ABNT/NBR 15287 – Informação e documentação – Projeto de pesquisa;
- j) ABNT/NBR 15437 – Informação e documentação – Pôsteres técnicos e científicos;
- k) ABNT/NBR 2108 – Informação e documentação – Número padrão internacional de livro (ISBN);
- l) ABNT/NBR 5892 – Informação e documentação – Norma para datar;
- m) ABNT/NBR 6021 – Informação e documentação – Publicação periódica técnica e/ou científica – Apresentação;
- n) ABNT/NBR 6022 – Informação e documentação – Artigo em publicação periódica científica impressa;

- o) ABNT/NBR 6023 – Informação e documentação – Referências – Elaboração;
- p) ABNT/NBR 6024 – Informação e documentação – Numeração progressiva das seções de um documento – Apresentação;
- q) ABNT/NBR 6025 – Informação e documentação – Revisão de originais e provas;
- r) ABNT/NBR 6027 – Informação e documentação – Sumário – Apresentação;
- s) ABNT/NBR 6028 – Informação e documentação – Resumo – Apresentação;
- t) ABNT/NBR 6029 – Informação e documentação – Livros e folhetos;
- u) ABNT/NBR 6032 – Informação e documentação – Abreviação de títulos de periódicos e publicações seriadas;
- v) ABNT/NBR 6033 – Informação e documentação – Ordem alfabética;
- w) ABNT/NBR 6034 – Informação e documentação – Índice – Apresentação;
- x) ABNT/NBR 9578 – Informação e documentação – Arquivos – Terminologia.

E dentre toda a informação científica que deve passar por um processo de normalização, subsidiado tecnicamente pelas normas técnicas, estão também os nomes de autores e de instituições, que também devem seguir padrões nos documentos acadêmicos científicos para que seja possível conhecê-los enquanto autores ou enquanto insumo utilizado para o desenvolvimento da pesquisa, ou seja, referencial teórico.

Entre as normas ABNT voltadas para a normalização de trabalhos acadêmico-científicos e mencionadas anteriormente, quatro delas fornecem orientações específicas a respeito da inserção do nome dos autores do documento. Embora outras normas listadas também mencionem a necessidade de inserir o nome dos autores, elas não fornecem detalhes de como essa inserção deve ser feita, por isso não serão destacadas neste momento.

Entre as quatro que orientam a forma com a qual os autores devem figurar no documento, estão (Quadro 1):

Quadro 1 – Orientações das normas sobre a inserção dos autores do documento

Norma	Orientações sobre a inserção do nome dos autores do documento
ABNT/NBR 6022 – Informação e documentação – Artigo em publicação periódica científica	<ul style="list-style-type: none"> a) o nome do autor deve ser inserido: prenome, abreviado ou não, e o sobrenome; b) quando houver mais de um autor, os nomes podem constar na mesma linha, separados por vírgula, ou em linhas distintas; c) deve constar um breve currículo de cada autor, com vinculação

impressa	institucional e endereço para contato (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018).
ABNT/NBR 6023 – Informação e documentação – Referências – Elaboração	<p>a) os autores devem ser indicados pelo último sobrenome (em letras maiúsculas), seguido pelo prenome e demais sobrenomes, abreviados ou não, de acordo com o que consta no documento;</p> <p>b) os autores devem ter seus nomes separados por ponto e vírgula e espaço;</p> <p>c) convém a padronização de prenomes e sobrenomes de autores, caso apareçam de formas diferentes em documentos distintos;</p> <p>d) se o documento tiver até três autores, todos devem ser indicados;</p> <p>e) se o documento tiver quatro ou mais autores, convém indicar todos, mas permite-se a indicação do primeiro seguido da expressão <i>et al</i>;</p> <p>f) autores que possuam nomes hispânicos, compostos, com grau de parentesco ou com sobrenomes com prefixos, devem seguir os exemplos: GARCÍA MÁRQUEZ, Gabriel; ASSAF NETO, Alexandre; ESPÍRITO SANTO, Miguel Frederico de; D’AMBROSIO, Ubiratan;</p> <p>g) quando a responsabilidade da obra for conjunta, a entrada deve ser realizada pelo nome do responsável, seguido da abreviação do tipo de participação (organizador, editor, coordenador, etc.) em letras minúsculas;</p> <p>h) obras publicadas sob pseudônimo devem exibir o nome deste na referência;</p> <p>i) obras psicografadas devem conter o nome do espírito em primeiro lugar;</p> <p>j) para obras adaptadas, consta o adaptador em primeiro lugar;</p> <p>k) obras com responsabilidade de empresas, organizações, associações, etc. devem ser indicadas pela forma conhecida ou como se destacada no documento, por extenso ou abreviada;</p> <p>l) convém a padronização de nomes para as entradas por pessoas jurídicas, caso apareça de forma diferente em documentos distintos;</p> <p>m) quando a responsabilidade do documento for de uma instituição governamental, o nome deve ser precedido pelo nome do órgão superior;</p> <p>n) quando a instituição que é vinculada a um órgão superior possui uma denominação específica, a entrada é feita diretamente por seu nome;</p> <p>o) quando a instituição for homônima, deve se indicar ao final, entre parênteses, a unidade geográfica que a identifica;</p> <p>p) obras resultantes de eventos possuem a entrada pelo nome do evento;</p> <p>q) quando a autoria do documento for desconhecida, a entrada é feita pelo</p>

	título (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018).
ABNT/NBR 10520 – Informação e documentação – Citações em documentos – Apresentação	<p>a) as citações, as chamadas pelos sobrenomes do autor, instituição ou título devem ser em letras maiúsculas e minúsculas, sendo que quando estiverem entre parênteses devem ser em letras maiúsculas;</p> <p>b) considerando o sistema autor data, a indicação da fonte é realizada pelo sobrenome de cada um dos autores do documento ou pelo nome da entidade responsável até o primeiro sinal de pontuação, seguido pela data de publicação do documento e página de citação; no caso de citação direta, são separados por vírgula e entre parênteses (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002).</p>
ABNT/NBR 12225 – Informação e documentação – Lombada – Apresentação	<p>a) o nome do autor deve ser impresso no mesmo sentido da lombada;</p> <p>b) caso o documento tenha mais de um autor, os nomes devem ser impressos um abaixo do outro na lombada, separados por sinais de pontuação, espaços ou outros sinais gráficos, permitindo a abreviatura ou omissão dos prenomes (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2004).</p>

Fonte: elaborado pela autora.

Ao verificar as orientações fornecidas pelas normas para inserção dos nomes de seus responsáveis, percebe-se a orientação comum a respeito dos sobrenomes dos autores, em letras maiúsculas, e dos prenomes, abreviados ou não, de acordo com o que consta no documento. Embora a norma seja enfática em relação ao modo como os autores devem ser inseridos, sua indicação pela inserção dos prenomes de acordo com o que consta no documento, pode criar variações devido à forma como foi diagramado e publicado este documento original, ainda que a norma também oriente que um autor com diferentes formas de entrada pode ter seu nome padronizado. A decisão de seguir o documento que possui em mãos para inserir os nomes dos autores, da forma que contam nele, volta a atenção para os editores e diagramadores dos documentos acadêmico científicos que estão sendo utilizados e sua responsabilidade com o controle de autoridades, se estão ou não seguindo também algum padrão. O mesmo ocorre com obras sob responsabilidade de instituições, entrada que a norma orienta que seja feito por seu nome conhecido ou pela forma que figura no documento.

Outro destaque está na orientação a respeito da colocação de todos os autores no momento de elaborar as referências bibliográficas do documento acadêmico científico: embora

seja conveniente indicar todos os autores da referência, permite-se que se indique apenas o primeiro seguido da expressão *et al.* (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018). Como existe a permissão por parte da norma para incluir apenas o primeiro autor, é fundamental refletir a respeito das consequências de se omitir autores das referências de documentos e como isso impacta em sua recuperação.

Contudo, quando se fala da normalização de trabalhos acadêmico-científicos, não se presume a utilização de apenas uma norma técnica, como a ABNT. A normalização da produção científica possui particulares e utilização de diferentes normas de acordo com as instituições e as áreas de conhecimento, ou seja, há casos em que as normas são adaptadas e utilizadas de uma forma específica em cada instituição, ou instituições que adotam outras normas que não necessariamente a ABNT.

O formato Vancouver, que é amplamente utilizado por publicações acadêmico-científicas na área de Ciências da Saúde, foi adaptado pela *National Library of Medicine* (NLM) dos Estados Unidos com base em trabalhos do *International Committee of Medical Journal Editors* (ICMJE), que teve a sua primeira reunião em 1978 na cidade de Vancouver, no Canadá (HOPPEN; MAGNUS; VANZ, 2018). A NLM publica o *Citing Medicine*, um guia que auxilia os autores a fazer as referências de suas publicações e os periódicos a estabelecer padrões. A norma Vancouver, diferente da ABNT não utiliza o sistema de citação autor-data, mas sim um modelo numérico. De acordo com o que estabelece o *Citing Medicine*, os nomes dos autores são referenciados sempre com os prenomes abreviados e os sobrenomes descritos sempre em minúsculo (PATRIAS; WENDLING, 2007), diferente das orientações da ABNT. Os prenomes abreviados podem contribuir esteticamente para o padrão da referência, contudo a conduta impacta na correta identificação e recuperação de informações a respeito dos autores.

Bellini et al. (2006) em um estudo a respeito das peculiares de cada área do conhecimento na normalização de documentos, argumenta que na área de Ciências Humanas é utilizada uma norma que indica a possibilidade de inserção dos prenomes dos autores por extenso considerando que seus documentos possuem uma maior abrangência, quanto a área temática, e por isso aumenta a possibilidade de especialistas homônimos; já na área de Ciências da Saúde, segundo sua argumentação, as especialidades são mais restritas e os autores possuem o hábito de publicar em grupo, o que tornaria o cruzamento entre o sobrenome do autor e a especialidade suficientes para encontrar informações e produções dos autores. Mesmo com as peculiaridades de cada área, é necessário que as estratégias que possibilitem uma rápida e eficaz identificação dos autores de produções acadêmico-científicas sejam otimizadas, sem a

necessidade de realizar maiores cruzamentos para fazê-lo.

A norma da *American Psychological Association* (APA), como o próprio nome sugere, é amplamente utilizada nos trabalhos acadêmico-científicos na área da Psicologia. Foi criada em 1929 em Washington e surgiu com o objetivo de melhorar a qualidade das referências citadas (HOPPEN; MAGNUS; VANZ, 2018). Assim como a ABNT, a APA utiliza o sistema autor-data para realizar as citações e referencia os nomes dos autores por seus sobrenomes, em letras minúsculas, e os prenomes abreviados.

As normas ABNT, Vancouver e APA não apenas alguns exemplos adotados à nível nacional e internacional que tratam a descrição dos autores, dentre outros diversos detalhes, de acordo com suas particularidades e que nem sempre convergem entre si. A comunidade científica, engajada em contribuir com o desenvolvimento da ciência, deve ser orientada a se atentar para as normas com intuito muito mais profundos do que simplesmente atender critérios exigidos por instituições de ensino superior ou periódicos científicos.

[...] a entropia ou a desordem que se observa na comunicação entre cientistas não resulta apenas do número excessivo de meios, mas da falta de uma normalização e de uma padronização adequada desses meios. Os canais da comunicação científica estão bloqueados por veículos que, sobre serem excessivos em número, continuam a ser produzidos sem nenhum respeito às recomendações internacionais e às normas nacionais que regulam a matéria (FONSECA, 1973, p. 55).

É útil que os pesquisadores entendam o real motivo de cuidar dos detalhes “formais” ou “normativos” de seus trabalhos, objetivando que seja feita sua recuperação enquanto referência, e suas argumentações e citações também estejam corretamente identificadas e sejam úteis para seus pares. É prudente também que os pesquisadores sejam orientados sobre a gestão de seu tempo de pesquisa, quanto ao tempo de escrita e ao tempo para se dedicar à sua normalização, embora atualmente estejam disponíveis diversos gerenciadores de referências, como EndNote, Mendeley, Zotero, etc. que podem tornar o processo de padronização um pouco menos árduo, embora não dispensem o olhar atento do autor do trabalho.

As inúmeras tarefas típicas do trabalho intelectual tanto de orientação como de criação, propriamente dita, demandam, portanto, atenção equilibrada do pesquisador ao duplo aspecto da qualidade: há que cuidar do conteúdo e há que atentar à qualidade de apresentação formal. E, em última instância, a qualidade formal é propiciada utilizando-se o suporte proporcionado pela normalização (RODRIGUES; LIMA; GARCIA, 1998, p. 152).

A tarefa de orientação e conscientização da comunidade científica a respeito da normalização e uso adequado dos nomes dos autores, por exemplo, poderia ser demandada para além dos profissionais bibliotecários que pertencem às instituições de ensino, mas ser parte da discussão proposta e abordada por manuais metodológicos. Assim como esses manuais trazem orientações a respeito de práticas de pesquisa desde, por exemplo, concepção de projetos, até detalhes a respeito de pesquisas quantitativas e qualitativas, seria benéfico que essas obras tão consultadas e utilizadas como referências por muitos pesquisadores, também pudessem discorrer sobre essas temáticas.

3.3 Uso da tecnologia para desambiguar autorias e instituições

Os avanços das ferramentas tecnológicas proporcionam soluções e melhorias de processos que fora do ambiente digital demandariam tempo e esforços de profissionais em atividades como, por exemplo, a construção do controle de autoridades, como visto anteriormente. Embora seja uma atividade essencial e que impacte diretamente no processo de padronização e recuperação da informação científica, a área de Ciência da Informação e os profissionais precisam se atentar para as alternativas que a parceria com outras áreas do setor tecnológico possa suscitar e que otimizem não somente o processo, mas também a qualidade de seu resultado.

O termo *Author Name Desambiguation* (AND) é encontrado na literatura científica e aborda técnicas ou discussões a respeito da aplicação de soluções tecnológicas que com a parceria dos profissionais e acadêmicos da Ciência da Informação e da Ciência da Computação buscam solucionar ou minimizar os problemas de ambiguidade nas autorias. Ao longo do trabalho este termo será comumente tratado também como desambiguação de nomes de autores.

Smalheiser e Torvik (2009) apontam quatro desafios para essa desambiguação: um pesquisador pode publicar com diferentes nomes ao longo de sua trajetória acadêmica; podem existir pesquisadores com nomes iguais; metadados incompletos por parte, por exemplo, dos periódicos científicos ou bases de dados, que podem armazenar os nomes dos autores sem outras informações como sua nacionalidade, graus acadêmicos e outras informações de identificação; e é crescente o número de publicações não apenas com múltiplas autorias, mas também com múltiplas áreas e instituições envolvidas. Este conjunto de fatores apontados suscitam reflexões à respeito da importância de técnicas e processos que consigam otimizar essa desambiguação aliadas a políticas estabelecidas por parte das fontes de informação

científica, como os periódicos científicos ou bases de dados, buscando que os metadados dos registros sejam indexados com uma riqueza de informações suficientes para então possibilitar a desambiguação realizada pelas ferramentas tecnológicas.

Esperamos que os projetos de desambiguação de nomes de autores comecem a dizer o caminho no qual nomes ou outros metadados são codificados nas publicações. Atualmente, cada editor codifica nomes de formas distintas; não há um padrão para representação de afiliações, e no geral não há captura geral de certos tipos de informação (ex. cargos, graus obtidos, etc.) que seriam muito úteis para a desambiguação. Além disso, atualmente o metadado não é público para outros lerem e exportarem, o que ajudaria na coordenação de esforços de desambiguação através de editores e disciplinas (SMALHEISER, TORVIK, 2009, p. 33, tradução nossa).

Mugnaini et al. (2012) desenvolveram um estudo em que apresentam um método automático para verificar a ocorrência de erros nos dados dos autores no Dedalus, a base de produção científica da Universidade de São Paulo (USP); para isso foi desenvolvido um algoritmo para processamento dos dados com o objetivo de identificar as distintas formas com as quais o nome de um docente aparece nos registros do Dedalus comparando com um outro banco de dados corporativo de todos os colaboradores da universidade, para que fosse possível verificar possíveis inconsistências. O estudo concluiu que existem registros de autoria com problemas e que isso impacta diretamente em indicadores de produtividade ou estudos sobre a produção científica institucional.

No estudo mencionado, o desenvolvimento de um método de verificação dos registros via algoritmos otimizou o tempo necessário para identificar quais são os registros que precisariam ser padronizados. Como visto anteriormente, com o aumento de conteúdos digitais, incluindo aqueles armazenados e oferecidos por bibliotecas digitais ou recursos para pesquisa, cresceu também a necessidade de distinguir e padronizar autorias. Smalheiser e Torvik (2009) apontam duas razões porque desambiguar nomes manualmente nem sempre é possível: muitas vezes as bibliotecas digitais captam os metadados de seus materiais de outros serviços de informação e essa variedade de fontes pode utilizar diferentes controles de autoridade, o que faz com que o aumento no número de materiais e variações torne inviável o processo de ajuste dos nomes; e buscas em ambientes *web* são realizadas em uma quantidade de fontes que não podem ser manualmente tratadas.

A quantidade de informações recuperadas em buscas realizadas em ambiente *web*, independentemente da assertividade e relevância, tornam o processo de tratamento

informacional ou desambiguação de autorias cada vez menos prático de ser realizado manualmente. Para Veve (2009), em contrapartida, a desambiguação automática de nomes nunca poderia ser realizada apenas por soluções tecnológicas, e que sempre o olhar de um humano seria necessário para concluir o processo com sucesso. Para Elliott (2010), uma abordagem híbrida, que trabalhe com soluções tecnológicas aliadas à verificação humana, parece ser uma excelente alternativa; contudo, é necessário que cada caso abordado leve em consideração suas particularidades.

A escolha de uma alternativa para solucionar ou auxiliar processo de controle das autorias deve ser feita de acordo com o contexto do recurso informacional como o acervo editorial de determinada biblioteca digital, e que essa tecnologia seja de fato estudada pelos responsáveis do projeto, não somente com o objetivo de solucionar a falta de padronização que levou à busca pela solução, mas também para procurar entender quais são os caminhos para evitar que este problema volte a acontecer.

Elliott (2010) comenta que a pesquisa sobre a desambiguação de nomes na comunidade científica é realizada por bibliotecários, cientistas da informação e por profissionais ligados à Ciência da Computação; e os esforços são feitos de maneira local e internacional, evidenciando que qualquer sistema com armazenamento de metadados em algum momento terá de lidar com essa questão e decidir qual método adotar para solucionar ou ao menos minimizar inconsistências.

Ao retomar os esforços realizados, sobretudo pelos profissionais bibliotecários e pesquisadores da área de Ciência da Informação, que foram essenciais para o desenvolvimento das atividades de catalogação e recuperação de materiais desde o início do tratamento informacional de materiais bibliográficos, é possível então vislumbrar na desambiguação de nomes de autores um processo que, diferentes destes, não está completamente voltado para a simples distinção de nomes, embora ela ainda seja indispensável. Smalheiser e Torvik (2009) afirmam que o processo de desambiguação vai muito além dos nomes, ele é voltado para a identificação dos indivíduos de uma maneira mais ampla:

Enquanto o reconhecimento do nome de uma entidade tenta identificar qual George W. Bush está sendo mencionado em um artigo específico, a desambiguação de autor incorpora informação além das publicações do indivíduo, inclui também recursos que envolvem computação extensiva e conhecimento externo de fontes externas. Assim, imaginamos o processo de desambiguação do nome do autor como indo consideravelmente além de tarefas de componentes, como classificação ou agrupamento, para fornecer uma análise aprofundada ou perfil da pessoa (SMALHEISER; TORVIK,

2009, p. 9, tradução nossa).

Na literatura científica, é possível encontrar diversos estudos que abordam técnicas ou metodologias para realizar a desambiguação de nomes de autores, sejam no formato de revisão, ou seja, listando as alternativas disponíveis nas épocas em que os estudos foram desenvolvidos, ou análises voltadas para fontes de informação específicas, como os métodos e esforços de bases de dados científicas. No presente trabalho serão mencionados alguns dos principais estudos e algumas das técnicas por eles abordadas.

Smalheiser e Torvik (2009), autores de diversos estudos sobre desambiguação de nomes de autores, comentam que parte das abordagens de pesquisa sobre a temática compartilham as linhas gerais com o *machine learning* ou aprendizado de máquina, com o objetivo de agrupar ou classificar trabalhos de acordo com seus autores; existem diversas abordagens estudadas ao longo dos anos, mas é preciso ter em mente que não podem ser comparadas quanto a performance, considerando que cada sistema é desenvolvido de acordo com um tipo de dado, mas no geral eles são aplicáveis a grande parte das bases de dados bibliográficas existentes.

Técnica também comentada na revisão de Smalheiser e Torvik (2009) é o *stylometry*, embora não tenha sido utilizada para literatura científica, trata-se de um processo que é utilizado para diferenciar os autores de acordo com seus estilos de escrita, incluindo análise e comparação de estrutura das frases, gramática, contexto da publicação (post de blog, artigo acadêmico, etc.), entre outros aspectos.

Outra técnica é a *record linkage*, que identifica registros em mais de uma fonte de informação e consegue conectá-los ao mesmo autor por meio da análise das variantes de seu nome e sua afiliação (CHURCHES et al.; 2002). Contudo, ao longo de sua trajetória acadêmica, os autores podem ter mais de uma afiliação e esse relacionamento entre endereço e autor pode apresentar falhas.

Smalheiser e Torvik (2009) apresentam em sua revisão um projeto próprio, chamado *Author-ity Project*, que teve como primeiro passo a desambiguação de autorias na base Medline, destacando o nível de dificuldade apresentado por esta fonte, que possui publicações científicas de uma mesma grande área do conhecimento, que é a Ciências da Saúde, e podem existir autores que tenham os mesmos nomes, publiquem sobre os mesmos assuntos e tenham afiliações iguais.

O *Author-ity Project* possui um modelo estatístico que é baseado em uma comparação de vetores que incorporam seis itens dos registros do Medline, as palavras do título, nome do periódico, nome dos co-autores, os descritores *Medical Subject Headings* (MeSH), idioma e

palavras da afiliação, além de verificar a estrutura do nome do autor (quais são as iniciais do nome, sufixo, etc); em um segundo momento, o *e-mail* dos autores também foi incorporado como um item (SMALHEISER; TORVIK, 2009).

Os autores do projeto concluem a revisão afirmando que não existe uma única técnica ou projeto para realizar a desambiguação de nomes de autores, tendo em vista que cada base de dados, biblioteca digital ou coleções possuem suas particularidades e problemas específicos que precisam ser particularmente estudados (SMALHEISER; TORVIK, 2009).

Outra questão que também precisa ser levada em conta é o surgimento e aperfeiçoamento de técnicas de desambiguação de nomes de autores que considerem a possibilidade de fazê-lo através de diversas fontes de informação, com o objetivo de conseguir de fato diferenciar os autores da literatura científica onde quer que eles estejam hospedados. As iniciativas voltadas especificamente para algumas bases de dados, tendo em vista suas necessidades e características próprias de suas informações, são grandes avanços, porém não se pode perder de vista a recuperação única dessa literatura científica para além das barreiras de cada base de dados.

[...] diferentes conjuntos de dados de desambiguação precisam ser ligados entre si – para exemplo, é desejável identificar os mesmos indivíduos e suas publicações quando eles aparecem em uma variedade de bancos de dados biomédicos públicos e proprietários (sobrepostos) e quando aparecem em outras disciplinas (bancos de dados de engenharia / ciência da computação) ou em outros tipos de publicações, como livros, revistas e patentes (SMALHEISER; TORVIK, 2009, p. 31, tradução nossa).

Elliott (2010), por sua vez, também fez uma revisão de abordagens ou técnicas de desambiguação de nomes de autores, ou AND, e diferentemente de Smalheiser e Torvik (2009), seu foco não foi apenas aquelas que envolviam a tecnologia de alguma forma, mas também listou algumas que envolviam trabalho essencialmente manual.

O *Library of Congress Authority File* (LCAF) contém registros de autoridade que foram manualmente criados e sua manutenção e atualização demanda um intenso trabalho manual que é compartilhado com bibliotecas participantes (ELLIOTT, 2010).

Outro projeto mencionado por Elliott (2010) é o *British Library's Name Project*, que foi uma resposta ao crescente número de repositórios institucionais no Reino Unido, e seu desenho corresponde à criação de um número único, ou *unique identifier* (ID) que faria a ligação entre diferentes formas dos nomes dos autores.

O *Friend of a Friend* (FOAF) é um sistema de codificação de informações sobre pessoas

de forma que várias informações sobre elas sejam compiladas em um único documento, podendo reunir nomes, e-mails, fotos e até páginas pessoais, utilizando o *Resource Description Format* (RDF) para codificar relacionamentos entre pessoas e informações; contudo não parece ser uma solução ideal para bibliotecas digitais, sobretudo considerando que elas frequentemente possuem apenas informações como o último nome e a primeira inicial dos autores (ELLIOTT, 2010).

Elliott (2010) finaliza sua revisão afirmando que a comunidade de bibliotecários e pesquisadores da Ciência da Informação reconheceram o problema da desambiguação de nomes de autores e têm buscado soluções, mas algumas delas ainda são baseadas em esforços manuais, enquanto outras investem em alternativas automatizadas.

Ferreira, Gonçalves e Laender (2012) também realizaram uma pertinente revisão a respeito das técnicas de desambiguação de nomes de autores, porém diferente de Smaiheiser e Torvik (2009) e Elliott (2010), os autores desse trabalho buscaram agrupar os estudos encontrados por seus respectivos métodos. Um dos destaques do estudo está no levantamento de desafios que devem ser considerados no momento de traçar soluções para a desambiguação de nomes em bibliotecas digitais: **poucos dados nas citações**, considerando que na maioria das vezes poucas informações ficam disponíveis como, por exemplo, autores (muitas vezes contém apenas o sobrenome e a primeira inicial), título da publicação, periódico (algumas vezes abreviado) e ano, dessa forma é fundamental que novas estratégias sejam traçadas para incluir informações adicionais; **muitos casos ambíguos**, considerando que vários métodos exploram heurísticas baseadas em co-autor, assumindo hipóteses de que muito raramente referências ambíguas terão coautores em comum que também têm nomes ambíguos, ou é raro que dois autores com nomes muito semelhantes trabalhem na mesma área de pesquisa, contudo quando essas hipóteses falham, os erros são muito difíceis de serem corrigidos, e essas falhas podem acontecer sobretudo quando autores possuem nomes asiáticos; **citações com erros**, que muitas vezes são difíceis de serem detectados; **eficiência**, considerando o crescente número de publicações científicas disponibilizadas atualmente; **diferentes áreas de conhecimento**, considerando que cada uma delas possui suas particulares e comportamentos distintos entre si como ter mais ou menos publicações em co-autoria, ou seja, os métodos de desambiguação adotados devem estar preparados para essas variações; **desambiguação incremental**, no sentido da desambiguação ser realizada sempre que uma nova publicação for incluída na biblioteca digital; mudança no perfil do autor, considerando que pesquisadores podem mudar de área e de temática trabalhada ao longo de sua trajetória acadêmica, e o método de

desambiguação deve prever esse tipo de mudança e contorná-la da melhor maneira, ainda que a maioria dos métodos atuais ignorem essa possibilidade; e **novos autores**, ou seja, o sistema deve ser capaz de identificar nomes ambíguos que já constam ou não cadastrados na base (FERREIRA, GONÇALVES, LAENDER, 2012).

O estudo realizado por Hussain e Asghar (2017) também faz uma revisão e discorre a respeito de diferentes técnicas e as classifica de acordo com os seus métodos, apresentando ao final alguns desafios observados: **casos de autores únicos**, uma vez que a maioria das técnicas de desambiguação de nomes de autores consideram que os nomes podem ser desambiguados por meio da identidade dos co-autores, mas essas técnicas apresentam resultados inferiores ou falharam na desambiguação quando a publicação possuía apenas um autor; **desequilíbrio na distribuição de publicações**, considerando que a maioria dos autores possuem menos publicações e a minoria dos autores possuem muitas publicações, esse fato precisa ser considerado no uso de técnicas de desambiguação; **escalabilidade**, considerando que a maioria das técnicas existentes para desambiguação de nomes de autores foram unicamente aplicadas em bancos de dados específicos e com limitações, e quando foram aplicados em larga escala apresentaram falhas no processo; **representação não padronizada dos nomes**, considerando que cada editora ou periódico científico adota uma maneira do nome em seus metadados como, por exemplo, último nome e apenas as iniciais dos demais nomes, o que dificulta a ação das técnicas de desambiguação; **publicações interdisciplinares**, ou seja, as técnicas precisam considerar as características específicas de publicação de diferentes áreas do conhecimento, incluindo as múltiplas autorias; **desambiguação incremental** e **novos autores**, com observações semelhantes às realizadas por Ferreira, Gonçalves e Laender (2012).

Os desafios propostos pelos dois estudos são fundamentais no momento da criação de uma proposta de método para a desambiguação dos nomes de autores, considerando que são pontos sensíveis e que impactam diretamente na recuperação exata de conteúdos nas fontes de informações nas quais estes métodos são aplicados.

É perceptível também a real necessidade de que se trabalhem juntos profissionais de diferentes áreas do conhecimento com potencial para contribuir com o desenvolvimento das técnicas de desambiguação de nomes de autores, uma vez que um dos desafios mencionados nos estudos está justamente nas particularidades da literatura das áreas, o que pode determinar ou não que uma técnica seja bem-sucedida se aplicada a uma fonte de informação especializada em determinada área do conhecimento. Para que se alcance uma técnica capaz de atuar em bases multidisciplinares ou em bases distintas, essa necessidade deve ser considerada.

Recomenda-se reforçar também a participação ativa de profissionais das áreas de Biblioteconomia e Ciência da Informação, que precisam estar preparados para utilização de novas técnicas ou contribuir com seu desenvolvimento, ocupando novos espaços e contribuindo para otimização de processos tão fundamentais e que são a base de seu desenvolvimento enquanto área de atuação e pesquisa.

À medida que a biblioteca e a Ciência da Informação se tornarem progressivamente mais centrados na pessoa, e não somente centrados no documento, esperamos ver ondulações que vão afetar o mundo da publicação, web semântica, design de motores de busca e a indexação de coleções de dados (SMALHEISER; TORVIK, 2009, p. 38-39, tradução nossa).

Especificamente no caso das instituições, a desambiguação por nomes de instituições, ou *institution name disambiguation* (IND), pode reunir diferentes formas do nome de uma instituição numa mesma entidade institucional, fazendo com que a normalização dos nomes institucionais consiga otimizar a precisão da recuperação de registros em bases de dados (HUANG et al., 2020).

Embora a lógica da desambiguação de autores e instituições, perpassasse pelas mesmas questões e objetivos gerais, que é oferecer qualidade na recuperação informacional que impacta diretamente na atribuição das publicações às instituições e conseqüentemente nos indicadores científicos, que são utilizados em diversos tipos de avaliação, *rankings* ou análises de fomento, destaca-se também que são muitos os estudos voltados exclusivamente para instituições.

Huang et al. (2020) afirmam que, os estudos relevantes sobre a temática podem se dividir em três categorias:

- a) **A IND baseada em similaridade de palavras:** verifica o grau de similaridade entre duas *strings*, geralmente por análise percentual; as métricas e algoritmos mais utilizados são similaridade por cosseno, distância Euclidiana, distância de Hamming e a distância de edição;
- b) **A IND baseada em regras e estatísticas:** combina modelo de regras, ou *rule model*, com técnicas para identificar similaridades de palavras via algoritmo;
- c) **A IND baseada em recurso genérico de classificação:** utiliza técnicas de aprendizado por máquina, considerando que cada instituição é representada por um perfil que utiliza características gerais independentes da lexicalização.

As técnicas utilizadas para desambiguar os nomes das instituições, assim como no caso dos nomes de autores, utilizam a tecnologia e o conhecimento que provém de áreas como a

Ciência da Computação, exigindo conhecimentos técnicos profundos para sua aplicabilidade.

Os métodos manuais ou automáticos desenvolvidos e utilizados para desambiguar as autorias em diferentes sistemas ou até mesmo nos catálogos de bibliotecas, além de serem essenciais na recuperação da informação bibliográfica por parte de seus potenciais usuários, são necessários o investimento e a aplicabilidade desses esforços para otimizar os processos avaliativos que são submetidos pesquisadores e instituições na comunidade científica. As avaliações, sobretudo em sua perspectiva quantitativa, implicam na recuperação da informação bibliográfica por seus nomes em diferentes fontes de informação científica e, se há problemas na recuperação, há impacto direto nos resultados da avaliação.

3.4 Impacto nos indicadores métricos científicos

A comunidade científica está envolta em processos de avaliação, incluindo as pesquisas, os pesquisadores e as instituições. A avaliação da ciência é uma temática que suscita diversas discussões, sobretudo em relação ao uso adequado e responsável das métricas, que costumam ser mais quantitativas do que qualitativas, o que gera uma reivindicação pertinente e que deve ser considerada pelos avaliadores, sejam eles pessoas ou instituições. Mesmo em meio à discussões, que precisam acontecer, a avaliação é intrínseca à atividade científica e em muitos desses processos são utilizadas as fontes de informação bibliográfica para coletar parte do insumo que será avaliado, para assim definir diversas questões de alto impacto, como a concessão de bolsas por agências de fomento e a contratação ou demissão de pesquisadores. E essa recuperação está diretamente relacionada aos nomes dos autores e das instituições e o impacto de possíveis irregularidades.

Destaca-se que o processo de avaliação da pesquisa científica deve ser realizado de acordo com diferentes níveis de medição do impacto, considerando que existem muitas métricas sendo utilizadas fora de contexto.

[...] é importante entender que o impacto pode – e deve – ser medido em diferentes níveis. Podemos falar sobre o impacto de um artigo, que é diferente do impacto do periódico onde esse artigo foi publicado, assim como do autor que o escreveu, ou da instituição a que pertence esse autor, pois cada uma dessas instâncias possui suas próprias medidas de impacto independentes (NASCIMENTO, 2017, p. 28, grifos da autora).

Para autores e instituições, dentre outras, são utilizadas métricas que envolvem

diretamente a quantidade de publicações e suas citações, geralmente para medir a produtividade e o alcance da pesquisa. Contudo, Haak, Meadows e Brown (2018) criticam veementemente a avaliação da pesquisa voltada exclusivamente para as publicações, reforçando que a pesquisa científica é e deve ser mais do que publicações e periódicos, enfatizando na necessidade de expandir o pensamento a respeito do tema também para os próprios pesquisadores e suas instituições.

Strotmann e Zhao (2012) afirmam que os estudos baseados em autoria formam um dos mais importantes aspectos da bibliometria, embora sua maior desvantagem seja a ambiguidade, um clássico problema da área de Ciência da Informação. Sendo assim, as avaliações da pesquisa científica por autores ou instituições precisa, necessariamente, recuperar suas produções e a forma mais utilizada para tal seria por seu nome. Dessa forma, se alinham essas necessidades de busca para sanar a necessidade informacional dos processos avaliativos em ciência com a discussão a respeito da recuperação eficaz da informação, considerando as ambiguidades de nomes que podem ser encontradas nas fontes de informação bibliográfica.

Smalheiser e Torvik (2009) reforçam que a desambiguação dos nomes de autores é útil para sanar questões que surgem cotidianamente na comunidade científica como a busca por parcerias ou colaborações com outros pesquisadores da mesma área de atuação, possibilitando inclusive a coleta muito além das publicações, mas também de materiais não publicados e que podem ser fundamentais para o desenvolvimento de novas pesquisas como ideias ou dados de pesquisa não armazenados em repositórios, e a avaliação do perfil de pesquisadores por parte de agências de fomento, incluindo a verificação de suas publicações, citações e demais informações a serem julgadas durante o processo de concessão das bolsas.

Ainda nesse sentido, Boudry e Durand-Barthez (2020) reiteram que a identificação correta dos pesquisadores e de sua produção científica beneficia a todas as partes envolvidas na pesquisa. Contudo, encontrar toda a produção científica de um pesquisador ou de uma instituição no contexto atual, de proliferação de publicações em periódicos online, é como procurar uma agulha no palheiro (JINHA, 2010).

No entanto, embora o problema de ambiguidade do nome do autor seja bem conhecido e vários métodos para AND tenham sido propostos e explorados, [...] a prática comum nos estudos bibliométricos baseados no autor têm sido o agrupamento de todos os autores com o mesmo sobrenome e a primeira inicial, trocando a alta recuperação por uma precisão reduzida ao identificar as obras do autor para análise. Esse uso ainda predominante da abordagem simplificada de AND provavelmente decorre em grande parte da maneira como os nomes dos autores há muito são referenciados em publicações periódicas e indexados

em bancos de dados bibliográficos, especialmente para referências citadas (STROTMANN; ZHAO, 2012, p. 1822, tradução nossa).

O papel das bases de dados bibliográficas em ambiente digital possui destaque devido ao seu potencial de reunir por meio de suas plataformas, grandes quantidades de publicações científicas que serão utilizadas como insumo para o desenvolvimento científico nas próximas décadas. Dessa forma, a qualidade das bases de dados é um fator primordial para garantir a eficiência do processo de comunicação científica e seus erros podem afetar diretamente a identificação, seleção, extração, classificação, ordenação e tabulação de dados, que podem ser utilizados para realização de estudos ou avaliações da ciência (RUIZ-PÉREZ; LÓPEZ-CÓZAR; JIMÉNEZ-CONTRERAS, 2002).

Nesse sentido, a recuperação eficaz dos nomes dos autores e instituições impacta diretamente no processo de avaliação da ciência, que utiliza indicadores cientométricos, bibliométricos e mais recentemente os altmétricos. No presente trabalho serão utilizadas definições gerais a respeito dos conceitos acima mencionados, embora sejam áreas de estudo com ricas definições e discussões a respeito, foram julgadas como adequadas ao contexto da pesquisa, considerando que o seu foco não está nos indicadores, mas sim o impacto exercido neles pelos nomes de autores e instituições. A bibliometria se atenta à análise das publicações, enquanto a cientometria analisa o processo de pesquisa de uma forma mais ampla.

A cientometria trata da medida quantitativa do conjunto das atividades científicas, válida para todas as disciplinas indistintamente. Seus dados incluem os montantes investidos na pesquisa-desenvolvimento, a formação do profissional de ciência e a produção de artigos e certificados. A bibliometria é um subconjunto da cientometria e limita-se à análise das publicações e de suas propriedades (GINGRAS, 2016, p. 23, grifos do autor).

Embora as instituições e os pesquisadores possuam seu conjunto de publicações científicas, elas comumente não estão reunidas e nem podem ser consultadas por uma única fonte de informação ou base de dados bibliográfica, o que acontece devido aos critérios de indexação de periódicos, que costumam ser mais ou menos rígidos a depender da fonte em questão. Geralmente, processos avaliativos são realizados com base na produção bibliográfica coletada de bases de dados referenciais, ou seja, fontes de informação que não possuem texto completo, embora elas nem sempre reflitam a totalidade da produção dos autores a depender da área de conhecimento por elas majoritariamente cobertas, como a Scopus ou Web of Science, ou seja, essas fontes são utilizadas para consulta de indicadores cientométricos e bibliométricos,

relativos às publicações, citações, fomentos recebidos, índice de colaborações etc.

Gasparyan et al. (2018) reforçam que Scopus e Web of Science são atualmente as bases de dados bibliográficas de maior prestígio para avaliação da pesquisa em diferentes áreas. Dessa forma, as informações contidas nessas fontes impactam diretamente na elaboração dos indicadores cientométricos e bibliométricos, independentemente da qualidade dos metadados por elas disponibilizados. Nesse sentido, embora as bases de dados recebam e disponibilizem em sua interface os metadados das publicações científicas baseadas no que é disponibilizado pelas revistas, a discussão também deve abarcar a eficiência das ferramentas por elas utilizadas para otimizar a qualidade de suas informações, tendo em vista sua utilização estratégica nos processos avaliativos institucionais.

Como anteriormente também discutido, “vários erros com nomes de autores podem ser rastreados através dos artigos acadêmicos publicados” (GASPARYAN et al., 2016, p. 171). Em um estudo sobre variações do nome de autores espanhóis em bases de dados, os autores concluíram que a produtividade está inversamente relacionada à recuperabilidade, ou seja, quando aumenta o número de variações de um nome, o número de publicações recuperadas por cada variante diminui (RUIZ-PÉREZ; LÓPEZ-CÓZAR; JIMÉNEZ-CONTRERAS, 2002).

Diante disso, Ruiz-Pérez, López-Cózar e Jiménez-Contreras (2002) afirmam que para evitar o problema desde sua origem, é preciso investir em recomendações aos autores e periódicos, para que a correção não seja realizada apenas na fase final do processo ou após a publicação; encorajando também os autores para que assinem seus nomes sempre da mesma forma ao longo de sua trajetória acadêmica.

Conforme detalhado pelo relato do pesquisador indiano Kurién (2008), suas publicações indexadas no PubMed apresentaram variações incorretas de seu nome e todas receberam citações, porém ao autor mudar seu nome para o padrão ocidental, essas publicações não apareceram mais atreladas ao seu nome nas bases de dados PubMed e Web of Science, ocasionando uma queda de 25% em seu *h-index*, métrica utilizada para medir a produtividade e alcance da produção científica. Ainda para o autor, pesquisadores e editoras devem ser igualmente responsabilizados pelos erros e inconsistências em registros bibliográficos e consequentemente no perfil dos autores e em *ranking* de citações (KURIÉN, 2008).

De acordo com Sugimoto e Larivière (2018, p. 50, tradução nossa), “a autoria é o coração dos indicadores bibliométricos: são autores cujos nomes e instituições aparecem em um documento, sendo mensurados”. Assim como discutido em relação à desambiguação de nomes de autores realizada por técnicas e métodos que envolvem a tecnologia, é importante

refletir que para a realização de indicadores utilizados em avaliações científicas, deve-se considerar também a necessidade de conseguir informações adicionais a respeito do autor, para além da publicação.

Um dos usos da pesquisa bibliométrica é extrair informações que não são apresentadas explicitamente nos bancos de dados tradicionais, como inferir padrões de novas tecnologias, a extensão da novidade e originalidade nas patentes, ou identificando o gênero do autor, idade acadêmica ou migração de trajetórias das publicações de pesquisa. A exposição de informações implícitas pode aumentar o valor dos dados e seu poder explicativo. **O nome do autor representa um item de dados a partir do qual mais informações podem ser deduzidas** (KARAULOVA; GÖK; SHAPIRA, 2019, p. 488, grifo nosso, tradução nossa).

Por sua vez, a altmetria, ou métricas alternativas, que também é utilizada nos processos avaliativos em ciência, pode ser definida como “indicadores da comunicação científica que ajudam a entender como os resultados de pesquisa são vistos e usados em ambientes *online*, complementando a análise tradicional baseada na contagem de citações” (NASCIMENTO, 2017, p. 56). Contudo, atualmente, a avaliação da pesquisa científica por indicadores altmétricos acontece no nível da publicação e essa métrica pode ser gerada apenas para o material e não para o autor ou para instituições. E para recuperar as publicações a serem medidas pela altmetria, as ferramentas utilizam o identificador persistente da publicação em vez de seu título, como será visto no próximo capítulo desse trabalho.

Contudo, altmetria ainda é pouco utilizada por diversos países, como no Brasil, não sendo compreendida em sua totalidade pelos pesquisadores. A pouca compreensão atrelada à ausência de interação e perfis nas redes acadêmica, faz com que o indicador altmétrico não seja cogitado frequentemente para integrar avaliações de pesquisa científica. Como as menções na *web* precisam ser feitas sempre por meio do identificador persistente, nem sempre este padrão é conhecido e muitas menções deixam de ser contabilizadas pelas ferramentas agregadoras deste tipo de métrica, como o Altmetric e PlumX.

[...] a falta de padronização e de qualidade dos metadados das publicações nacionais impede o correto reconhecimento de itens individuais da produção acadêmica e o rastreamento das citações recebidas *online*. A baixa participação da comunidade científica brasileira na *web* e em ambientes *online* é um outro fator que inibe a adoção de métricas alternativas (SILVA; GOUVEIA, 2020, p. 15, grifos do autor).

Nesse sentido, fontes de informação, grupos editoriais ou independentes, mas de alguma

forma ligados à comunidade científica, e pesquisadores criaram espaços digitais para que se pudesse acrescentar informações a respeito dos pesquisadores e de sua pesquisa para além das publicações, seja em forma de seus metadados ou texto completo. Nesses ambientes, é possível vislumbrar uma interação entre a comunidade acadêmica e, em algumas delas, reunir toda a trajetória acadêmica de um pesquisador, independentemente de sua contribuição e das fontes nas quais elas foram feitas, sem restrições.

3.5 Redes acadêmicas em ambiente digital: espaço para manutenção da visibilidade

O ambiente digital é atualmente intrínseco às atividades desenvolvidas pelos pesquisadores na comunidade científica, tendo se consolidado devido a sua ampla capacidade de disseminação e alcance da informação. Neste contexto, são cada vez mais frequentes as redes acadêmicas que possuem a proposta de reunir produtos da trajetória científica dos pesquisadores além de possibilitar a troca de experiências entre os pares. As redes voltadas para pesquisadores têm se consolidado com um espaço que possibilita a manutenção de sua visibilidade enquanto pesquisadores e conseqüentemente a visibilidade de suas pesquisas, atuando como um facilitador e promotor de sua divulgação.

A produção científica tradicionalmente disponível por meio de bases de dados científicas em ambiente digital, ainda impõe algumas barreiras sobretudo se for considerado que parte delas possuem acesso restrito, possível apenas mediante assinaturas. Existem também os repositórios digitais, que muitas vezes são mantidos pelas instituições, mas se pouco divulgados não conseguem atingir seu público. Conforme afirma Boudry e Durand-Barthez (2020), essas ferramentas não supriram a falta de um espaço para comunicação direta entre pesquisadores, o que era possível, por exemplo, apenas por meio da disponibilidade do *e-mail* de contato do autor correspondente.

A entrada em cena das tecnologias de informação e comunicação (TIC) no final do século XX e início do século XXI, especialmente da *Internet* e da *Web*, vem produzindo sensíveis alterações nos processos tradicionais de comunicação científica, **alterando padrões e comportamentos**, introduzindo uma série de **mudanças e abordagens**, possibilitando novas formas de produção, circulação, disseminação, recuperação e **uso da informação** [...] (PRÍNCIPE, 2013, p. 197, grifos nossos).

Neste contexto, por volta de metade dos anos 2000, surgem as primeiras *Academic Social Networks* (ASN), redes sociais acadêmicas ou redes acadêmicas, como serão chamadas

no presente trabalho, que através de um cadastro rápido de informações pessoais e acadêmicas, permite o contato direto entre os pesquisadores com perfil ativo, via plataformas, recebimento de alertas quando novas publicações de interesse estiverem disponíveis, realização de conexões e colaborações entre pares, além de responder questionamentos de colegas e também os realizar via redes (BOUDRY; DURAND-BARTHEZ, 2020; MIKKI et al., 2015).

As redes acadêmicas se consolidaram como um espaço dinâmico e interativo que oferece aos pesquisadores a oportunidade de uma troca mais próxima com colegas da mesma área de atuação e acesso também às publicações científicas que colegas disponibilizam via plataformas, respeitando os direitos de acesso de acordo com o periódico científico. Trata-se de uma forma alternativa à busca por artigos científicos via bases de dados, que em linhas gerais demandam a construção de uma estratégia de busca para sua adequada recuperação. Por meio das redes acadêmicas, baseadas nos interesses e pessoas que os pesquisadores seguem, o acesso a estes conteúdos fica facilmente disponível, lembrando o *feed* de redes sociais não acadêmicas, no qual sempre é exibido os conteúdos que mais podem interessar ao usuário.

[...] um sistema de nodos e elos; uma estrutura sem fronteiras; uma comunidade não geográfica; um sistema de apoio ou um sistema físico que se pareça com uma árvore ou uma rede. A **rede social**, derivando deste conceito, passa a representar um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados. (MARTELETO, 2001, p. 72, grifo nosso).

O indicador de participação e interação dos pesquisadores em ambientes *online* pode ser medido pelos indicadores alométricos, como anteriormente mencionado, evidenciando que embora a informação esteja disposta em novas plataformas que não são as tradicionais bases de dados bibliográficas, seu alcance e visibilidade também podem ser mensurados pelas chamadas métricas alternativas.

Contudo, em relação à recuperação correta de pesquisadores por seus nomes, as redes acadêmicas não possuem recursos sofisticados para otimizar essa recuperação, ao menos não até o presente momento. Mas considerando a fluidez e dinamicidade no desenvolvimento do ambiente digital de uma forma ampla e de suas plataformas, é possível que algo ainda seja realizado proximamente com este objetivo.

ASNs [...] não oferecem, no entanto, ferramentas para recuperar pesquisadores de forma inequívoca. Gratuitos, mas “estanques”, esses sites têm a comum característica de uma operação documental extremamente básica. Ou seja, a chave de pesquisa é o nome do autor, o que obviamente

condiciona a exibição da resposta menos aproximada possível. No entanto, esta chave é limitada a cadeias de caracteres brutos nessas redes, deixando a porta aberta para atomização excessiva de identidades, muitas vezes limitada devido à falta de respostas relevantes [...] (BOUDRY; DURAND-BARTHEZ, 2020, p. 2, tradução nossa).

Podem ser mencionadas como exemplos de redes sociais acadêmicas, as plataformas *ResearchGate* e *Academia.edu*, que possuem certo nível de interação entre os pesquisadores e possibilitam o acesso ao perfil e publicações científicas de pares.

Embora não sejam plataformas que obtenham um nível de interação e comunicação facilitada como as anteriormente mencionadas, algumas plataformas como Google Scholar, *ResearcherID* e ORCID têm sido cada vez mais notadas entre os pesquisadores e utilizadas em estudos (BOUDRY; DURAND-BARTHEZ, 2020; MIKKI et al., 2015; GASPARYAN et al., 2017) para verificar a presença de pesquisadores de instituições e compreender como se dá essa atuação.

Nenhuma das redes será detalhada, considerando que não é este o objetivo do trabalho, mas é importante mencionar que muitos pesquisadores se atentam para a presença de um perfil no Google Scholar considerando a visibilidade conferida por um buscador acadêmico, além de, como mencionam Gasparyan et al. (2017), muitos pesquisadores que não possuem publicações indexadas e conseqüentemente visíveis em outras prestigiadas bases de dados bibliográficas, encontram no Google Scholar uma maneira de também se posicionarem em ambiente digital. Quanto a presença em ambientes como o *ResearcherID* e ORCID, que são considerados identificadores persistentes, maiores detalhes serão discutidos a respeito no próximo capítulo.

No Brasil, embora não seja uma rede que possibilite uma interação direta via plataforma, não possua a função de identificador persistente ou forneça um indicador próprio de citações, há o Currículo Lattes, que faz parte da Plataforma Lattes mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O Currículo Lattes é, como o próprio nome sugere, um espaço no qual os pesquisadores brasileiros podem inserir suas informações acadêmicas e profissionais e serem localizados nessa plataforma *online*, que é considerada uma referência nacional e exigida durante os processos seletivos para pesquisadores, submissão de projetos de pesquisa ou concessão de fomento.

A presença digital do pesquisador, embora já seja uma realidade, se trata de uma tarefa árdua e que demanda alto investimento de tempo, sobretudo considerando a existência de diferentes plataformas que podem ou não ser interoperáveis. Essa interoperabilidade atuaria no sentido de informações serem intercambiáveis entre elas e evitarem que o preenchimento de

informações tenha que ser feito manualmente e individualmente em cada uma delas.

Jamali, Nicholas e Herman (2016) realizaram um estudo com o objetivo de entender como se dá o conceito de reputação acadêmica no ambiente digital e as práticas de pesquisadores europeus diante de plataformas emergentes, como as redes sociais acadêmicas. O estudo mostrou que a disseminação da pesquisa via *blogs* e Twitter ou via redes acadêmicas obteve o menor nível de importância em relação sua influência na reputação acadêmica, contudo isso provavelmente se deve a um aspecto que deve ser considerado, conforme os autores pontuam, que seria idade dos pesquisadores, considerando que poucos tinham menos de 30 anos.

Dessa forma, é possível observar que a idade dos pesquisadores e sua familiaridade e uso de ambientes digitais pode influenciar no uso de redes sociais acadêmicas e consequentemente com a disseminação da pesquisa científica na *web*.

Foi amplamente considerado que as plataformas de reputação terão um impacto duradouro e merecem atenção, especialmente por jovens acadêmicos que se beneficiam mais do que acadêmicos mais velhos com as ferramentas de reputação *online* (JAMALI; NICHOLAS; HERMAN, 2016, p. 47, tradução nossa).

Um estudo realizado por Mikki et al. (2015) para verificar a presença digital de pesquisadores da Universidade de Bergen (Noruega) nas plataformas *ResearchGate*, *Academia.edu*, *Google Scholar Citations*, *ResearcherID* e *ORCID*, concluiu que 77% dos pesquisadores não possuíam perfil nas cinco redes, apenas 17% tinham perfil em duas redes, 5% tinham perfil em três redes, 0,5% tinham perfil em quatro redes e apenas um pesquisador tinha perfil em todas as redes analisadas na pesquisa.

Ortega (2015) em seu estudo com perfis de autores do *Spanish National Research Council*, também concluiu que apenas 72% dos pesquisadores mantinham presença em apenas uma das redes.

De acordo com o povoamento desses sites acadêmicos, os resultados mostram que a maioria dos pesquisadores está incluída apenas em um sítio acadêmico, confirmando um número limitado de perfis sobrepostos. Esse comportamento pode ser motivado pelo esforço e manutenção que exigem (ORTEGA, 2015, p. 47, tradução nossa).

Embora os estudos tenham sido realizados em 2015 e 2016, é preciso levar em conta a dinamicidade das redes: certamente tais redes foram otimizadas e provavelmente o resultado

dessas pesquisas atualmente seria outro, é necessário refletir a respeito da quantidade de ambientes em que os pesquisadores têm a possibilidade de marcar presença em ambiente *online*, mediante o alto custo de manter diferentes perfis ou plataformas atualizados com suas informações acadêmicas e profissionais. O investimento de tempo do pesquisador na pesquisa se vê reduzido diante da necessidade de divulgá-la e posicioná-la em tais plataformas.

Ao mesmo tempo, como se trata de ambientes que proporcionam essa interação e visibilidade da pesquisa e do pesquisador, embora não forneçam rebuscadas ferramentas de busca, com o objetivo de identificar mais facilmente os pesquisadores, o fato de disponibilizar as publicações nas plataformas ligadas ao perfil, pode ampliar a precisão na identificação da autoria, considerando que ela não tenha ficado clara quando recuperada, por exemplo, por uma base de dados bibliográfica.

Todas as discussões a respeito de atribuição autoral, uso adequado de normas e padrões de registro de nomes de autores, as alternativas que a tecnologia abarca para desambiguar homônimos, o impacto em indicadores e a presença nas redes acadêmicas suscita discussões a respeito dos desafios que a comunidade científica enfrenta em relação à autenticidade de conteúdos e identificação correta de seus autores.

São iniciativas que, ao seu modo e com seus objetivos específicos, buscam de alguma maneira ligar pesquisas aos seus pesquisadores, conferindo a elas a adequada visibilidade. Contudo, essas iniciativas com tal objetivo podem também auxiliar no processo de identificação de falhas na atribuição de autorias de publicações científicas, ou ao menos uma otimização desse processo, que também podem impactar na recuperação informacional.

3.6 Autenticidade e autoridade de conteúdos em ambiente digital: desafios da comunidade científica

Atribuir corretamente as autorias de conteúdos em ambientes digitais se faz um desafio, sobretudo ao considerar a velocidade de disseminação de conteúdos por diferentes plataformas que, na maioria dos casos, não possuem uma estrutura adequada de identificação e rastreamento de seus criadores e de sua autenticidade, ocasionando o compartilhamento desenfreado de conteúdos.

A informação científica, de igual maneira, é compartilhada em larga escala em ambiente digital, entre os membros da academia, mas também pode ser acessada e compartilhada pela sociedade de forma geral, por meio de iniciativas de divulgação científica.

Contudo, o compartilhamento desenfreado de conteúdos sem autenticidade comprovada, sejam eles de cunho científicos ou não, acabam contribuindo com as chamadas *fake news*, conferindo-as ainda mais força e poder na sua disseminação. Essas notícias falsas ou conteúdos de idoneidade suspeita, conseguem disseminar calúnias, difamações e informações erradas que contribuem para a desinformação da sociedade em grande escala. Embora essas notícias falsas e a desinformação sejam algo comum na sociedade mesmo antes do contexto digital, o termo ganhou visibilidade em 2016, quando foi mencionado pelo então candidato à presidência dos Estados Unidos da América, Donald Trump, ao chamar de *fake news* notícias as quais não aprovava. Jardim e Zaidan (2018, p. 3) conceituam: “designam-se como *fake news* notícias e informações falsas que reproduzem o padrão convencional de notícias verdadeiras, porém que ausentam de dados editoriais e normas que validam sua veracidade e credibilidade”.

A informação científica também é vítima das *fake news* e muitos conteúdos são compartilhados em redes sociais como se tivessem validade científica ou como se fossem ditos por supostos pesquisadores vinculados a universidades ou instituições renomadas de pesquisa, nacionais ou internacionais. De um modo geral, a sociedade pode não ter elementos ou conhecer caminhos para atestar se a informação é falsa ou não e se o suposto pesquisador realmente o é e se tem condições e conhecimento para dissertar sobre o tema da notícia.

Neste ponto, é importante discutir a respeito da garantia de autenticidade de conteúdos científicos em meio digital, de forma a oferecer condições para que a população e também para a comunidade científica consigam se certificar a respeito do pesquisador responsável pela informação e seu histórico de atuação e pesquisa.

Não apenas em relação a falsa informação científica disseminada, mas a informação científica idônea também encontra dificuldades de atestar sua autenticidade e sua autoria, no nível de seus pesquisadores e instituições.

Na literatura científica, é possível encontrar diversos pesquisadores discutindo a respeito da ética na atribuição de autorias nas publicações científicas e às diretrizes dos periódicos científicos a respeito da temática. Em linhas gerais, as discussões dão ênfase para aos papéis dos autores em cada uma das publicações e a forma com a qual é decidido incluir um pesquisador como de fato autor de um artigo e suas respectivas posições, o que pode variar de acordo com a área do conhecimento em questão.

Nesse sentido, Sugimoto e Larivière (2018, p. 50-51, tradução nossa) apontam que: “As práticas de atribuição de autoria variam por disciplina [...] diferenças na atribuição de autorias

são vistas não apenas no nível de disciplina, mas também atravessam especialidades e continentes”. Um estudo analisou 600 periódicos indexados no Journal Citation Reports (JCR) e 62% possuíam uma diretriz voltada para orientação aos autores a respeito da atribuição autoral e outras condutas éticas relacionadas ao assunto (RESNIK et al., 2016).

Schroter et al. (2020) realizaram uma pesquisa com autores correspondentes da coleção de periódicos científicos *BMJ Journals* no período de 2014 a 2018 e comentam:

Compreender a prática de autoria é uma importante questão ética, porque a autoria apropriada garante crédito e responsabilidade pela pesquisa. A prática ética da autoria é essencial para a promoção e manutenção da integridade científica da pesquisa biomédica. Nós descobrimos que as diretrizes e critérios de autoria são conhecidos pela maioria dos pesquisadores e sua aplicação é considerado benéfica ao preparar manuscritos. No entanto, a má conduta de autoria ainda prevalece; até aqueles que são novos na pesquisa relataram experiências com ela. Assim, não é simplesmente uma questão de os autores precisarem ser informados sobre diretrizes e critérios, mas de ter a oportunidade de aplicá-los em um ambiente de apoio que é adequado para sua disciplina [...] apesar de um alto nível de conhecimento das diretrizes e critérios de autoria, eles não são tão amplamente usados. Mas o incentivo explícito por parte das instituições para discutir a autoria no início e com frequência pode resultar em decisões que são percebidas como mais justas (SCHROTER et al., 2020, p. 6-7, tradução nossa).

A comunidade científica precisa compreender o papel ético das diretrizes para autores, com o objetivo de atribuir de maneira precisa as autorias às suas publicações científicas. Considerando o movimento e as práticas de ciência aberta, recentemente tem sido adotada em diversos periódicos científicos a taxonomia *Contributor Roles Taxonomy* (CRediT), que descreve as atribuições, papéis e responsabilidades dos autores em cada publicação científica. A taxonomia inclui catorze itens: conceituação; metodologia; *software*; validação; análise formal; investigação; recursos; curadoria de dados; escrita – primeira redação; escrita – revisão e edição; visualização; supervisão; administração do projeto e obtenção de financiamento. A utilização da CRediT aliada ao uso de identificadores persistentes, que serão detalhados adiante, pode otimizar a identificação da contribuição dos autores na publicação, com a possibilidade de agregar indicadores qualitativos à avaliação da pesquisa científica.

As agências de fomento, a exemplo dos National Institutes of Health dos Estados Unidos (NIH), e da Comissão Europeia, têm muito a se beneficiar com o uso de identificadores persistentes de autoria, como o ORCID, para reduzir a entrada de dados e facilitar o rastreamento e recuperação do impacto da pesquisa. Se a estes identificadores estiver vinculada a contribuição dos autores através da taxonomia CRediT, as agências poderão com maior

eficiência obter dados sobre a verdadeira contribuição de um pesquisador que submete um auxílio à pesquisa, sem ter que recorrer a outras métricas de impacto (NASSI-CALÒ, 2018).

Mas não somente em relação aos papéis e responsabilidades dos autores na publicação, é vantajoso refletir sobre a oportunidade que essas políticas possuem de orientar também a respeito da forma correta de inserir os nomes dos autores na publicação, incluindo suas formas de abreviação. Essa conduta por parte das editoras ou periódicos científicos poderia auxiliar no processo de padronização dessas autorias assim que esses artigos fossem disponibilizados para consulta acadêmica em ambiente digital, via bases de dados bibliográficas.

Além da importância da orientação a respeito da disposição dos nomes dos autores, diretrizes para orientação a respeito das instituições a serem inseridas como afiliações autorais também precisam obedecer critérios que, se não forem previamente estabelecidos pelas universidades ou instituições de pesquisa por meio da publicação e divulgação das chamadas nominais institucionais, que podem conter a afiliação padronizada do nome da instituição para serem utilizadas em publicações científicas, impactam diretamente na recuperação da produção científica das instituições.

Nos últimos anos, diversos estudos têm se atentado para o aumento no número de afiliações dos autores nas publicações científicas, também chamado como múltiplas afiliações. Essa pode ser uma consequência na máxima *publish or perish* a qual pesquisadores e instituições muitas vezes são submetidos, ou seja, um sistema de avaliação focado em produtividade. Um espaço que deveria ser utilizado para inserir a afiliação institucional do autor considerando sua real contribuição na pesquisa realizada, muitas vezes é utilizado como uma espécie de currículo, onde alguns autores inserem todas as suas afiliações, considerando que possui vínculo com mais de uma instituição, ainda que sua participação na pesquisa não tenha envolvido diretamente sua atuação em todas elas. Por exemplo, se um autor que está vinculado às instituições x, y e z publica um artigo, ele pode colocar todas essas afiliações vinculadas a seu nome, mas na verdade este artigo é resultado de sua pesquisa de doutorado a qual realiza na instituição y, então esta publicação deveria conter apenas a afiliação y.

Um recente estudo de Hottenrott, Rose e Lawson (2021, p. 1, tradução nossa) sobre múltiplas afiliações, afirma que: “[...] a afiliação institucional afeta o valor atribuído aos pesquisadores individuais por meio do prestígio institucional, com consequências para a pesquisa e as trajetórias de carreira”. Entende-se que muitos pesquisadores podem inserir múltiplas afiliações com o objetivo de se agregar valor à sua trajetória e à própria publicação,

ainda que esta não tenha relação direta com a instituição inserida na afiliação.

Bachelet et al. (2019) publicaram um estudo que analisou publicações de 2016 indexadas na base de dados Scopus, buscando por todos os autores que apresentavam múltiplas afiliações e pelo menos uma delas era de universidades do Chile, com o objetivo de identificar se as afiliações poderiam ser validadas ou se evidenciavam a inserção de informações falsas. O estudo concluiu que 38% dos autores não puderam ter seu vínculo institucional validado, representando cerca de 40% das publicações, sendo que a utilização do identificador persistente ORCID não foi útil para localizar evidências de afiliação pois apenas 14% deles tinham perfil ativo; ou seja, a afiliação não validada pode indicar má conduta que impacta diretamente em diversos atores da comunidade científica como universidades ou *rankings*, considerando a contagem inflacionada de publicações e reduzindo a real contribuição de instituições em artigos científicos.

A preocupação com a descrição correta e dentro dos padrões que possibilitem a recuperação da produção científica institucional deve acontecer dada sua alta relevância. Contudo, é indispensável que discussões a antecedam e prevejam tópicos como as múltiplas afiliações, ou seja, antes de saber como inserir uma afiliação no nível descritivo, o pesquisador necessita entender quais afiliações devem ou não ser de fato inseridas na submissão de um artigo científico.

Retomando os impactos da informação em seu caráter descritivo, Gasparyan et al. (2016) listam uma série de opções para auxiliar no processo de identificação dos autores de conteúdos científicos, dentre elas estão estratégias como a criação de perfis em redes sociais e acadêmicas, conforme abordado anteriormente, além da fundamental consideração a respeito da vinculação institucional e a escrita dos nomes com precisão (Quadro 2). Todas as estratégias mencionadas são válidas e devem ser de alguma forma consideradas pelos autores, com o objetivo de identificá-los corretamente e conseqüentemente gerar indicadores científicos de qualidade, considerando o impacto que a ambigüidade ou falta de precisão nos nomes de autores causam na recuperação informacional.

Quadro 2 – Opções para identificar e avaliar autores científicos

Escrito por precisão / abreviatura primeiro, nome do meio e sobrenome
Afiliações vinculadas
Contatos postais e de <i>e-mail</i>
Assinatura

Fotos em artigos ou em plataformas institucionais ou individuais
Notas biográficas em portais acadêmicos, institucionais ou sociais, incluindo páginas na Wikipedia
Vídeos com resumos de artigos
<i>Open Researcher and Contributor ID (ORCID)</i>
Registros acadêmicos em bases de dados bibliográficas ou plataformas (PubMed, Scopus, Web of Science e Google Scholar)
Perfil em redes acadêmicas e corporativas (ResearchGate e LinkedIn)
Perfil em redes sociais (Facebook, Twitter)
Estilo de escrita acadêmica

Fonte: adaptado de GASPARYAN et al., 2016, p. 172, tradução nossa.

A correta identificação de pesquisadores e de suas pesquisas possui impacto direto na autenticidade e autoridade de conteúdos científicos em ambiente digital, perpassando por diversas questões que remetem ao estabelecimento de diretrizes por parte das editoras e periódicos, além de outras iniciativas que prevejam a orientação de pesquisadores para condutas apropriadas, não somente em relação à forma com a qual informam seus nomes e de suas afiliações, mas também em relação à ética atrelada ao processo.

Sendo assim, surgem diferentes estratégias na comunidade científica que possuem o objetivo de facilitar esse processo de identificação de autores e suas instituições, contornando a falta de padronização requerida em seus nomes por meio da atribuição de códigos alfanuméricos únicos, que por sua natureza, tornam-se inequívocos em sua representação.

4 IDENTIFICADORES PERSISTENTES

Garantir a recuperação da informação científica implica diretamente no desenvolvimento científico, considerando que se trata de seu principal insumo. Como visto, a recuperação da informação científica por linguagem natural prejudica a relevância e pertinência do conteúdo entregue como resultado de busca ao pesquisador nas fontes de informação. A estratégia desenvolvida e aperfeiçoada ao longo dos anos esteve na construção de linguagens controladas para indexação e no controle de vocabulário, utilizados para recuperação por temáticas ou por autores e instituições. Contudo, o ambiente digital potencializou a quantidade de informação disponibilizada e diversos atores foram envolvidos no processo de produção, armazenamento e disponibilização de conteúdos científicos, tornando menos eficiente o processo de controle de entradas, ou seja, a forma com a qual esses materiais eram descritos.

O processamento automático dos nomes de autores, por exemplo, baseado exclusivamente na forma em que o nome se encontra no documento é um problema desafiador, sobretudo ao considerar que cada editora ou periódico científico possui seus próprios padrões e processo de edição específicos; ressaltando ainda mais a necessidade de a comunidade científica estabelecer e adotar padrões a serem seguidos nesse sentido (KENDALL; YEE; HARDLY, 2017).

O controle de vocabulário e de autoridade continua sendo um trabalho fundamental para registrar esses conteúdos de forma padronizada também em ambiente digital, mas uma alternativa ainda mais inequívoca de recuperação da informação tem sido discutida na comunidade científica na última década, embora o conceito em si não seja tão recente assim: o uso de identificadores.

Com a produção científica global crescendo rapidamente, está se tornando cada vez mais difícil distinguir os autores. Um sistema de numeração universal poderia ajudar os cientistas a se manterem atualizados na literatura, ajudar as universidades a rastrear mais prontamente a produtividade do pessoal e permitir que as agências de financiamento monitorem melhor o retorno que estão obtendo com seus investimentos. Um número de identificação eficaz também pode facilitar a localização de informações sobre as afiliações, colaboradores, interesses de um autor ou simplesmente seu paradeiro atual (ENSERINK, 2009, p. 1662, tradução nossa).

Especificamente no contexto acadêmico-científico, a tendência pela identificação de documentos começou a ser observada na década de 1960, surgindo então os primeiros sistemas

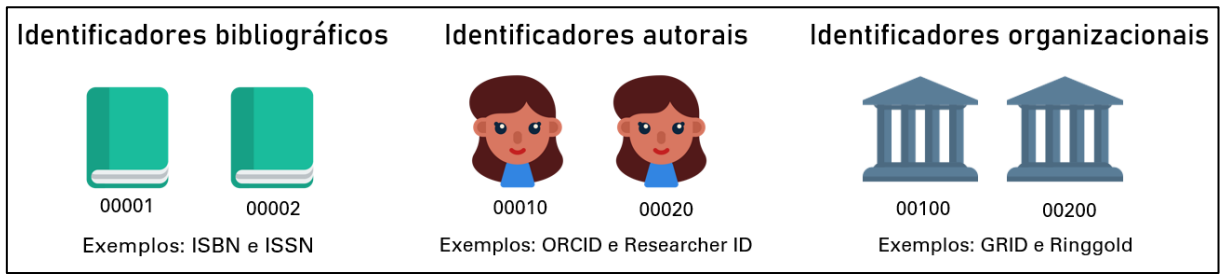
numéricos a partir das possibilidades oferecidas pelo desenvolvimento da informática (CAMPELLO, 2019). Por anos os identificadores foram desenvolvidos para identificar patentes, itens materiais (como elementos químicos), livros, periódicos, artigos e autores (BOUDRY; DURAND-BARTHEZ, 2020). É interessante pontuar que Garfield (1955) na década de 1950 já falava sobre o desenvolvimento de um código único para a ciência, considerando uma possível identificação de publicações por código.

Conforme afirma Campello (2019), é uma prática milenar da Biblioteconomia a identificação de documentos a partir da descrição bibliográfica, mas no contexto atual, com as bases de dados que contém registros com alto padrão de descrição e os sistemas de metadados utilizados para recuperação de documentos, são fundamentais os instrumentos de identificação e acesso às publicações.

É possível observar a necessidade e avanço na criação desses identificadores para recursos acadêmico-científicos, sobretudo com o desenvolvimento da *web*. Contudo, é preciso destacar que, embora o ambiente digital tenha sido propulsor do avanço dos identificadores, eles já eram utilizados em alguns materiais, como livros e periódicos, que não estavam em formato digital. Nesse sentido, é indispensável realizar uma distinção entre os conceitos de identificadores e identificadores persistentes: “um identificador é uma sequência de caracteres que identifica uma entidade. O termo ‘identificador persistente’ é comumente utilizado no contexto de objetos digitais acessíveis na internet” (MCMURRY et al., 2017, tradução nossa).

Persistent identifier (PID) ou identificador persistente, é como um nome, geralmente uma pequena *string* ou sequência de caracteres de busca, dado a alguma entidade em um contexto específico, ou seja, pode ser utilizado para nomear/identificar diferentes entidades, como números de acesso, livros ou sequências genéticas, sendo único e globalmente utilizado para recuperar a descrição de um recurso ou ele próprio (ANANTHAKRISHNAN et al, 2020).

Os identificadores persistentes, que na prática são códigos alfanuméricos com distintas e específicas configurações de acordo com seu contexto e finalidade, otimizam a recuperação dos recursos que representam. Na Figura 1 é possível observar alguns exemplos de como os identificadores persistentes podem auxiliar na desambiguação de publicações, autores ou instituições com nomes iguais ou muito semelhantes; fato que em uma busca por seus nomes poderia gerar resultados mesclados devido à essa semelhança, mas os códigos alfanuméricos, exemplificados na imagem que serão abordados mais adiante, podem diferenciá-los de forma eficaz.

Figura 1 – Exemplos de identificadores persistentes e sua utilização prática

Fonte: elaborado pela autora.

No Quadro 3, é possível observar alguns dos identificadores utilizados no contexto acadêmico-científico e que contemplam a representação de recursos de pesquisa, listados de acordo com o recurso o qual identifica e data de lançamento, ou seja, data em que foram disponibilizados para utilização, contendo também endereço *web* para acesso.

Quadro 3 – Exemplos de identificadores utilizados na comunidade científica

Nome	Ano de lançamento	Recurso que identifica	Link
Archival Resource Key Identifiers (ARKs)	2001	Objetos informacionais	https://arks.org/
ArXiv Author ID	2005	Pesquisadores	https://arxiv.org/help/author_identifiers
AuthorClaim	1999	Pesquisadores	http://authorclaim.org/
Digital Object Identifier (DOI)	1994	Publicações	https://www.doi.org/
Global Research Identifier Database (GRID)	2015	Instituições	https://www.grid.ac/
Handle	1995	Objetos informacionais	https://www.handle.net/
International Geo Sample Number (IGSN)	2007	Amostras físicas	https://www.igsn.org/
International Standard Name Identifiers (ISNI)	2012	Pesquisadores e instituições	https://isni.org/

International Standard Book Number (ISBN)	1967	Livros	https://www.isbn-international.org/
International Standard Serial Number (ISSN)	1967	Periódicos	https://www.issn.org/
Currículo Lattes	1999	Pesquisadores	https://lattes.cnpq.br/
Legal Entity Identifiers (LEI)	2012	Entidades jurídicas	https://www.lei-worldwide.com/
Open Citation Identifier (OCI)	2010	Citações bibliográficas	https://opencitations.net/oci
Open Researcher and Contributor ID (ORCID)	2012	Pesquisadores	https://orcid.org/
Persistent Identification of Instruments (PIDINST)	2020	Instrumentos	https://www.rd-alliance.org/groups/persistent-identification-instruments-wg
Persistent Uniform Resource Locator (PURL)	1995	Objetos de informação	http://www.purlz.org/
PubMed Author ID	2010	Pesquisadores	https://pubmed.ncbi.nlm.nih.gov/
PubMed Central ID	Não identificado	Publicações	https://www.ncbi.nlm.nih.gov/pmc/
PubMed ID	Não identificado	Publicações	https://pubmed.ncbi.nlm.nih.gov/
Research Activity Identifier (RAiD)	2017	Atividades de pesquisa	https://www.raid.org.au/
Research Organization Registry (ROR)	2019	Instituições	https://ror.org/
ResearcherID	2008	Pesquisadores	http://publons.com/
Ringgold IDs	2003	Instituições	https://www.ringgold.com/ringgold-identifier/

ScopusID	2006	Pesquisadores	https://www.scopus.com/
Uniform Resource Identifier (URI)	2001	Objetos informacionais	-
Uniform Resource Name (URN)	2005	Objetos informacionais	-
Virtual International Authority File (VIAF)	2003	Autores	http://viaf.org/

Fonte: elaborado pela autora.

Os identificadores possuem mais de 20 anos e continuam fundamentais para auxiliar no processo de recuperação e descoberta da literatura científica, sobretudo ao considerar estudos como os de Lawrence et al. (2001) e Dellavalle et al. (2003), que identificaram a perda de acesso às referências citadas por artigos científicos analisados quando eram informados endereços na *web*, sugerindo então a aplicação de *Uniform Resource Locators* (URLs) que contenham um identificador persistente, como o DOI, para a recuperação dos locais de armazenamento das referências citadas ao longo do tempo, sem que esse tipo de informação se torne inacessível.

Diante das transformações e avanços inerentes ao ambiente digital, considerando os conceitos de objeto digital e de recurso de informação disponível em ambiente digital, faz-se necessária uma reflexão a respeito do impacto do uso dos identificadores nesse contexto.

Em ambiente digital, quando é discutida a importância da recuperação da informação científica, inclusive por meio dos identificadores persistentes, não se trata apenas das publicações. No contexto da pesquisa científica, os identificadores persistentes podem ser utilizados para identificar recursos como as próprias publicações, *preprints*, dados de pesquisa, pesquisadores ou instituições.

Nesse mesmo sentido, Haak, Meadows e Brown (2018, p. 2, tradução nossa) reiteram que “para entender melhor a pesquisa, precisamos conhecer não somente as publicações, mas também as pessoas, organizações e recursos envolvidos. Assim como com as publicações, isso significa que precisamos estabelecer chaves únicas – identificadores – para estas entidades”.

No mesmo sentido:

A identificação única é o método solucionável por máquina na Web e demonstra um compromisso a longo prazo com a persistência que é fundamental para prover acesso a dados citados e os metadados a eles

associados. Existem vários esquemas de identificadores na Web que atendem a esses dois critérios. Os melhores identificadores para citação de dados em uma comunidade particular de prática serão aqueles que atendem a esses critérios e são nela amplamente utilizados (STARR et al., 2015, p. 22, tradução nossa).

E considerando a necessidade de recuperação dos conteúdos científicos atrelados a diferentes chaves de busca, como pessoas, organizações e recursos, foram criados ao longo dos anos diversos identificadores com características distintas e se adequam especificamente para o tipo de recurso de pesquisa ao qual foram criados.

Nesse sentido, Cousijn et al. (2021) trazem uma “escala de maturidade” para demonstrar o quão madura é a infraestrutura de identificadores ligados ao universo da pesquisa em relação às entidades representadas. Analisados no mês de maio de 2018: maduro, indicando que a infraestrutura é de uso comum da comunidade científica; emergente, indicando serviços pilotos ou que ainda não estão amplamente disponíveis para a comunidade científica; e imaturo, indicando que apenas há discussões superficiais a respeito e ainda não há consenso sobre como implementar o identificador para o recurso. O Quadro 4 possui alguns dos resultados encontrados:

Quadro 4 – Panorama do uso de identificadores para diferentes recursos de pesquisa

Recursos de pesquisa	Maturidade da infraestrutura do identificador
Publicações, pesquisadores e dados;	Maduro
Citações, conferências, organizações, financiamentos, projetos, pesquisa, <i>software</i> , simulação computacional, instrumentos/plataformas, arquivos, amostras geológicas ou biológicas, artefato cultural e pessoa histórica ou mítica;	Emergente
Repositório de dados, experimentos, análises, licença de <i>software</i> , período ou local histórico, registros de ensaios clínicos e não clínicos, plano de gestão de dados, protocolos e <i>workflow</i> .	Imaturo

Fonte: adaptado de COUSIJN et al., 2021.

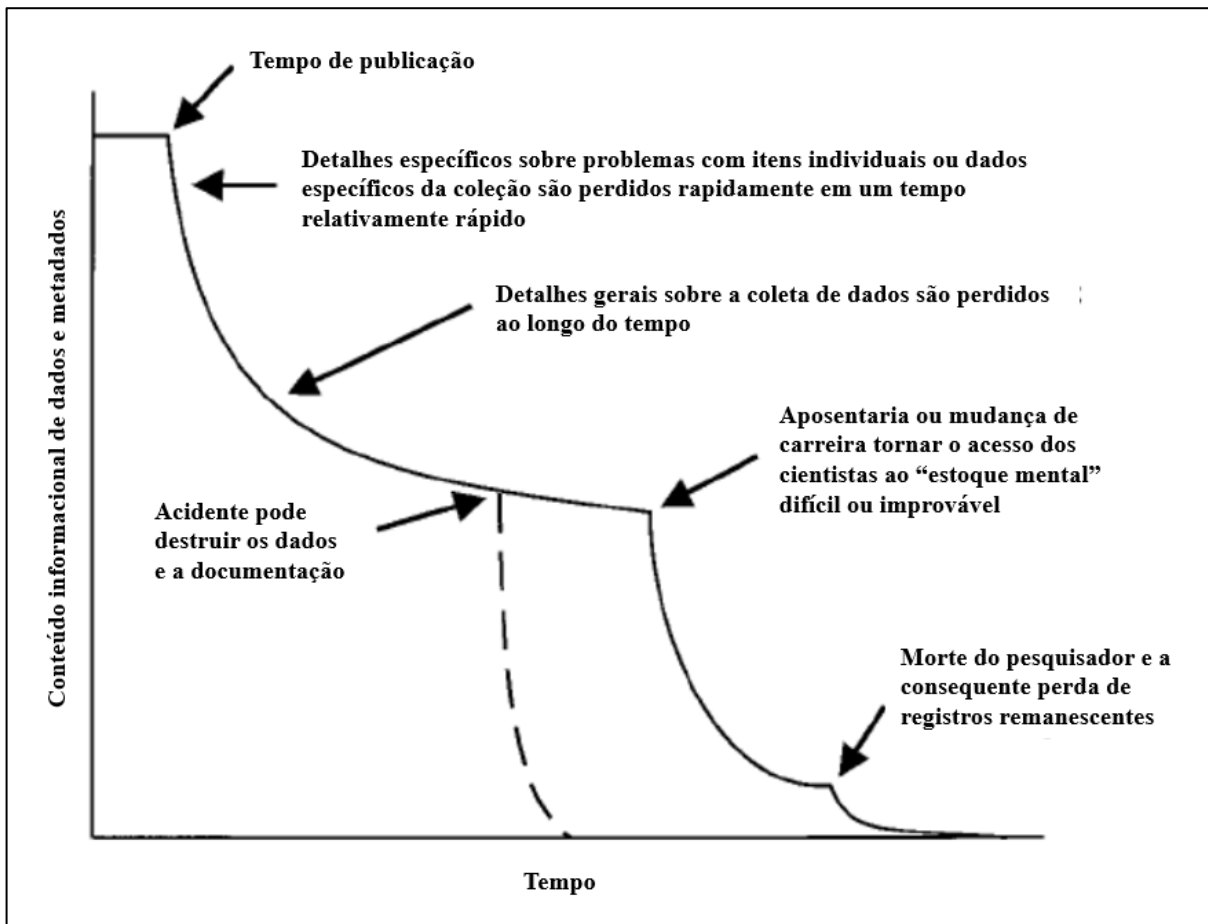
Os identificadores persistentes, por si só, não são capazes de representar a entidade de forma efetiva, necessitando que estejam ligadas outras informações expressas por meio dos metadados, e a outros serviços e plataformas, interligando esforços e tornando a proposta mais

integrada ao contexto científico de maneira geral. Cousijn et al. (2021) afirmam que os metadados associados que descrevem a entidade, seus padrões, os serviços associados e a comunidade envolvida são cruciais para garantir que os identificadores agreguem valor à infraestrutura de pesquisa.

Michener et al. (1997) afirmam em seu estudo que embora tenha mais de vinte anos, o principal motivo para investir no desenvolvimento dos metadados é a pouca memória humana; e nesse sentido apresentam um modelo de degradação do conteúdo informacional associados aos dados e metadados (Figura 1).

Na Figura 2, é possível observar o alto conteúdo informacional disponível assim que a publicação é realizada, e a perda de informações e detalhes sobre a pesquisa e o conteúdo publicado ao longo do tempo, iniciando por detalhes específicos e depois gerais.; A situação se agrava caso algo aconteça na trajetória do pesquisador, como seu afastamento da comunidade científica; e a qualquer momento, conforme detalha a figura, acidentes podem acontecer e todo o conteúdo da pesquisa ser perdido.

Figura 2 – Conteúdo informacional de dados e metadados



Fonte: adaptado de Michener et al. (1997, p. 332).

Nesse sentido, é importante refletir sobre o uso dos identificadores e a recuperação por eles proporcionada, se há de fato um manejo de dados a eles atrelados que possibilite que os metadados e dados sejam a eles satisfatoriamente vinculados.

Outros insumos utilizados durante o processo de pesquisa, e não somente as publicações, também precisam ser disponibilizados, representados e recuperados. Um exemplo são os dados de pesquisa que, conforme detalhado em tópico anterior sobre o tema, se forem disponibilizados poderão ser utilizados por outros pesquisadores, contribuindo significativamente com o desenvolvimento científico, uma vez que a publicação oferece apenas os percursos e resultados de pesquisa.

Contudo, o reuso desses dados fica condicionado à sua disponibilização por parte de autores, instituições e periódicos científicos. Nesse contexto, os princípios *Findable*, *Accessible*, *Interoperable* e *Reusable* (FAIR) são como um guia para editoras e produtores de dados sobre sua gestão, ajudando-os a potencializar o valor agregado obtido pela publicação em formato digital (WILKINSON et al., 2016).

Segundo Lecardelli, Monteiro e Lucas (2019), o processo de gestão de pesquisa está envolto em um ciclo de ações as quais propiciam sua recuperação tanto por humanos quanto por máquinas, portanto é sugerida a adoção dos Princípios FAIR, garantindo a qualidade dos dados de pesquisa ao mesmo tempo em que confere as condições necessárias para que sejam utilizados e citados corretamente, e sua sigla significa: **localizável**, quando o dado possui um identificador persistente e único; **acessível**, quando o dado pode ser recuperado por humanos ou máquinas; **interoperável**, quando o dado pode ser processado por máquinas a partir da utilização padronizada de vocabulários e ser semântica e sintaticamente acessível; e **reutilizável**, quando o dado possui metadados ricos e trabalhados adequadamente e que possam ser recuperados de forma eficaz.

Em outras palavras, os princípios FAIR são um modelo a ser seguido para a implementação de recursos digitais de pesquisa interoperáveis e que possam ser localizados e reutilizados pela comunidade científica (JUTY et al, 2019).

O depósito de dados sem a consciência desses princípios poderia implicar em dados armazenados sem perspectiva de uma recuperação eficiente e que de fato consiga atingir seu objetivo de contribuição para demais pesquisadores.

Nesse contexto, os recursos de pesquisa em ambiente digital, ou objetos digitais, para serem localizados e utilizados, devem estar disponíveis e representados. E a representação deles por meio dos identificadores persistentes, contribui diretamente para o primeiro princípio FAIR, o *findability*.

Alinhado aos princípios FAIR, de acordo com Ananthakrishnan et al. (2020), os identificadores são úteis se são únicos, persistentes, bem determinados por seus criadores e bem descritos para localizar os objetos. Os autores reúnem todas essas características:

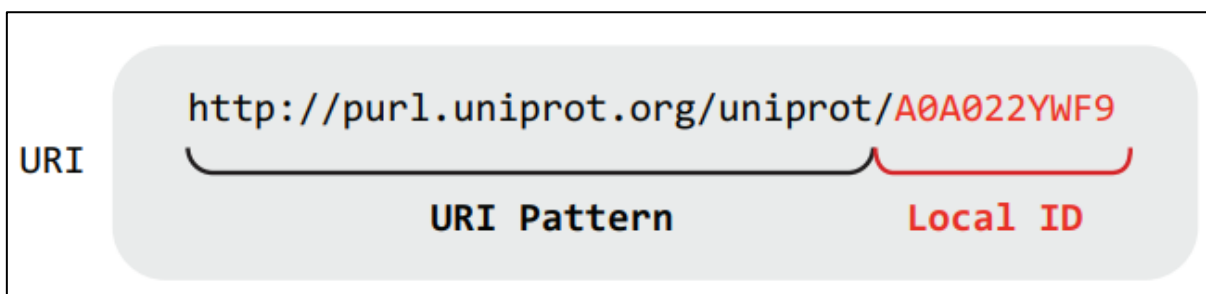
- a) **Univocidade:** o identificador deve ser globalmente único dentro de um dado contexto e sempre deve se referir a um único objeto ou um conjunto de objetos; uma vez que o identificador foi atrelado a um objeto, ele não pode ser realocado e utilizado para identificar outro;
- b) **Persistência:** o identificador não deve mudar, independentemente de o dado existir ou não futuramente, considerando sua hospedagem em ambiente digital;
- c) **Resolução:** o identificador deve ser mapeado para a página do recurso que identifica. A forma mais comum e recomendada para tornar os identificadores acionáveis é a URL;
- d) **Descrição:** um identificador formatado como URL deve redirecionar para uma página HTML, no mínimo, com metadados que descrevam o objeto representado. Para aumentar sua eficácia, é interessante que os metadados também sejam lidos por máquinas em um formato conhecido e não-proprietário (ANANTHAKRISHNAN et al, 2020).

Como anteriormente dito, os identificadores que foram desenvolvidos e começaram a ser aplicados fora do ambiente digital, como os de livros e periódicos, não possuem algumas das características mencionadas, sobretudo à sua resolução e descrição. Conforme afirmam Meadows e Haak (2018), a transformação para o mundo digital tem dependido dos identificadores, em grande parte, para se tornar legível por máquina.

Os identificadores em ambiente digital não devem apenas apresentar as características acima mencionadas, mas também devem possuir uma *Uniform Resource Identifier* (URI), que se trata de um tipo de protocolo de transferência de hipertexto com o qual o usuário consegue colar o endereço na barra de seu navegador de *internet* para então acessar o recurso que o identificador representa, sendo composta de duas partes (Figura 3): *URI Pattern*, um conjunto fixo de caracteres; e o *Local ID*, que é um código numérico único que garante que o recurso seja encontrado na fonte a qual pertence, neste caso, na própria base de dados do identificador

(BERNERS-LEE, 1998; MCMURRY et al, 2017). Uma URI começa com um nome que especifica o método utilizado para resolvê-la, como o “http” ou o “https”, fornecendo um protocolo para localizar e acessar o recurso informacional ao qual se refere; sendo que a URI também está incorporada no requisito *access* dos princípios FAIR (JUTY et al, 2019).

Figura 3 – Anatomia de um identificador baseado na *web*



Fonte: Mcmurry et al (2017, p. 2).

Lee e Stvilia (2014), enumeram uma série de requisitos a serem considerados na avaliação de qualidade dos identificadores, com base em publicações de diferentes pesquisadores, sendo os requisitos: singularidade; persistência; simplicidade; opacidade, verificabilidade; compatibilidade; interoperabilidade; acionabilidade; granularidade/flexibilidade; autoridade; escalabilidade; e segurança.

Por sua vez, o Quadro 5 mostra quais são as características desejáveis para que um identificador persistente seja de fato pertinente à comunidade científica:

Quadro 5 – Características desejáveis para identificadores

Características	Definição
Inequívoco	Um ID local deve estar associado a apenas uma entidade local. Uma URI deve estar associado a apenas uma entidade global;
Único	Uma entidade deve ser identificada por uma única URI;
Estável (identificador)	A URI e o ID local devem permanecer os mesmos ao longo do tempo;
Estável (entidade)	O identificador não pode ser atribuído a uma entidade diferente, ainda que a entidade original possa evoluir desde que um histórico seja documentado;
Possuir versões documentadas	A definição e os principais metadados da entidade podem sofrer mudanças, então o identificador deve ter um controle de versões e ter seu histórico documentado;

Persistência	O identificador não deve ser deletado, ainda que possa ser descontinuado;
<i>Web-resolvable</i>	A URI deve direcionar efetivamente para o local onde na <i>web</i> onde as informações sobre o registro podem ser acessadas;
Conversível	A URI e o Local ID devem ser interconversíveis;
Definido	O conjunto de identificadores devem ser descritos por meio de um padrão formal/expressão regular;
<i>Web-friendly</i>	O ID não deve necessitar de um manuseio especial quando usado em URL ou formatos de exportação como, por exemplo, XML;
Livre atribuição	O identificador deve ser atribuído sem nenhum custo aos indivíduos que o aderem e nele depositam dados;
Acesso e uso abertos	O identificador deve ser capaz de ser referenciado e acionado de forma transparente, em qualquer lugar, por qualquer pessoa e por qualquer motivo, como em uma pesquisa pública. Restrições aos dados podem ser aplicadas, mas não são recomendadas;
Documentado	O esquema de identificadores deve ser documentado.

Fonte: adaptado de Mcmurry et al. (2017, p. 5).

Os identificadores persistentes devem possuir a característica de interoperabilidade e capacidade de leitura de outros sistemas e ter a estrutura necessária para que os demais façam o mesmo com seus dados.

Contudo, Cousijn et al. (2021) discutem que a efetividade dos identificadores é comprometida se existirem falhas em seu processo de implementação, ou seja, um conjunto de fatores pode fazer com que a comunidade científica perca facilmente o interesse e engajamento pelo esforço de utilização dos identificadores em seus fluxos de pesquisa como, por exemplo, a não verificação de perfis no ORCID, DOIs que não estão registrados na CrossRef ou baixa qualidade dos metadados a eles associados.

Para que os identificadores persistentes alcancem seu objetivo de auxiliar na desambiguação de nomes de autores, instituições ou publicações, é preciso que ele tenha adesão por parte da comunidade científica, incluindo seus pesquisadores, instituições de pesquisa, editores, periódicos e instituições vinculadas ao desenvolvimento científico de forma geral. E esse apoio será alcançado também pelas facilidades de integração de dados e robustez em sua construção, otimizando o tempo do pesquisador no preenchimento de múltiplas plataformas e demonstrando que colabora com a visibilidade ao mesmo tempo que é capaz de conectar diversas informações que provém de outros provedores de dados.

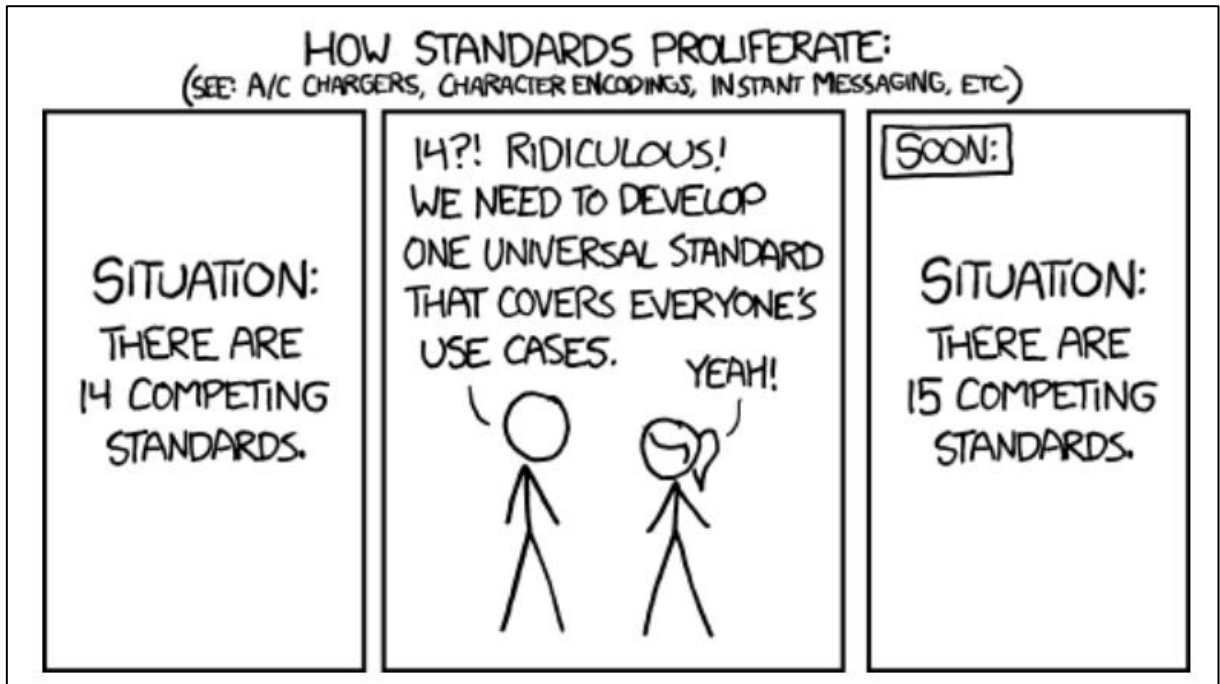
Car, Golodoniuc e Klump (2017) discutem a respeito da quantidade de identificadores persistentes que surgiram nas últimas duas décadas, desenvolvidos por diferentes comunidades e com objetivos sobrepostos, mas todos tinham uma característica em comum: seguiram o mesmo ciclo de vida e entraram em um estado de paralisia, no qual continuaram existindo, mas sem acrescentar mais à área. Beck, Ritz e Wittenburg (2016) e Huber e Klump (2016) chamaram este estado de paralisia de “estágio zumbi”.

Dappert et al. (2017) afirmam que a variedade de sistemas de identificação persistente reflete as diferenças nos tipos de entidades sendo representadas, as exigências da comunidade científica e fatores históricos, ou seja, não existe um único identificador persistente que regule todas essas questões, embora o ecossistema de identificadores apoie a diversidade e as iniciativas de ciência aberta, sendo também a favor da interoperabilidade.

Observa-se que, em relação aos identificadores, realmente há na comunidade científica uma rápida disseminação de padrões e tentativas de otimizar o fluxo de comunicação científica e recuperação informacional, mas isso resulta em diferentes iniciativas que pouco contribuem para que um padrão seja de fato estabelecido e amplamente difundido e adotado pelos pesquisadores e todos os envolvidos em pesquisa. O desenvolvimento de identificadores, persistentes ou não, continuará e o surgimento ou desuso de muitos deles é uma consequência natural desse processo, sobretudo em ambiente digital.

A Figura 4 demonstra essa questão, ilustrando que a insatisfação diante de tantos “padrões” estabelecidos e a vontade de criar um padrão considerado universal, contribui para a criação de um novo “padrão”, aumentando mais uma vez a quantidade de recursos disponíveis e descentralizando uma possível solução aplicável no geral.

Figura 4 – Proliferação de "padrões"



Fonte: Huber e Klump (2016, p. 16).

É indispensável que estudos sobre identificadores persistentes continuem sendo feitos para que seja possível entender a melhor forma deles avançarem e sobreviverem às mudanças tecnológicas e aos processos em ambiente digital (CAR; GOLODONIUC; KLUMP, 2017).

E nesse sentido, existe um evento realizado especificamente para tratar de identificadores persistentes, o PIDapalooza, sendo que o nome do evento é inspirado no conhecido festival de música Lollapalooza. O evento teve sua primeira edição realizada em 2016, depois foi realizado também nos anos de 2018 e 2021.

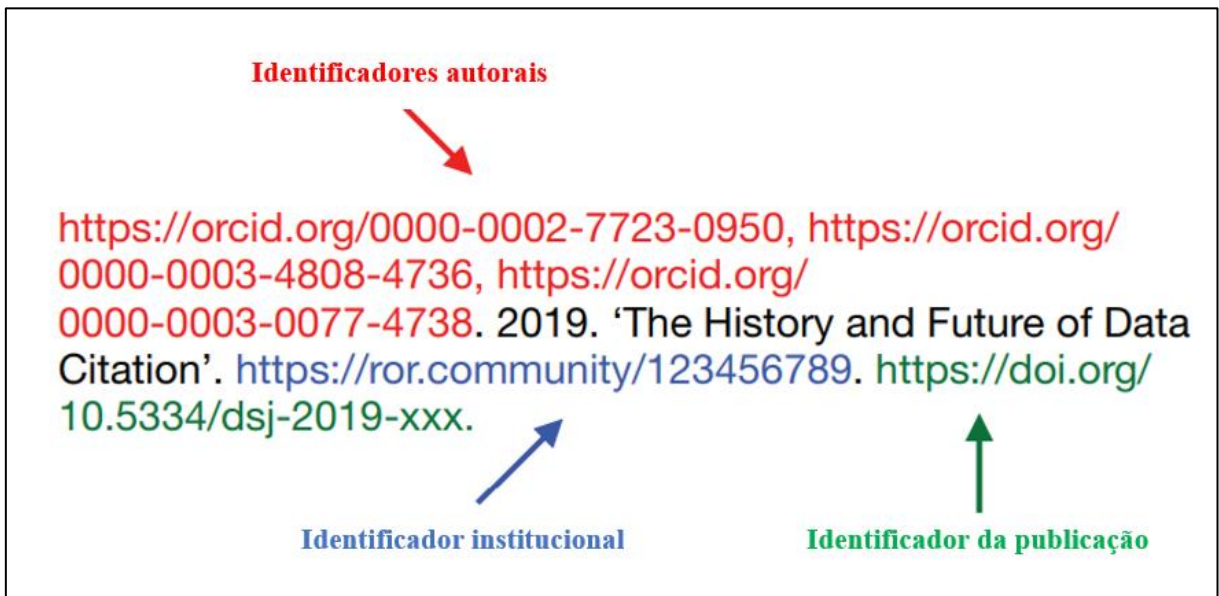
O PIDapalooza, organizado por California Digital Library, Crossref, DataCite, NISO e ORCID, tem o objetivo de reunir entusiastas que utilizam ou criam identificadores persistentes voltados para a comunidade científica, com sessões de discussão informais e interativas, buscando responder questionamento como: quais tipos de PIDs serão necessários no futuro; como eles devem ser utilizados; e qual é a melhor forma de convencer pesquisadores a adotarem os identificadores (MEADOWS, 2017).

Um evento como esse, específico sobre identificadores persistentes, pode suscitar discussões pertinentes sobre seu futuro e sua aplicabilidade, despertando novas possibilidades que otimizem a interoperabilidade entre sistemas que de fato beneficiem todos os recursos de pesquisa disponíveis na comunidade científica.

Parsons, Duerr e Jones (2019) em seu estudo sobre o futuro das citações na comunidade científica, tradicionalmente realizadas para atribuir créditos aos trabalhos utilizados como insumo para desenvolvimento de outros, mencionam que as pessoas estão começando a repensar a forma que os identificadores funcionam e a necessidade de se manter a persistência dos objetos e suas referências, de forma conjunta; nesse sentido, afirmam que muitos gestores de dados estão olhando para os identificadores baseados em conteúdo, ou *content-based identifiers*, em detrimento dos identificadores baseados em entidades, sendo utilizados para rastreamento e consulta precisa de dados.

Inserir apenas os identificadores persistentes para montar uma citação de uma publicação científica (Figura 5) pode torná-la mais precisa e inteligível por máquina, porém que não satisfaz completamente o olhar humano (PARSONS; DUERR; JONES, 2019).

Figura 5 – Citação imaginária para uma publicação fazendo o uso completo de identificadores persistentes



Fonte: adaptado de Parsons, Duerr e Jones (2019, p. 6).

Dessa forma, os diversos atores da comunidade científica necessitam trocar informações entre si, colaborando para o desenvolvimento de todo esse sistema, também por meio de identificadores. Com o desenvolvimento de novas formas de uso dos identificadores, acredita-se que novos padrões podem surgir em relação à menção de citações e atribuição correta de autorias e vínculos a entidades, facilitando a leitura de máquinas e de humanos.

4.1 Identificadores bibliográficos

Mesmo antes da consolidação da tecnologia como integrante da rotina de pesquisa na comunidade científica, o volume informacional crescente demandava por formas mais eficientes de organização e recuperação dos conteúdos disponíveis pelos pesquisadores.

Dessa maneira, o cenário propiciou o surgimento de sistemas numéricos de identificação de publicações: chamados neste trabalho de identificadores bibliográficos, especificamente serão abordados o ISBN, DOI e ISSN.

Nesse contexto, tem-se o ISBN como o pioneiro dos identificadores bibliográficos, sendo amplamente conhecido devido à sua atribuição para a identificação de livros. A ISO por meio de seu comitê de documentação iniciou um estudo sobre a possibilidade de adaptar o sistema *Standard Book Numbering*, que tinha sido desenvolvido no Reino Unido em 1967 pela rede de livrarias W. H. Smith para identificar corretamente as edições de seus livros, para uso internacional; surgindo então o ISBN em 1970 como norma ISO 2108/1972, funcionando como uma agência internacional e agências de registros em diversos países (CAMPELLO, 2019).

O ISBN pode ser atribuído a diferentes materiais: livros físicos, *e-books*, *audiobooks*, *CD-ROMs*, microformas etc. (INTERNATIONAL STANDARD BOOK SYSTEM, 2017). Seu número identificador era composto por nove dígitos mais um dígito de controle, totalizando 10 dígitos, separado por hifens em quatro segmentos (CAMPELLO, 2019), mas a partir de 2007, o número ficou composto por 13 dígitos em vez de 10.

Os 13 dígitos indicam o título, o autor, o país, a editora e a edição de uma obra (Figura 6): os três primeiros dígitos são determinados pelo GS1, uma organização que gerencia dados dos códigos de barra de produtos, sendo conhecido como código GTIN; os próximos dois dígitos identificam o país, que por muitos anos foi identificado com o número 85, identificando o Brasil, mas a partir de 2018 o 65 também começou a ser utilizado, devido à alta demanda; os próximos cinco dígitos identificam o editor ou marca que vai realizar o registro; os outros dois números identificam a publicação; e o último dígito, chamado de dígito de controle, garante que o identificador seja único e é calculado utilizando um algoritmo (CÂMARA BRASILEIRA DO LIVRO, 2021).

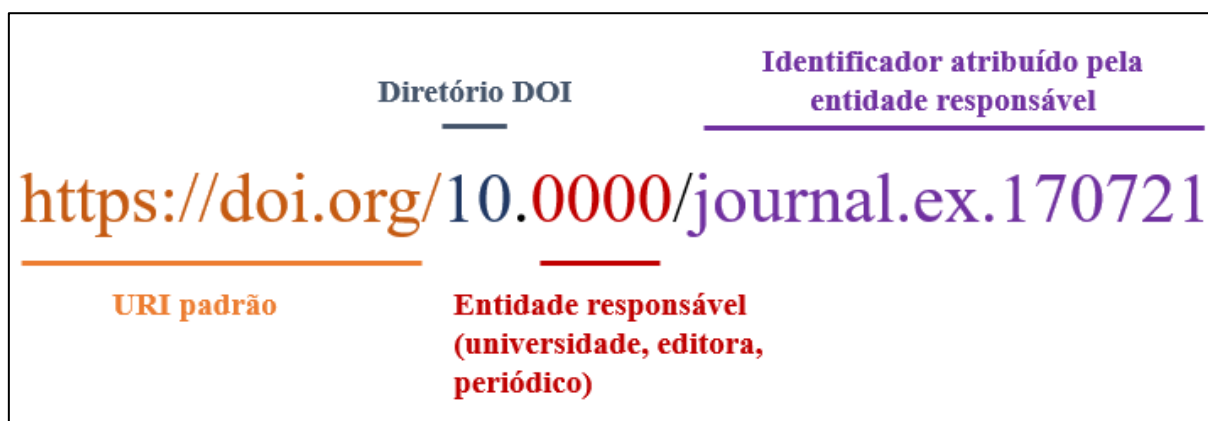
Figura 6 – Estrutura do ISBN

Fonte: Câmara Brasileira do Livro (2021).

Se para livros tem-se o ISBN, em ambientes digitais os documentos também são identificados com código numéricos, como o DOI. O DOI pode ser atribuído a objetos disponibilizados em meio digital sendo implementado largamente por grande parte das editoras e organizações de pesquisa, facilitando a identificação e a recuperação de documentos. O sistema foi iniciado em 1997, ocasião em que foi apresentado na Feira Internacional do Livro em Frankfurt (CAMPELLO, 2019). O DOI pode ser atribuído a publicações científicas, documentos de texto, fotografias, *softwares*, áudio etc. (LEE; STVILIA, 2014).

O DOI é de responsabilidade da *International DOI Foundation*, mas possui diversas agências de registro que são as responsáveis por de fato coletar os metadados e registrar o identificador, as mais amplamente conhecidas na comunidade científica são a CrossRef e a DataCite.

Esse identificador persistente é composto por um nome DOI, que tem sua URI completa formulada com o padrão https://doi.org/<DOI_name>, no qual o nome DOI possui um prefixo e um sufixo separados por uma barra (Figura 7), sendo que o prefixo é um código da instituição ou periódico responsável pelo registro do DOI e o sufixo do identificador que é atribuído por ela (JUTY et al., 2020).

Figura 7 – Itens que compõem o DOI

Fonte: elaborado pela autora.

O identificador atribuído pela entidade responsável, ao final do DOI, pode ser determinado pela própria entidade de forma livre, embora seja recomendado que se adote um padrão, por exemplo, há periódicos científicos que padronizam com o número do volume, fascículo ou data de publicação.

No momento de registro de um DOI junto a uma das agências de registro, como a CrossRef, é necessário o envio do próprio identificador, de acordo com o padrão atribuído pelo periódico, junto com os metadados do conteúdo, como uma publicação científica. É pertinente que os metadados sejam ricos em informações, tanto em relação ao próprio artigo que está sendo identificado, quanto às informações adicionais como, por exemplo, os identificadores persistentes de autor, que posteriormente serão utilizados para alimentar os perfis de autores de maneira automática, como será mencionado mais adiante.

E além de identificadores voltados para publicações, sejam elas livros ou outros documentos em formato digital, são atribuídos códigos numéricos para identificar as fontes informacionais que publicam atualmente grande parte da literatura científica disponível, que são os periódicos científicos ou outras publicações seriadas.

O ISSN surgiu da ideia de criação de um sistema de numeração padronizada que identificasse os periódicos científicos em 1967 durante a 16ª Conferência Geral da UNESCO e da Assembleia Geral do Conselho Internacional de Uniões Científicas, pois durante o evento os participantes suscitaram a preocupação de encontrar uma forma que possibilitasse o cadastro e identificação de periódicos à nível internacional, considerando que se trata da forma mais utilizada para difundir literatura científica (CAMPELLO, 2019).

O ISSN é geralmente conhecido devido a possibilidade de busca por seu número em muitas bases de dados bibliográficas, quando existe o objetivo de recuperar todas as publicações científicas atreladas a determinado periódico.

O registro do ISSN inclui informações, relativas aos periódicos, como: o próprio ISSN, título, título abreviado, outras versões do título, títulos anteriores, periodicidade, idioma, local de publicação, editor ou formato de publicação (CAMPELLO, 2019). O ISSN possui oito dígitos, dividido na exibição por dois grupos separados por um hífen, precedido das letras ISSN (REYNOLDS, 2015).

Existem também outros identificadores bibliográficos como o *International Standard Music Number* (ISMN) para publicações musicais ou o *International Audiovisual Number* (ISAN) para publicações audiovisuais, além de outros que já foram ao menos mencionados no tópico anterior deste trabalho. Todos estes identificadores possuem características próprias e adequadas para representar as particularidades de seus materiais.

Os identificadores de publicações, sejam persistentes ou não, contribuem para a correta recuperação da informação científica, pela facilidade de identificação de um conteúdo ou um conjunto deles. Embora atualmente haja a necessidade de encontrar novas formas ou metodologias de organização da ciência em ambiente digital, antes já se faziam necessárias técnicas que otimizassem o tempo de profissionais e pesquisadores. Essa necessidade hoje se estende também para outros espectros da comunidade científica como a de identificação da produção científica de instituições e de autores, que enfrenta diversos problemas com a variações de seus nomes e que impactam diretamente em avaliações e visibilidade de sua pesquisa, e também de seu perfil como cientista nas respectivas áreas de conhecimento e atuação.

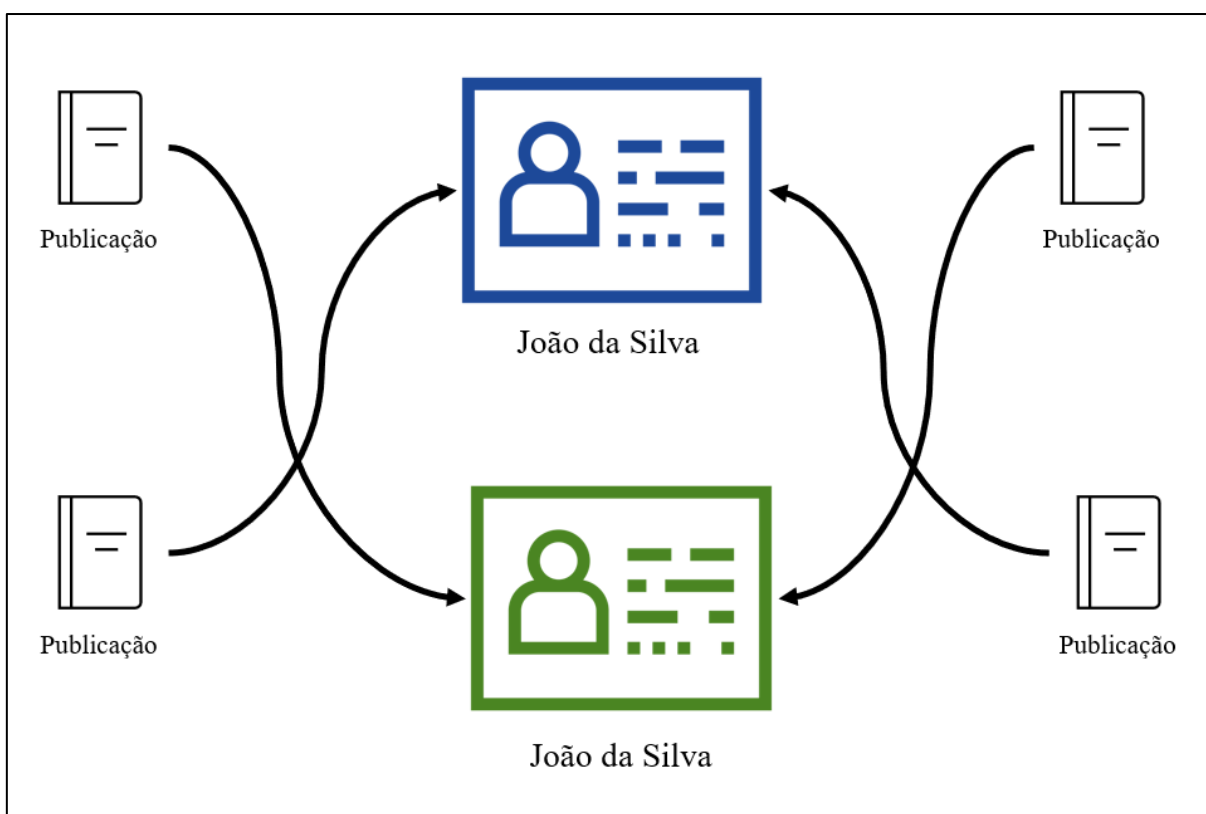
4.2 Identificadores autorais

A trajetória de um pesquisador na academia demanda estudo e dedicação, além do incentivo por parte das universidades, instituições de pesquisas e agências de fomento. Em muitos momentos da jornada, o pesquisador se depara inevitavelmente com as avaliações de seu desempenho, que normalmente são medidas por meio de sua produção científica e seu respectivo alcance. São muitas as discussões no meio científico a respeito dos métodos empregados nos processos avaliativos e se vê necessária a aplicação de uma metodologia que vislumbre não apenas a visibilidade da pesquisa, mas que também consiga visualizar o

pesquisador e seu percurso. Essa é uma tarefa que, embora ideal, encontra alguns desafios, sobretudo em relação à correta identificação desses pesquisadores na comunidade científica.

Assim como atribuídos a outros recursos acadêmico-científicos, como publicações, dados de pesquisa e instituições, aos autores também podem ser atribuídos identificadores únicos, com o objetivo de identificá-los de forma inequívoca em relação a outros com nomes iguais ou semelhantes (Figura 8) e de si próprios, caso ocorra alguma mudança em seus nomes, informação óbvia para sua identificação, mas sensível e suscetível a variações que podem determinar a sua recuperação.

Figura 8 – Identificação de autoria com nomes iguais



Fonte: elaborado pela autora.

Em 2006 surgiram alguns dos primeiros trabalhos que cogitaram a criação e utilização de um número identificador único para autores, ou *unique author identification number* (UAIN), relatando seu impacto direto na comunidade científica (FALAGAS, 2006; JOLY, 2006). Aerts (2008) discutiu sobre a existência de identificadores para publicações, como é o caso do DOI, e a ausência de identificadores para autores, naquele contexto e período, mostrando seu questionamento a respeito do tópico.

Construir um sistema de identificadores de autor não apresenta apenas desafios técnicos, mas também implica em decisões sobre abertura de dados, privacidade, colaboração e modelos de negócio, por este motivo as iniciativas demoraram consideravelmente para serem levantadas pela comunidade científica (FENNER, 2011).

A utilização dos identificadores é útil para autores, não apenas em relação aos aspectos de avaliação de suas pesquisas, embora também seja um fator a ser considerado, mas vislumbrando otimizar a troca de experiências e relacionamentos entre pares na comunidade científica, otimizando a construção de parcerias que podem impactar diretamente no desenvolvimento científico de suas áreas. Nesse sentido, ao acessar um artigo que está corretamente identificado com seus autores, aumentam as chances de visualizar o percurso de pesquisa e outras publicações realizadas, e a possibilidade do contato com interessados em colaborações acadêmicas.

Seria também de grande utilidade que os identificadores autorais fossem incorporados no *workflow* dos periódicos científicos (HAAK et al., 2012), de forma que o identificador já seja solicitado ao autor no momento da submissão da publicação na revista, tornando-se uma exigência para publicação em periódicos que passam pelo processo de revisão por pares (KENDALL; YEE; HARDY, 2017). Para que os identificadores sejam efetivos, eles precisam ser amplamente adotados pela comunidade científica, não apenas pelos próprios pesquisadores, mas também por outros processos e instituições diretamente envolvidas no ciclo da pesquisa: submissão, depósito de dados de pesquisa, instituições de fomento etc. (HAAK et al., 2012).

Considerando o tempo dispendido para realizar o controle de autoridades em grandes volumes de informação, como é o caso dos artigos de periódicos, as bases de dados internacionais, que ocupam uma posição de alto destaque e relevância na busca por literatura científica, precisam se engajar na consolidação e atribuição de identificadores autorais. “Quando os catalogadores podem criar *links* entre fontes de dados pré-existentes, em vez de perder tempo na criação de uma nova autoridade, o controle se torna muito mais viável” (SANDBERG; JIN, 2016, p. 539, tradução nossa). Na prática, seria a atribuição do identificador a cada uma das autoridades presentes no registro bibliográfico, fazendo com que este e outros registros do autor sejam recuperados por aquele único código. Essa prática está intimamente relacionada ao conceito de *Linked Data*, conforme anteriormente abordado.

Nesse sentido, a *Online Computer Library Center* (OCLC) foi pioneira no controle de autoridades e no estabelecimento de identificadores únicos para autores, pois o *Virtual International Authority File* (VIAF) é consequência dos esforços da OCLC em prover uma base

de autoridades internacional (FOX, 2016). O VIAF foi criado em 2003 e possui a colaboração de diversas bibliotecas.

Cada autor que consta no VIAF possui sua *string* única, atribuída por meio de uma URI, que o identifica e possibilita a recuperação de todas as formas de seu nome, nomes relacionados, países de publicação e as suas próprias publicações (FOX, 2016).

Como normalmente se pensaria, a OCLC não esteve à frente de pesquisas sobre o tema, pensando apenas na recuperação da informação pelos usuários das bibliotecas. O que evidencia a necessidade das universidades em levantar a produção acadêmica de seus pesquisadores e as dificuldades encontradas diante da falta de controle de autoridade.

A OCLC investiu muito em pesquisas para representar pesquisadores com identificadores e *linked data*. O foco da motivação esteve nas necessidades das universidades em rastrear a produção científica, ao invés das necessidades dos usuários finais das bibliotecas (SANDBERG; JIN, 2016, p. 541, tradução nossa).

Foram criados identificadores para autores, não restritos ao meio acadêmico, pelo próprio ISNI e pela *Library of Congress*, dos Estados Unidos, ambos utilizam *Linked Data* em alguma medida. Depois surgiu a necessidade de criar identificadores voltados especificamente para acadêmicos, para recuperar suas pesquisas científicas, então editoras e organizações desenvolveram identificadores de autor para este público, sendo que os repositórios *Research Papers in Economics* (RePEc) e arXiv foram os primeiros a incluírem os identificadores para os autores, em 1999 e 2005, respectivamente (BOUDRY; DURANT-BARTHEZ, 2020). Outros identificadores autorais foram criados nos anos seguintes, sendo que o ScopusID foi criado em 2006, o *ResearcherID* em 2008 e o ORCID em 2012.

[...] são muitos interesses diferentes nas identidades de autor/pessoa e seus usos, sugerindo que haverá muitos sistemas diferentes e múltiplas identidades para cada autor. No domínio da comunicação científica, haverá uma mistura de sobreposições de redes de publicação. A menos que um sistema cresça para dominar, as diferentes redes de publicação vão identificar os autores utilizando identificadores diferentes. No entanto, será muito mais fácil combinar vários identificadores para cada autor do que desambiguar múltiplos autores com o mesmo nome [...] a capacidade de registrar identificadores estrangeiros no registro do autor dentro de um repositório e a capacidade de combinar artigos entre repositórios, permitirá a junção dos dados além dos limites do repositório (WARNER, 2010, p. 3, tradução nossa).

No contexto acadêmico-científico, ficam claros os distintos interesses de grupos editoriais e organizações em criar identificadores de autor, resultando em diferentes iniciativas com o mesmo objetivo, sendo que muitas delas são direcionadas apenas para seu “universo de atuação”, ou seja, seu próprio repositório ou base de dados associada.

Embora Warner tenha alertado a respeito dessa questão ainda em 2010, quando já tinham surgido alguns dos identificadores atrelados a grandes grupos como o da Elsevier, com o *ScopusID*, e da *Clarivate Analytics*, com o *ResearcherID*, estava por surgir esse sistema “dominante” ou mais amplamente adotado e utilizado, que seria o ORCID, em tese não vinculado diretamente a nenhum grupo ou empresa. Seu crescimento, utilização e ampla aceitação na comunidade científica atingiu um patamar não conhecido pelos demais identificadores.

A utilização de diferentes identificadores de autor desperta uma múltipla atenção dos pesquisadores e a sensação de que existem diversas plataformas com o mesmo objetivo, como de fato é verdade, considerando que demanda tempo para que estes autores verifiquem seus diferentes perfis e o alimentem ou façam ajustes, quando se fizerem necessários. Esse aspecto remonta às discussões a respeito do posicionamento dos pesquisadores nas redes acadêmico científicas: muitas plataformas, funcionalidades e relacionamentos, o que demanda muito tempo na manutenção de todos esses ambientes, fazendo-os escolher por aqueles que mais se adaptam, ou que fazem sentido em sua trajetória ou área de pesquisa.

Embora essa situação não seja perfeita, poderá ajudar na montagem de uma rede que possibilite uma ligação entre publicações e seus autores, oferecendo suporte para novas ferramentas que poderão abranger várias fontes desse tipo de informação (WARNER, 2010). É necessário que os estudos a respeito dos identificadores e perfis de pesquisadores na *web* não deixem de avançar e questionar os atuais padrões e ferramentas, para que novas e otimizadas formas de os identificar possam surgir.

Warner (2010) também reflete sobre a rapidez de adoção e aceitação dos identificadores autorais pela comunidade científica. As amplas possibilidades, conforme mencionado, também podem comprometer uma maior adesão dos membros da academia. Essa falta de engajamento dos pesquisadores com os identificadores é utilizada inclusive como justificativa àqueles que utilizam as técnicas de desambiguação de nomes de autores.

Antes desses esforços produzirem resultados para cobrir uma parcela considerável do mundo dos autores ou publicações, a desambiguação de nomes de autores (AND), o processo de identificar indivíduos e suas obras em

um campo da literatura, seja ele automático ou manual, é necessário para a bibliometria baseada nos autores (STROTMANN; ZHAO; 2012, p. 1822, tradução nossa)

Críticas são realizadas quanto a efetividade dos identificadores autorais considerando aqueles que dependem ativamente de seus pesquisadores. Dervos et al. (2006) apresentaram um sistema universal de identificação autoral denominado *The Universal Author Identifier System* (UAI_Sys), que forneceria um identificador único, uma *string* com 16 números, a ser utilizado pelos pesquisadores pelo resto de suas vidas e que seria alimentado com metadados pelos próprios autores. Este projeto se assemelha muito ao identificador persistente ORCID, utilizado atualmente na comunidade científica.

Mas com base no projeto UAI_Sys, Smaiheiser e Torvik (2009), fazem uma crítica ao comentar que, embora tecnicamente simples e viável, ele falha por não considerar o real comportamento do pesquisador e esperar que todos participem preenchendo suas informações acadêmicas.

Realmente pode ser utópico imaginar que todos os pesquisadores, dentro de suas rotinas de pesquisas, terão tempo disponível para investir no preenchimento dos perfis de identificadores ou de redes acadêmicas. Contudo, uma conscientização por parte das instituições de pesquisa a respeito dessa adoção e uso poderia ajudar muito os pesquisadores.

Álvarez de Toledo (2013) afirma que é preciso informar os pesquisadores e convencê-los a respeito da necessidade de aderir e revisar seu identificador digital e a olhar com atenção à forma com a qual seu nome é exibido nas publicações científicas, considerando que isso impacta diretamente em suas métricas de autor, como o *h-index*; ainda que, alguns erros não sejam cometidos especificamente pelos autores, mas sim pelas bases de dados, então os identificadores podem auxiliá-los a resolver este problema.

Mas talvez o maior desafio seja cultural. Quer o ORCID ou algum outro sistema de identificação do autor se torne o padrão aceito, as novas métricas possibilitadas precisarão ser levadas a sério por todos os envolvidos no sistema de recompensa acadêmica – agências de financiamento, administrações universitárias e comitês de promoção e estabilidade. Cada papel na ciência deve ser reconhecido e recompensado, não apenas aqueles que produzem publicações de alto perfil (CREDIT..., 2009, p. 825).

A adoção e ampla utilização dos identificadores persistentes de autor é um desafio na comunidade científica, sobretudo devido ao surgimento de diversas iniciativas com o mesmo objetivo e que demandam retrabalho e dispendem tempo do pesquisador. É preciso que as

pesquisas sobre a temática avancem e que sua utilização seja realizada por diversos atores na comunidade científica, de maneira a proporcionar uma interoperabilidade de informações entre servidores e plataformas, garantindo autenticidade dos conteúdos inseridos nos perfis dos autores e demandando menos tempo de manutenção.

4.3 Identificadores organizacionais

As universidades ou instituições de pesquisa devem manter altos padrões de qualidade e constante desenvolvimento, sempre visando seu reconhecimento e contribuindo ativamente com o desenvolvimento científico em suas respectivas áreas de atuação. Existem identificadores persistentes utilizados também para organizações e não somente para recursos como publicações ou autores, considerando que elas também podem sofrer mudanças em seus nomes ao longo do tempo, além de serem mencionadas nas afiliações institucionais de diferentes formas.

A atribuição do nome da instituição em uma publicação científica e sua consequente indexação em bases de dados bibliográficas, impacta diretamente, assim como dito anteriormente, na elaboração de indicadores métricos científicos, geralmente fazendo com o que indicador da organização de pesquisa tenha números abaixo do real, devido a publicações com variações que não foram automaticamente identificadas por essas fontes informacionais. Huang et al. (2020, p. 381, tradução nossa) afirmam que “a unificação das informações institucionais é essencial para a recuperação da informação, análise bibliométrica e avaliação da pesquisa baseada no nome da instituição”.

As formas com a qual a instituição é inserida pelos autores no momento de submeter as publicações em periódicos científicos variam por diferentes motivos, dentre eles estão a falta de padronização entre os periódicos científicos ao solicitarem esse tipo de informação, considerando que alguns podem solicitar as afiliações que constem nos respectivos departamentos de origem ou até peçam para que a tradução do nome da instituição seja realizada, condicionando o pesquisador a seguir suas regras, ainda que essas entrem em conflito com o entendimento dele sobre afiliações padronizadas.

Desde a década de 1990, De Bruin (1990) já afirmava que eram grandes as variações nas afiliações ou endereços que constavam nas publicações científicas, tanto com relação ao nome das próprias instituições quanto aos nomes de cidades e países nas quais se localizavam, gerando diferentes entradas para a mesma organização nos índices das bases de dados,

prejudicando os estudos bibliométricos e a identificação correta da instituição, considerando que muitos poderiam não conhecer e associar variações não usuais a determinada organização.

Outra razão muito comum para a inserção da afiliação institucional sem um padrão está no desconhecimento por parte dos autores de qual é o nome pelo qual a instituição deve ser mencionada nas publicações científicas. Este é um papel fundamental das universidades e instituições de pesquisa: possuem nominais para afiliação institucional, ou seja, documentos que busquem orientar seus autores de maneira adequada para a inserção da afiliação no momento de realizar a publicação científica, considerando suas diferentes especificidades. Além de possuírem, essas nominais precisam ser amplamente divulgadas e acessíveis ao corpo docente e discente.

As orientações e a adoção da forma padronizada do nome institucional são fundamentais e implicam diretamente na recuperação de sua produção científica nas fontes de informação, sobretudo nas bases de dados bibliográficas. As orientações podem mitigar estas variações decorrentes da ação dos pesquisadores, mas os erros que impactam na recuperação da produção institucional nas fontes informações podem sofrer impacto também de erros ortográficos, de digitação e até de erros cometidos pelo *Optical Character Recognition* (OCR) na edição (HUANG et al., 2020).

O uso de identificadores persistentes também para instituições tem a proposta de recuperar sua produção científica de uma forma mais prática e exata, sem a necessidade de mapeamento de variações do nome e inserção de todas elas na busca.

De acordo com Dappert et al. (2017), as organizações são difíceis de identificar devido às suas estruturas hierárquicas e categorização internas, além de sua reestruturação dinâmica ao longo do tempo.

Assim como os demais identificadores mencionados, os identificadores de organização possuem características semelhantes, sempre representam o recurso informacional por meio de um código numérico.

Bilder, Brown e Demeranville (2016) afirmam que os identificadores de recursos de pesquisa e de pesquisadores são dois importantes pilares na infraestrutura de comunicação científica, contudo um terceiro pilar é necessário para trazer de fato eficiência ao processo, que são os identificadores organizacionais; e embora muitas iniciativas tenham surgido, nenhuma se tornou equivalente, em termos de eficiência e adequação, aos demais identificadores de outros recursos, mostrando que não há uma infraestrutura adequada à essa área.

Considerando essa necessidade por um identificador organizacional, CrossRef, DataCite e ORCID se uniram na criação do projeto de um identificador organizacional aberto e independente (CRUSE; HAAK; PENTZ, 2016). Este projeto se consolidou recentemente, em 2019, no identificador organizacional ROR.

Para agências de fomento que precisam identificar e monitorar os resultados de pesquisas que financiaram, o auxílio dos identificadores na recuperação das publicações otimiza o processo, além de conectar pesquisas, instituições e pesquisadores por meio dos metadados fornecidos e registrados no DOI da publicação (MEADOWS; HAAK; BROWN, 2019).

Nesse sentido, os identificadores organizacionais ou institucionais também colaboram com a interoperabilidade na comunidade científica nacional e internacional, uma vez que os metadados das publicações podem conter os identificadores da publicação, dos autores e da instituição; conectando todos os esforços e trocando informações entre perfis de todos os envolvidos naquele recurso.

Um único PID possibilita estabelecer interoperabilidade entre sistemas, compartilhamento de recursos e correlação de informações institucionais, assim como formar uma base de dados de especificações institucionais com informações valiosas, e prover suporte de dados para uma recuperação precisa e análise estatística dos resultados de pesquisa científica institucionais (HUANG; LI; SUN, 2020, p. 383, tradução nossa).

De acordo com Haak (2014), essa conexão entre recursos tem potencial para transformar a forma com a qual a pesquisa é atualmente atribuída, citando o exemplo das organizações que desejam validar o curso de graduação de um estudante, e o fazem normalmente pedindo uma cópia de seu diploma, mas: existem universidades que atribuem identificadores, como o DOI, aos seus trabalhos acadêmicos; existem estudantes que incluem um identificador de autor, como o ORCID no momento de depositar seu trabalho; e existem organizações que incluem os metadados deste trabalho entregue no ORCID do estudante com seu identificador organizacional, como o Ringgold, fazendo com que a informação de que o estudante possui o curso de graduação completo, evidenciado pela entrega do trabalho pela instituição, seja validada.

Dessa forma, embora os identificadores organizacionais se mostrem fundamentais para tornar o processo de comunicação científica e interoperabilidade entre sistemas mais robusto, a comunidade científica ainda engatinha em iniciativas que realmente se mostrem consolidadas,

aplicáveis e adotadas pela comunidade científica. É importante que as pesquisas sobre a temática avancem para que novas propostas ou avanço de iniciativas já existentes consigam direcionar as organizações para o uso de seus próprios identificadores, se possível, persistentes, considerando o ambiente digital e a sua dinâmica.

5 ANÁLISE E APLICABILIDADE DE IDENTIFICADORES ORGANIZACIONAIS E AUTORAIS

Os identificadores e identificadores persistentes são códigos alfanuméricos utilizados como estratégia para otimizar o processo de recuperação da informação científica em fontes de informação disponíveis, sobretudo, em ambiente digital.

Sua estrutura está diretamente vinculada com padrões que os possibilitam interoperabilidade e conexão entre si, de modo que o perfil de autores e instituições, por meio dos identificadores autorais e organizacionais, seja constituído com a maior gama de informações possíveis a respeito de sua trajetória na comunidade científica, incluindo sua produção científica, que é identificada por meio dos identificadores bibliográficos.

A adesão aos identificadores é uma preocupação da comunidade científica e das organizações que mantêm os próprios identificadores, sejam eles pertencentes ou não a grupos editoriais com interesses majoritariamente financeiros. As iniciativas se mostram fundamentais enquanto seu potencial de auxiliar no processo de recuperação da informação científica. Contudo, são muitos os identificadores disponíveis para os pesquisadores e para as instituições, gerando certa angústia na comunidade científica de forma geral em relação à priorização de algum deles, se for o caso, para sua alimentação e manutenção periódicas.

Com base em todos os aspectos discutidos ao longo deste trabalho, desde os conceitos sobre organização e representação da informação científica em ambiente digital, perpassando por todos os aspectos da recuperação e visibilidade da pesquisa, do pesquisador e das instituições e finalizando com as principais características e requisitos que caracterizam os identificadores e identificadores persistentes, abordando ao longo de toda a discussão também as críticas quando realizadas pela literatura científica da área, é fundamental que seja feita uma análise a respeito da aplicabilidade dos identificadores, especificamente os autorais e organizacionais, na organização da informação científica.

A análise será realizada em duas fases: por meio do detalhamento de algumas iniciativas, de identificadores organizacionais e autorais, realizado de forma exploratória e mediante argumentação embasada em literatura científica; e por meio da verificação da possibilidade de busca pelos identificadores em uma amostra de fontes de informação científica.

É pertinente ressaltar que os identificadores e identificadores persistentes são altamente dinâmicos e alvo de constantes mudanças de versões e otimizações em sua estrutura. Dessa forma, todas as apontamentos e observações realizadas e as ferramentas ou plataformas

escolhidas como objeto de estudo estão contextualizadas de acordo com a data de sua realização, e que isso deve ser considerado na leitura do texto e utilização futura dessas informações sobre a temática em outras pesquisas.

5.1 Iniciativas

A análise das iniciativas terá como base os tópicos e descrições do Quadro 6. As características analisadas nem sempre serão aplicáveis a todos os identificadores persistentes, sobretudo ao considerar que a análise contempla identificadores autorais e organizacionais, que embora possuam características gerais em comum, possuem também especificidades dos recursos que representam.

É fundamental destacar que a análise se concentra apenas nas características que são passíveis de verificação por meio de observação de utilização, sem análise profunda de aspectos técnicos; e todas as observações encontradas a respeito das características serão realizadas ao longo do detalhamento dos identificadores, nos próximos subtópicos, sem pontuar de forma literal cada um dos itens.

Quadro 6 – Características dos identificadores persistentes para sua análise de aplicabilidade na organização e acesso à informação científica

Característica	Descrição
Univocidade	Identificação apenas com uma única URI.
<i>Web-resolvable</i>	Direcionamento para o local exato em que as informações sobre o recurso são acessadas.
Atribuição livre	Atribuição sem nenhum custo às entidades que o aderem.
Acesso e uso abertos	Acesso realizado livremente por qualquer pessoa de qualquer local.
Interface amigável	Possibilitar a fácil identificação do pesquisador por meio de nome, área de pesquisa e/instituição afiliada e do conjunto de itens de pesquisa atrelados ao perfil.
Rede acadêmica	Oferecer ferramentas que favorecem o relacionamento e troca de experiências ou estimula o contato entre pares.
Inserção de outros identificadores	Possibilitar que o pesquisador insira em seu perfil o seu ID de outro identificador.

Integração de dados entre identificadores	Possibilitar o intercâmbio de registros entre identificadores, otimizando o tempo do pesquisador no preenchimento e manutenção das plataformas.
---	---

Fonte: elaborado pela autora.

Serão analisados o cumprimento básico de aspectos técnicos, aqueles que foram abordados em discussão anterior, e todas as demais características necessárias e desejáveis para que eles cumpram seu papel de maneira assertiva. Também será verificada a presença de opções de busca em fontes de informação científica como repositórios de dados de pesquisa, repositórios de *e-prints* e em bases de dados.

5.1.1 International Standard Name Identifier (ISNI)

O ISNI, uma iniciativa sem fins lucrativos, é certificado pela ISO 27729 “*Information and documentation — International standard name identifier (ISNI)*” e mantido pela *ISNI International Agency* (ISNI-IA), que foi criada em 2010 com o apoio de membros como *Conference of European National Librarians* (CENL), *International Confederation of Societies of Authors and Composers* (ICSAC), *International Federation of Reproduction Rights Organizations* (IFRRO), OCLC, ProQuest e *Societies' Council for the Collective Management of Performers' Rights* (SCAPR). Possui a proposta identificar por meio de um código numérico organizações e pessoas, não necessariamente ligadas à comunidade científica, diferente dos demais identificadores abordados nesse trabalho.

Atualmente, o ISNI possui mais de 12 milhões de registros e mais de 104 fontes informacionais, com o objetivo de atribuir identificação a pessoas que criem conteúdo, como pesquisadores, artistas, escritores, inventores, produtores, editores etc. (ISNI INTERNATIONAL AGENCY, 2021c). De acordo com documento publicado pela OCLC, o “ISNI é o identificador mais utilizado para organizações e é disseminado através do Wikidata (a central de armazenamento de dados estruturados de todas as Wikipédias), o VIAF e outros sistemas de identificação como o ORCID” (SMITH-YOSHIMURA et al, 2016, p. 8, tradução nossa). O VIAF é uma fonte informacional de destaque para o ISNI, considerando seus milhares de registros de autoridade de origem internacional.

O ISNI não pode ser administrado como os demais identificadores disponíveis, como o ISBN, o qual as editoras fazem sua atribuição, mas o ISNI deve ter controlado de forma

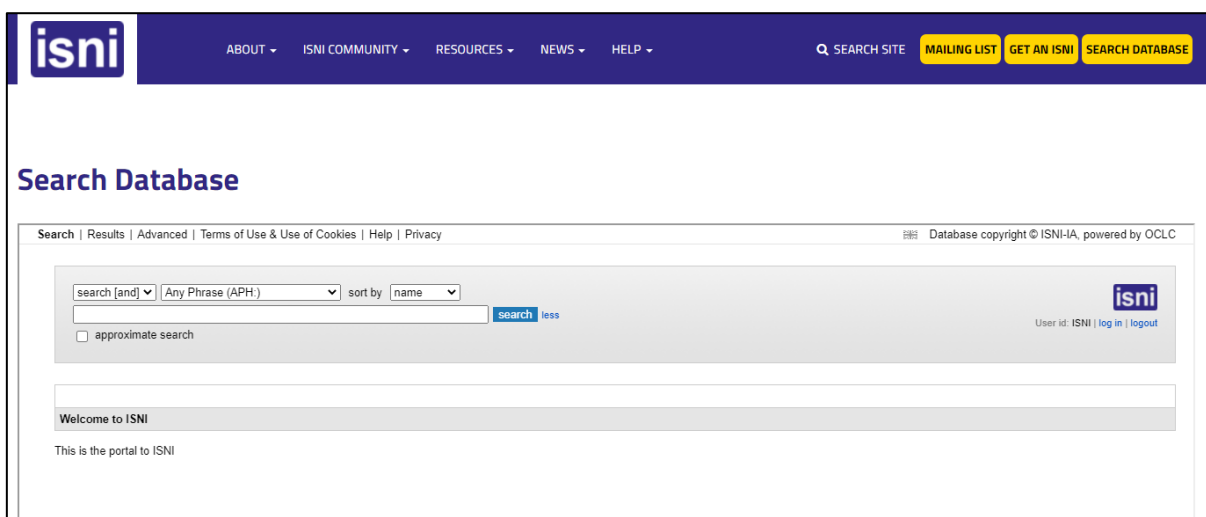
centralizada, considerando que uma mesma pessoa pode possuir diferentes atuações enquanto criador de conteúdo, por exemplo, um cantor que também é compositor e escreve poemas ou livros infantis; o que possivelmente faria com que uma mesma pessoa recebesse diferentes atribuições por diferentes organizações em diferentes segmentos de sua atuação (GATENBY; MACEWAN, 2011).

Como já visto, não é uma tarefa fácil realizar e controlar a correta identificação de pessoas, independentemente de sua atuação, seja por meio do uso de controle de autoridades ou garantindo o correto uso de identificadores. Nesse sentido, a tarefa de coletar e administrar dados, garantindo sua desambiguação assertiva, pode ser realizada de maneira cooperativa entre organizações, como é o caso das Agências de Registro (AG) do ISNI, responsáveis por coletar metadados com alta qualidade (GATENBY; MACEWAN, 2011).

Não é possível criar um registro no ISNI de forma rápida e prática, como acontece com outros identificadores, apenas acessando seu site na *web* e inserindo uma série de informações básicas. Para criar um ISNI, é necessário que a pessoa ou organização interessada se certifique, consultado a base do ISNI, de que não há registro em seu nome fornecido por organizações parceiras e, caso negativo, tenha comprovantes de sua atuação enquanto criador e entre em contato com uma das AGs: Bibliothèque et Archives nationales du Québec (BAnQ), Bibliothèque Nationale de France, British Library, Identification Agency, National Assembly Library of Korea, Numerical Gurus, Quansic, Ringgold e Sound Credit (ISNI INTERNATIONAL AGENCY, 2021b).

Um aspecto que deve ser mencionado a respeito do ISNI, é que ele permite que organizações externas sejam Membros ISNI ou Agências de Registro mediante uma taxa de assinatura anual, e as duas opções oferecem acesso à base de dados e API para pesquisa e atribuição de ISNI, mas apenas as organizações que se unirem à ISNI como Agências de Registro podem atribuir registros também à clientes externos à sua própria organização, com permissão para, inclusive, cobrar uma taxa pela realização desses serviços, para custear gastos e manter modelos de negócio (ISNI INTERNATIONAL AGENCY, 2021c).

O base de busca de registros no ISNI é de acesso livre, sem que seja necessário nenhum tipo de login ou autorização prévia (Figura 9). É possível escolher filtros para realização da busca, considerando o uso de operadores *booleanos*, incluindo uma opção para buscar diretamente pelo número ISNI, caso o usuário o saiba antes da pesquisa.

Figura 9 – Base de busca de registros no ISNI

The screenshot shows the ISNI Search Database interface. At the top, there is a dark blue navigation bar with the ISNI logo on the left and menu items: ABOUT, ISNI COMMUNITY, RESOURCES, NEWS, HELP. On the right side of the bar are links for SEARCH SITE, MAILING LIST, GET AN ISNI, and SEARCH DATABASE. Below the navigation bar, the main content area is titled "Search Database". It features a search form with a dropdown menu for "search [and]", a text input field for "Any Phrase (APH)", and a dropdown for "sort by name". A "search" button and a "less" link are also present. There is an unchecked checkbox for "approximate search". In the top right corner of the search area, there is a small ISNI logo and the text "User id: ISNI | log in | logout". Below the search form, there is a "Welcome to ISNI" message and the text "This is the portal to ISNI". At the top right of the search area, it says "Database copyright © ISNI-IA, powered by OCLC".

Fonte: ISNI INTERNATIONAL AGENCY, 2021.

Os metadados essenciais do ISNI utilizados para a identificação de pessoas ou organizações que constam em sua base de dados são: o nome da identidade pública, variações de seu nome, classe ou classes de criação, função ou funções de criação e URI (GATENBY; MACEWAN, 2011).

Ao realizar a busca por um nome de criador de conteúdo, por exemplo, é possível identificar inicialmente se existem mais de uma opção ou entrada para aquela chave de busca escolhida. Na Figura 10 é possível verificar que, no caso da busca realizada por “Roberto Carlos”, tem-se duas respostas fornecidas pela base de dados. O usuário então pode selecionar aquela que deseja visualizar mais detalhes que o ajudem a identificar corretamente a pessoa que procura, evidenciando a importância dos metadados complementares, considerado que a apenas o número identificador não é suficiente nesse momento de busca.

Figura 10 – Busca realizada na base do ISNI

The screenshot displays the ISNI Search Database interface. At the top, there is a navigation bar with the ISNI logo and menu items: ABOUT, ISNI COMMUNITY, RESOURCES, NEWS, and HELP. On the right side of the navigation bar, there are links for SEARCH SITE, MAILING LIST, GET AN ISNI, and SEARCH DATABASE. Below the navigation bar, the main heading is "Search Database".

The search interface includes a search bar with the text "Roberto Carlos" and a "search" button. There are also options for "search [and]", "Any Phrase (APH)", and "sort by name". A "User id: ISNI | log in | logout" link is visible in the top right corner of the search area.

Below the search bar, there is a "shortlist" section with tabs for "title data" and "search history". The "results" section shows "search [and] (Any Phrase (APH)) Roberto Carlos | 2 hits".

The search results are listed as follows:

1. [Braga, Roberto Carlos \(1941-\)](#)
Seu corpo / Café da manhã / Os seus botões / Falando sério / O côncavo e o convexo / Eu e ela
2. [Roberto Carlos da Silva \(1973-\)](#)
Orfeão da Feira cem anos de história em imagens, 1911-2011

On the left side of the results, there is a "filter" section with options for "Type of identity", "Creation class", and "Creation role". The "Creation class" filter is currently set to "roberto carlos [2]".

Fonte: ISNI INTERNATIONAL AGENCY, 2021.

Ao escolher um dos itens fornecidos pela base de dados, é possível verificar detalhes a respeito do registro e um conjunto de informações que auxiliem no processo de identificação assertiva como (Figura 11): URI; nome e suas variações; data de nascimento ou fundação, esta última no caso de organizações; classe de criação como, por exemplo, a musical; função de criação exercida como, por exemplo, composição e performance; nomes relacionados, que podem auxiliar na identificação do contexto no qual aquela autoridade está envolvida; títulos como, por exemplo, seus discos e composições; contribuições realizadas; e notas como, por exemplo, as fontes das quais essas informações foram coletadas. No caso de organizações registradas no ISNI, outras informações como a localidade ou nacionalidade são igualmente relevantes para sua correta identificação.

Figura 11 – Resultado de busca na base do ISNI

Fonte: ISNI INTERNATIONAL AGENCY, 2021.

A URI do ISNI, presente em cada um dos registros de entidade juntamente com o código identificador de 16 dígitos, é persistente e é exibida no seguinte formato “<https://isni.org/isni/0000000081873460>”, sendo que uma vez inserida em um navegador *web* direciona para o registro da entidade e seu conjunto de metadados.

Embora o ISNI tenha organizações especializadas como fontes informacionais, erros ainda podem acontecer e duas entradas serem encontradas para um mesmo criador de conteúdo, seja ele uma pessoa ou instituição. Uma vez que a informação já foi disponibilizada, é difícil corrigi-la, sendo necessário entrar em contato com o ISNI para que um dos registros duplicados seja descontinuado e relacionado ao registro original (GATENBY; MACEWAN, 2011), ou seja, aquele que deve ser mantido como padrão para identificação da entidade.

Outra opção de busca e metadado presente nos registros do ISNI, é o identificador autoral ORCID, que é voltado para pesquisadores e que será abordado em detalhes em um próximo tópico deste trabalho. Considerando que o ISNI, embora não seja voltado exclusivamente para o registro de entidades ligadas de alguma forma à comunidade científica, possui perfis de pesquisadores e de instituições de pesquisa, por isso faz sentido a presença do identificador autoral. Ao realizar uma busca-teste no ISNI, utilizando apenas como chave de busca o número 0 no filtro para pesquisa do ORCID, com o objetivo de que um índice de opções de IDs fosse recuperado, é possível verificar a forma com a qual é exibido no registro do ISNI

(Figura 12). A URI ORCID presente no registro direciona para o perfil do pesquisador na página *web* do identificador.

Figura 12 – Identificador ORCID atrelado ao registro ISNI

The screenshot displays the ISNI Search Database interface. At the top, there is a navigation bar with the ISNI logo and links for ABOUT, ISNI COMMUNITY, RESOURCES, NEWS, and HELP. On the right, there are buttons for SEARCH SITE, MAILING LIST, GET AN ISNI, and SEARCH DATABASE. The main content area is titled "Search Database" and includes a search bar with the query "ORCID Identifier (ORCID:)" and a "search" button. Below the search bar, there is a table with one result. The result details are as follows:

shortlist	title data	search history	
	results search (ORCID Identifier (ORCID:)) 0000000150992093 1 hits		1

Below the table, there are tabs for "labels", "sources data", and "marc21". A yellow box on the left contains a message: "Please help us improve this record. If you have any supplemental information about the identity listed here, please click in this box to go to the contribution form. Thank you in advance!". The main details for the record are:

- ISNI: 0000 0004 6350 0109
<https://isni.org/isni/0000000463500109>
- Name: Jiang, Shouyong
- ORCID: orcid.org/0000-0001-5099-2093
- Creation class: text
- Creation role: author
- Related names: De Montfort University (isAffiliatedWith)
- Titles: Evolutionary algorithms for static and dynamic multiobjective optimization
- Notes: <http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.716183>
- Sources: ETHOS

Fonte: ISNI INTERNATIONAL AGENCY, 2021.

Observa-se que dos metadados do registro no ISNI, ainda conforme mostra a Figura 12, somente o ORCID, notas e fontes estão diretamente vinculados a sites externos. Dentre os demais itens como, no caso do exemplo acima, o nome da universidade a qual o pesquisador que o registro está identificando não possui nenhum *link* para o registro ISNI dessa instituição, considerando que ela possui um registro. O mesmo poderia acontecer com o título, caso este seja identificado nos metadados da fonte informacional também com um identificador como, por exemplo, o DOI. Considerando o objetivo dos identificadores persistentes quanto à sua interoperabilidade de dados, é fundamental questionar se essas conexões entre os conteúdos dispostos nas plataformas poderiam otimizar a recuperação de outros que estão a eles relacionados.

Conforme brevemente mencionado, o ISNI possui um relacionamento com o identificador autoral ORCID e possui como uma de suas fontes informacionais o identificador Ringgold. A união de esforços e a o intercâmbio de dados é fundamental para que iniciativas distintas, como as mencionadas neste caso, possam contribuir entre si e oferecer um sistema de comunicação científica articulado e que ofereça dados de identificação com qualidade.

5.1.2 Ringgold Identifier

O Ringgold Identifier (Ringgold ID) é um identificador voltado para organizações acadêmicas criado em 2003 e constituído por dígitos numéricos, normalmente entre 4 e 6 dígitos no total. É uma Agência de Registro do ISNI, representando essa última para registro das organizações de maneira a possibilitar que essas informações sejam intercambiáveis.

O formato do identificador é “Ringgold ID 396788” e não se trata de um serviço gratuito. Atualmente é possível acessar a base online de identificadores gratuitamente mediante uma inscrição como convidado, mas que permite apenas sete dias de acesso, o limite de dez pesquisas diárias, o máximo de 10 resultados por pesquisa e visualização restrita apenas do ID do registro juntamente com o seu identificador ISNI (RINGGOLD, 2021b).



Contudo, a Ringgold não possui como serviço oferecido apenas o Ringgold ID, pois existem uma gama de outros serviços que afirmam contribuir com a correta identificação de organizações de pesquisa. O serviço que está diretamente ligado com a correta identificação de instituições de pesquisas, potencializando a organização da informação científica, é o *Identify Database*.

O *Identify Database* é uma base de dados de referência com dados de organizações de pesquisa, a qual se for integrada aos sistemas internos de instituições pode otimizar a normalização de entradas de instituições, oferecendo um conjunto de mais de 25 metadados estruturados, incluindo nome institucional padronizado, nomes alternativos, localidade, URL, tamanho da organização, assuntos, entre outros (RINGGOLD, 2021a). Se trata de um serviço atualmente utilizado por associações de pesquisa, periódicos ou editoras científicas.

Quando o serviço é integrado a sistemas organizacionais internos, possibilita verificação de informações padronizadas a respeito de, por exemplo, seus clientes. Outro exemplo: caso um grupo editorial ou periódico científico utilize o serviço, será possível que ele consiga associar os IDs e toda a gama de metadados de seus clientes, ou seja, aqueles que utilizam seus serviços ou são assinantes, conseguindo retirar relatórios mais precisos a respeito de seus nomes, quem são e suas localidades; e no caso de periódicos científicos, especificamente, será possível normalizar as afiliações institucionais que, muitas vezes, são campos de texto livre e que são suscetíveis a variações de nomenclatura, conforme discutido anteriormente neste trabalho. Os metadados que são disponibilizados por meio do serviço constam na Figura 13:

Figura 13 – Exemplo do conjunto de informações de um registro Ringgold

Organization: University of California Berkeley
Ringgold ID: 1438

Optional Fields ▾  View Hierarchy ?  Send Feedback  Download

Location Information

Field	Value
City	Berkeley
Admin Area Level 1	California
Admin Area Level 2	Alameda County
Post Code	94720
Region	North America
Country	United States

Details

Field	Value
Type	academic
Self Description	University
Reach	Regional
Activity Status	Active
Governance	Government
Governance	Academia
Economic Model	Non-Profit
Level	0
Number of Sites	Multiple
Ringgold Tier	A4
Size	36142
Staff	1852

External Metadata

Field	Value
ISNI	0000 0001 2181 7878
IPEDS	110635
Open Funder Registry ID	http://dx.doi.org/10.13039/100006978

Carnegie Data

Field	Value
size_set_2005	16
basic_2005	15
size_set_2010	16
basic_2010	15
size_set_2015	16
basic_2015	15

Customer Summary

Name	Count	Agent ID	Agent Name	Quantity
<i>There are no records for this section</i>				

URLs

Type	URL
Primary Domain	www.berkeley.edu
Email Site	calmail.berkeley.edu
Former Site	boathall.berkeley.edu
Former Site	csua.berkeley.edu
Former Site	rjf2.biol.berkeley.edu
Former Site	sims.berkeley.edu
Former Site	socrates.berkeley.edu
Former Site	www.berkeley.edu

Alternate Names

Type	Year	Altname
AB		UC Berkeley

Mission





Mission
Research/Perform
Research/Sponsor
Advocacy and Standards/Political
Advocacy and Standards/Social Aid and Welfare
Advocacy and Standards/Environmental
Teaching/Higher Education
Teaching/Graduate Education
Teaching/Adult Education

Subjects









Subjects

There are no records for this section

Inferred Subjects

Ringgold Id	Inferred Subject
500207	Applied Sciences 
113017	Applied Sciences/Agriculture and Food 
124490	Applied Sciences/Agriculture and Food/Crop Science/Biofuels 
118550	Applied Sciences/Agriculture and Food/Forestry 
224036	Applied Sciences/Bioscience/Bioengineering 
207037	Applied Sciences/Bioscience/Bioinformatics 
207037	Applied Sciences/Bioscience/Genomics 

Consortia Membership

Ringgold ID	Organization		
93275	Califa Group		
33320	California Digital Library		
122433	Center for Research Libraries		
225245	Organization for Tropical Studies		

Consortium Members [Display in Search](#)

Ringgold ID	Organization
<i>There are no records for this section</i>	

Fonte: RINGGOLD, 2021c.

O serviço *Identify Database* possibilita que sejam exibidas hierarquias das instituições com Ringgold ID, considerando que elas podem ser cadastradas, caso ainda não façam parte do banco de dados da plataforma; além de permitir a solicitação de criação de IDs para instituições que porventura não a possuam, mas que os clientes necessitam utilizar em seus sistemas internos, sempre passando por um processo de curadoria e análise por parte da Ringgold (RINGGOLD, 2021c).

Com uma proposta bem diferente dos demais identificadores persistentes, a Ringgold possui um perfil voltado para o comercial, sem oferecer nenhum de seus serviços e dados gratuitamente. Contudo, oferece a possibilidade de integração de dados entre sistemas, conferindo maior assertividade e normalização de informações das instituições, não se configurando apenas como mais um número a ser utilizado e atribuído sem nenhum outro objetivo prático aplicado.

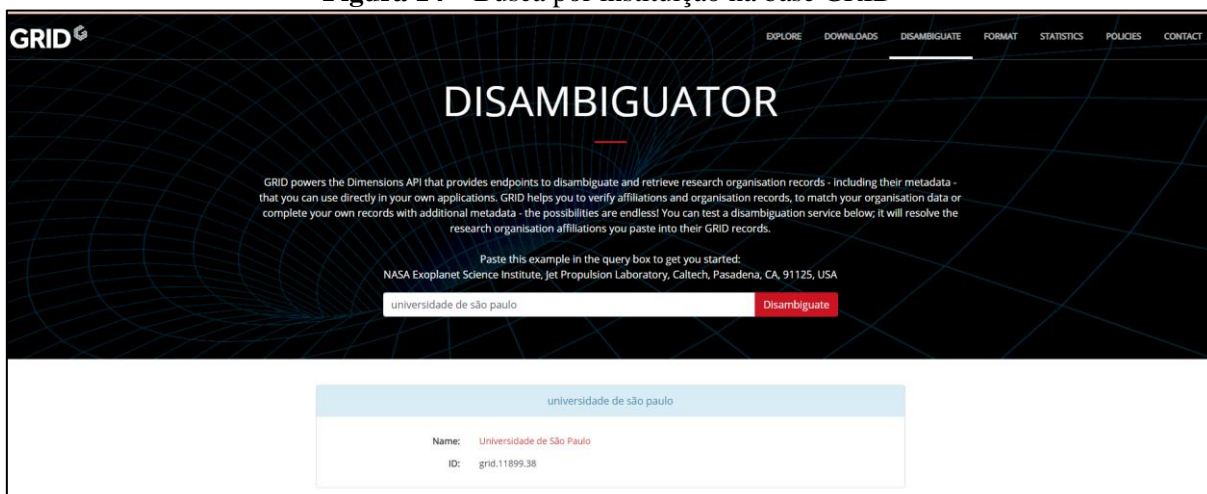
5.1.3 Global Research Identifier Database (GRID)

A *Global Research Identifier Database* (GRID) foi criada em 2015, pela empresa Digital Science, a qual possui sede em Londres e possui diversos produtos relacionados e voltados à pesquisa e comunidade científica. A GRID é uma base de dados que cataloga organizações de todo o mundo gerando identificadores numéricos, com o objetivo de desambiguá-las umas das outras.

Atualmente a base de dados possui mais de 100 mil instituições cadastradas em seu banco de dados, originadas de 219 países e que podem ser instituições de diferentes tipos, como educacionais, voltadas ao cuidado em saúde, empresas, organizações sem fins lucrativos, arquivos, organizações governamentais, entre outras. As instituições são adicionadas na GRID por meio de dados extraídos de afiliações de publicações científicas e dados de fomento concedidos, e os registros das instituições levam em consideração o nome pelo qual a organização é mais conhecida nessas fontes e em seu site *web* (GRID, 2021).

A base de dados pode ser consultada livremente em seu site *web* e a busca pode ser realizada pelo nome da instituição (Figura 14), sendo exibido o resultado da busca juntamente com o número do identificador.

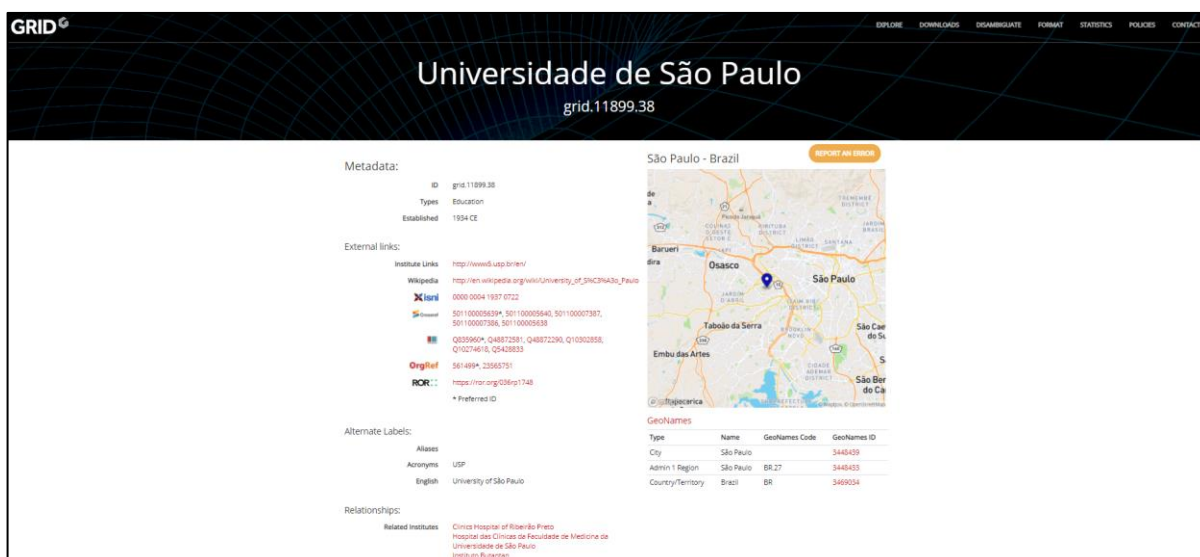
Figura 14 – Busca por instituição na base GRID



Fonte: GRID, 2021.

Ao selecionar a instituição, é possível visualizar o perfil da instituição com diversas informações como (Figura 15): ID GRID; tipo de instituição; coordenadas de geolocalização; *links* para fontes externas que falam sobre a instituição, como o seu próprio site, página na Wikipédia, *link* para o perfil no ISNI, Crossref, OrgRef e *link* para o perfil no *Research Organization Registry* (ROR); acrônimos institucionais ou variações do nome institucional em outro idioma; e instituições relacionadas, geralmente hierarquicamente envolvidas. O conjunto de informações disponibilizado auxiliam no processo de desambiguação de instituições ao mesmo tempo em que ilustra o relacionamento da base de dados e seu identificador com outros identificadores da comunidade científica.

Figura 15 – Perfil da Universidade de São Paulo na base GRID



Fonte: GRID, 2021.

Na página do perfil da instituição na base GRID, é possível visualizar um botão disponibilizado acima das informações de geolocalização para reportar erros. É uma oportunidade que a base de dados oferece para realizar possíveis correções no nome ou em informações ali disponíveis.

Embora na página da instituição na GRID não seja exibida uma URI de referência para acessar aquele perfil, mas apenas o número GRID, a URL disponível possui o formato “https://www.grid.ac/institutes/grid.11899.38”, que ao ser copiada e colada em um navegador *web*, direciona para a página do perfil institucional.

Além de disponibilizar abertamente a base de dados para consulta, conforme exemplo acima mencionado, a GRID oferece a possibilidade de *download* da listagem de instituições e de seus respectivos IDs, sendo que essa listagem é atualizada e publicada em seu site quatro vezes ao ano. É interessante mencionar que é atribuído o identificador persistente DOI à listagem para *download* publicada. A listagem é oferecida em formato JSON, TTL e CSV, sendo que este último pode ser visualizado em um programa como o Excel, do Pacote Office (Figura 16).

Figura 16 – Visualização de listagem de instituições da GRID no Excel

ID	Name	City	State	Country
grid.1001.0	Australian National University	Canberra	Australian Capital Territory	Australia
grid.1002.3	Monash University	Melbourne	Victoria	Australia
grid.1003.2	University of Queensland	Brisbane	Queensland	Australia
grid.1004.5	Macquarie University	Sydney	New South Wales	Australia
grid.1005.4	UNSW Sydney	Sydney	New South Wales	Australia
grid.1006.7	Newcastle University	Newcastle upon Tyne		United Kingdom
grid.1007.6	University of Wollongong	Wollongong	New South Wales	Australia
grid.1008.9	University of Melbourne	Melbourne	Victoria	Australia
grid.1009.8	University of Tasmania	Hobart	Tasmania	Australia
grid.1010.0	University of Adelaide	Adelaide	South Australia	Australia
grid.1011.1	James Cook University	Townsville	Queensland	Australia
grid.1012.2	University of Western Australia	Perth	Western Australia	Australia
grid.1013.3	University of Sydney	Sydney	New South Wales	Australia
grid.1014.4	Flinders University	Adelaide	South Australia	Australia
grid.1016.6	Commonwealth Scientific and Industrial Research Organisation	Canberra	Australian Capital Territory	Australia
grid.1017.7	RMIT University	Melbourne	Victoria	Australia
grid.1018.8	La Trobe University	Melbourne	Victoria	Australia
grid.1019.9	Victoria University	Melbourne	Victoria	Australia
grid.1020.3	University of New England	Armidale	New South Wales	Australia
grid.1021.2	Deakin University	Geelong	Victoria	Australia
grid.1022.1	Griffith University	Brisbane	Queensland	Australia
grid.1023.0	Central Queensland University	Rockhampton	Queensland	Australia
grid.1024.7	Queensland University of Technology	Brisbane	Queensland	Australia
grid.1025.6	Murdoch University	Perth	Western Australia	Australia
grid.1026.5	University of South Australia	Adelaide	South Australia	Australia
grid.1027.4	Swinburne University of Technology	Melbourne	Victoria	Australia
grid.1029.a	Western Sydney University	Richmond	New South Wales	Australia
grid.1031.3	Southern Cross University	Lismore	New South Wales	Australia
grid.1032.0	Curtin University	Perth	Western Australia	Australia
grid.1033.1	Bond University	Gold Coast	Queensland	Australia
grid.1034.6	University of the Sunshine Coast	Sunshine Coast	Queensland	Australia
grid.1035.7	Warsaw University of Technology	Warsaw		Poland
grid.1037.5	Charles Sturt University	Bathurst	New South Wales	Australia
grid.1038.a	Edith Cowan University	Perth	Western Australia	Australia
grid.1039.b	University of Canberra	Canberra	Australian Capital Territory	Australia
grid.1040.5	Federation University	Ballarat	Victoria	Australia

Fonte: elaborado pela autora.

Em 2019 foi criado o ROR, que também é um identificador organizacional, espelhando os dados da base GRID. Nesse contexto, a Digital Science decidiu que irá descontinuar a base GRID no último semestre de 2021, para que o ROR ocupe seu espaço. A Digital Science é conhecida por diversas iniciativas e serviços relacionados à pesquisa científica que tem liderado nos últimos anos como, por exemplo, o Altmetrics, que gera métricas alternativas, e a base de dados Dimensions.

O ROR surgiu da ideia de criação de um identificador organizacional do qual toda a comunidade científica pudesse se beneficiar. A iniciativa contou com a colaboração de diferentes instituições, dentre elas a Crossref, ORCID, DataCite e a Digital Science, que concedeu os dados da GRID para implementação no projeto.

O ROR possui o objetivo de fornecer dados abertos e gratuitos, que não podem ser adquiridos e, mesmo que ofereçam serviços pagos futuramente, os dados continuarão livremente abertos para serem utilizados no desenvolvimento de projetos em benefício à comunidade científica (ROR, 2021).

A busca por instituições pode ser feita livremente no ROR e os resultados de busca são exibidos juntamente com informações como nome, nome no idioma inglês, site, país, tipo de organização e outros identificadores institucionais na base GRID, ISNI, Crossref e Wikidata (Figura 17).

Figura 17 – Busca por instituições no ROR

The screenshot shows the ROR website interface. At the top left is the ROR logo. To its right is a search bar with the placeholder text 'Search Registry...'. Further right are navigation links: ABOUT, SCOPE, FACTS, SUPPORTERS, RESOURCES, and BLOG. Below the search bar, it displays '6,633 Organizations'. A search result is highlighted in a white box with a light gray border. The result shows the ROR URI: <https://ror.org/036rp1748>. Below the URI, the organization name is 'Universidade de São Paulo' followed by 'USP, UNIVERSITY OF SÃO PAULO'. Underneath, there are two columns of information: 'WEBSITE' with the URL 'http://www5.usp.br/en/' and 'OTHER IDENTIFIERS' which lists 'GRID grid.11899.38', 'ISNI 0000000419370722', 'Crossref Funder ID 501100005639', and 'Wikidata Q835960'. At the bottom of the result box, there are two colored tags: 'BRAZIL' (blue) and 'EDUCATION' (orange).

Fonte: ROR, 2021.

O registro da instituição também exibe a URI do identificador persistente, que pode ser colada em um navegador *web* que direciona para o perfil da organização no identificador.

Ao considerar que a maior parte das informações de afiliação autoral disponíveis em publicações na comunidade científica são fornecidas pelos próprios autores no momento de submissão ao periódico, geralmente em campo de texto livre, é preciso que as iniciativas de identificadores organizacionais reflitam sobre as possibilidades de tornar esse processo mais assertivo e que possibilite uma saída ou disponibilização padronizada dessas informações nos artigos publicados. Se, porventura, a informação de afiliação não for inserida nos metadados da publicação, o item será indexado em fontes informacionais sem esse conteúdo, impossibilitando sua recuperação por ele.

Nesse sentido, o ROR espera que, assim que a Crossref, agência oficial de registro dos DOIs, começar a aceitar o ROR nos metadados do editor, esses dados poderão ser utilizados pelos próprios editores para entender a origem de seus autores e para as instituições, que poderão verificar assertivamente onde seus autores estão publicando (LAMMEY, 2020). Atualmente, essa verificação pode ser realizada, mas de maneira manual e/ou menos assertiva, ou seja, sujeito aos erros e variações de nomenclatura do nome da instituição.

A Crossref já mencionou intenção de adotar os identificadores organizacionais como uma alternativa para garantir qualidade em seus metadados:

Estamos particularmente interessados em oferecer suporte a identificadores abertos e liderados pela comunidade, como ORCID e ROR, mas também queremos oferecer suporte aos metadados que nossos membros desejam distribuir. Os identificadores organizacionais serão particularmente úteis pois nos ajudaram a preencher os registros com ROR IDs no futuro, gerando metadados de afiliação de melhor qualidade [...] atualmente coletamos a afiliação como uma única *string* – nós vamos separar isso para oferecer suporte às afiliações e adicionar suporte para identificadores organizacionais como o ROR (FENNEY, 2019, [s. p.], tradução nossa).

Lammey (2020) menciona que algumas organizações já incluíram o ROR em seus *workflows* e menciona especificamente o caso de um repositório de dados de pesquisa chamado Dryad, no qual o autor no momento de depositar seus dados e consequentemente inserir sua afiliação institucional, um menu *dropdown* é exibido com opções de organizações vindas do ROR para a escolha do autor; e, por sua vez, o repositório registra DOIs de seus dados de pesquisa via DataCite, fornecendo juntamente os ROR IDs junto com os demais metadados, fazendo com que a pesquisa por identificadores organizacionais seja possível via busca no DataCite.

A mudança definitiva de todas as informações da base GRID para o ROR, que acontecerá ainda em 2021, pode subsidiar novas análises e discussões a respeito da aplicabilidade desse identificador organizacional na organização da informação científica. Assim como foi perceptível em outros identificadores, para que o identificador tenha uma utilidade prática, é essencial que ele seja utilizado por outros membros e organizações da comunidade científica e possibilite a interoperabilidade de dados, sem figurar apenas como mais um número disponível que representa instituições. A base GRID e ROR possuem uma interface clara de informações e buscam conectar a organização identificada a outros identificadores delas disponíveis.

5.1.4 ORCID

O identificador *Open Researcher and Contributor ID* (ORCID), foi criado em 2012 por uma organização sem fins lucrativos de mesmo nome. O projeto em si teve início em 2009 com Thompson Reuters e o grupo editorial Nature à frente dele, engajados criar um identificador que pudesse interagir com outros sistemas acadêmico-científicos.

Atualmente com mais de 11 milhões de perfis (SILVA, 2021), se trata de um identificador numérico de dezesseis dígitos que é capaz de diferenciar pesquisadores, o qual é aplicado em uma URI e possibilita que o usuário consiga, através deste *link*, acessar diretamente seu perfil ou o perfil de outros pesquisadores por este mesmo endereço. A URI do ORCID possui o seguinte formato, alterando apenas o número de acordo com o pesquisador: <https://orcid.org/0000-0003-3644-8600>. O identificador também oferece a opção de gerar um *QR Code* com o *link* para o perfil do pesquisador, de modo que este poderá utilizar em sites, blogs ou apresentações ao longo da trajetória acadêmica ao invés de inserir o número ORCID propriamente dito.

O número uma vez atribuído a um pesquisador, não é atribuído a nenhum outro, garantindo sua univocidade. O ORCID também é um identificador persistente, considerando que não deve mudar, garantindo que o número que esteja disponível em um recurso estático ainda possa ser buscado na *web* e redirecionado para o recurso ao qual originalmente representa.

O identificador também possibilita uma visualização direta da página onde se encontram todas as informações a respeito do recurso, que neste caso é o perfil do autor, e é capaz de acessar uma página com metadados, se necessário, em um formato que seja legível por máquinas e não-proprietário.

Com o objetivo de se tornar interoperável e referência enquanto identificador autoral

persistente, uma vez que não possui declarado vínculo com empresas que possuem interesse direto por determinadas bases de dados ou grupos editoriais, como no caso de outros identificadores existentes, o ORCID tem sido adotado mundialmente desde seu surgimento, propondo a criação gratuita do perfil no qual o pesquisador pode inserir os detalhes de sua trajetória acadêmica e corporativa, sendo exigido por diversas revistas nacionais e internacionais como requisito para submissão de publicações ou agências de fomento, para submissão de projetos.

A ORCID é gerenciada por dezesseis diretores que são membros da comunidade científica, que possuem o objetivo de conferir transparência à comunidade acadêmica (CRESS, 2019). O ponto de partida do ORCID foi o *ResearcherID*, identificador que na época era atrelado à Thompson Reuters, considerando que esta empresa cedeu o software para sua implementação (GARCÍA-GÓMEZ, 2012).

O perfil do pesquisador no ORCID oferece condições para inserção de diversas informações a respeito da trajetória acadêmica e profissional, incluindo uma pequena biografia, nome preferido e utilizado nas publicações científicas, espaço para inserir *link* para perfis em outros identificadores, *link* para sites ou *blogs*, *e-mails* para contato e palavras-chave que, por exemplo, representem sua linha de pesquisa (Figura 18).

Figura 18 – Exemplo de perfil do pesquisador no ORCID

Thamyres Vieira dos Santos

ORCID iD
<https://orcid.org/0000-0003-3644-8600>
 Visualizar a versão pública

Gerenciar Conta Proxy

Exiba seu iD em outros sites
 Visualização de impressão do registro público
 Obtenha um Código QR para seu iD

Também conhecido como
 Santos TV

Pais
 Brasil

Palavras-Chave
 informação científica, comunicação científica, identificadores autorais, identificadores organizacionais, informação em saúde, divulgação científica, estudos métricos da informação, bibliometria, cientometria

Websites
 Currículo Lattes

Outras IDs
 ResearcherID: Q-3719-2017
 Loop profile: 1390142
 Scopus Author ID: 57224725549

E-mails

Biografia
 Possui graduação em Biblioteconomia pela Universidade de São Paulo (2014), especialização em Teorias e Práticas da Comunicação pela Faculdade Cásper Líbero (2016), especialização em Gestão de Serviços de Informação pela Fundação Escola de Sociologia e Política de São Paulo (2019), mestrado em Ciência da Informação pela Universidade de São Paulo. Bibliotecária no Hospital Israelita Albert Einstein atuando especificamente com os indicadores de Produção Científica Institucional. Interesse pelos temas: comunicação científica, indicadores bibliométricos, métricas alternativas, informação científica, divulgação científica, organização da informação, informação em saúde, identificadores autorais e organizacionais.

Emprego (3)
Educação e qualificações (4)
Posições convidadas e distinções (0)
Adesão e serviço (0)
Financiamento (0)
Trabalhos (11 of 11)

Fonte: ORCID, 2021.

Atualmente é possível inserir informações a respeito de seis aspectos da trajetória acadêmica e corporativa dos pesquisadores: **emprego**, para informações de atuação profissional; **educação e qualificações**, para histórico acadêmico, incluindo cursos de especialização; **posições convidadas e distinções**, para posições convidadas como, professor emérito, ou prêmios recebidos; **adesão e serviço**, para membresia de associações ou serviços prestados; **financiamento**, para fomentos de pesquisa recebidos; e **trabalhos publicados**, para toda a produção intelectual do pesquisador.

Todas essas informações podem ser manualmente inseridas pelo pesquisador, mas algumas delas possibilitam a interoperabilidade também com outras fontes de informação, ou seja, é possível que o pesquisador importe previamente suas informações sem a necessidade de realizar um preenchimento manual.

ORCID fornece uma interface de programação de aplicativo (API) que permite a troca de informações entre sistemas de comunicação científica assim como sistemas de submissão de manuscritos dos editores, repositórios institucionais, sistemas de rastreamento de fomento e sistemas de informação do pesquisador. A API destina-se a reduzir a redigitação de informações, melhorar os dados compartilhados entre sistemas e facilitam a manutenção dos registros (THOMAS; CHEN; CLEMENT, 2015, p. 333).

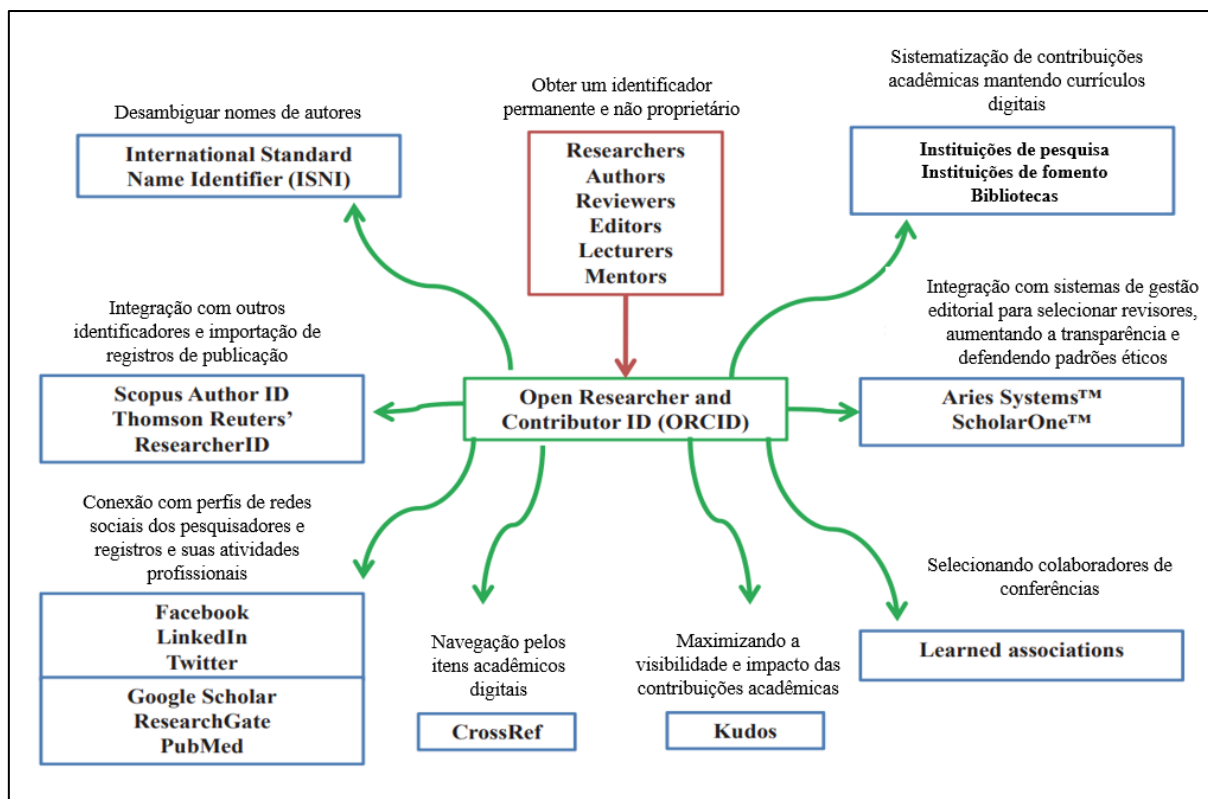
As APIs da ORCID permitem o intercâmbio de informações entre sistemas de comunicação científica e o identificador de maneira que, mediante autorização do pesquisador, dados podem ser trocados entre as duas partes. Por exemplo, caso um pesquisador faça a vinculação de seu identificador com a instituição, esta pode receber notificações em seu repositório institucional ou outro sistema de gestão de publicações, sempre que um novo artigo for adicionado em seu perfil e vice-versa, a instituição também pode inserir publicações em seu perfil, validando-as.

São diversos os identificadores de diferentes recursos de pesquisa utilizados na comunidade científica e existem diferentes formas de conseguir conectar as informações entre eles, de maneira a tornar o processo de comunicação e troca interoperável, basta que as necessidades da comunidade se concentrem em integrar todos esses identificadores em *workflows* e permitir esse compartilhamento, em benefício de todos (HAAK, 2014).

Na Figura 19, é possível observar as conexões que podem ser realizadas entre o ORCID e outras fontes de informação, contribuindo para que o identificador seja povoado por conteúdos já armazenados por terceiros e que possam ser imputados no registro, otimizando o tempo do

pesquisador. É fundamental pontuar que nem todas as conexões oferecem necessariamente uma integração de dados, ou seja, migrar o que está em uma plataforma para a outra. Em algumas delas, apenas é possível inserir *links* para perfis em plataformas externas.

Figura 19 – Integrações do ORCID com redes, bases de dados e plataformas de comunicação científica



Fonte: adaptado de GASPARYAN; AKAZHANOV; VORONOV; KITAS, 2014, p. 1455.

As integrações e possibilidades de conexão são úteis no momento do pesquisador inserir sua produção científica no identificador, sendo que este pode inserir manualmente ou via integrações, que atualmente são permitidas com as seguintes plataformas: Airti, ArXiv, Bielefeld Academic Search Engine, CrossRef, DataCite, Deutsche Nationalbibliothek (DNB), Europe PubMed Central, Google Scholar (via carga de arquivo em formato BibTex), HAL Université Savoie Mont Blanc, ISNI, Japan Link Center (JaLC), MLA International Bibliography, OpenAIRE Explore, PubMed, Redalyc, Research Data Australia, Scopus, The Lens e US Department of Energy. E para integração de informações de financiamento de pesquisa, é possível realizar uma integração com os dados da plataforma Dimensions Wizard. Essas são as opções atuais, mas sem dúvidas a dinâmica de mudanças em plataformas como o ORCID é intensa, resultando em novas opções ou em retirada de opções atuais ao longo do tempo.

Dentre todas as integrações permitidas, destaca-se também a interoperabilidade de

informações entre Ringgold, ISNI e ORCID. Como anteriormente mencionado, no momento de preencher o perfil no identificador, o pesquisador pode inserir manualmente suas informações de vínculo empregatício e educação. Contudo, no momento de realizar o preenchimento manual das informações, o pesquisador vai se deparar com um menu *dropdown* de opções ao começar a digitar o nome da instituição nesses campos (Figura 20). Esse menu utiliza os dados do Ringgold *Identify Database*, serviço mencionado em tópico anterior desse trabalho.

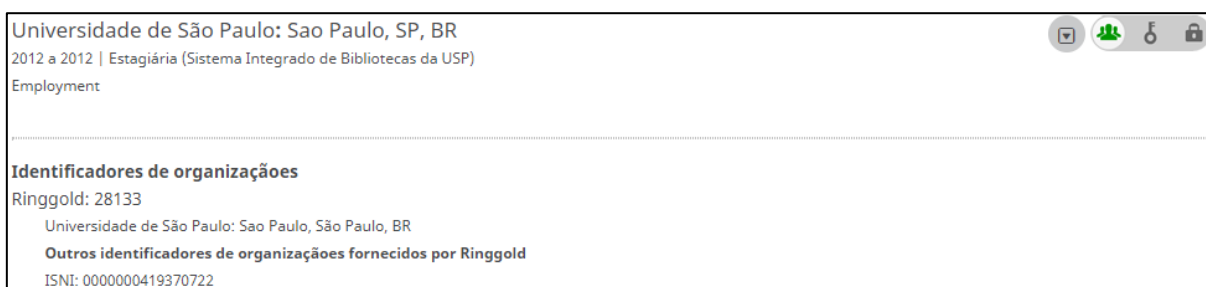
Figura 20 – Menu *dropdown* com organizações no ORCID

A imagem mostra a interface de usuário do ORCID para adicionar um emprego. O formulário é intitulado "ADICIONAR EMPREGO" e contém os seguintes campos e opções:

- Organização*:** Campo de texto com o valor "universidade de são pau" e um menu "dropdown" aberto mostrando sugestões de instituições. As sugestões incluem:
 - Universidade de São Paulo BR
 - Universidade de São Paulo Câmpus de São Carlos, Sao Carlos, São Paulo, BR
 - Universidade de São Paulo, Sao Paulo, São Paulo, BR
 - Universidade São Marcos, São Paulo, BR
 - Universidade de Santo Amaro, Sao Paulo, SP, BR
- Departamento:** Campo de texto com o valor "Adicionar departamento".
- Função/título:** Campo de texto vazio.
- URL:** Campo de texto com o valor "Adicionar URL".
- Data de início:** Campos de seleção para Ano, Mês e Dia.
- Data final (deixar em branco se corrente):** Campos de seleção para Ano, Mês e Dia.
- Set visibility:** Botões para definir a visibilidade (público, privado, restrito).
- Botões:** "Adicionar à lista" e "Cancelar".

Fonte: ORCID, 2021.

Esse serviço oferecido pelo Ringgold e contratado pelo ORCID garante maior assertividade no preenchimento das informações de organização, evitando que aconteçam problemas semelhantes aos que vemos nas afiliações institucionais das publicações: um preenchimento de campo de texto livre, suscetível a conteúdo sem padronização. E o preenchimento automatizado por essa ferramenta garante que a organização seja corretamente identificada quando o registro e a informação do perfil do autor são públicas, sendo exibida a instituição e maiores detalhes a seu respeito, como o Ringgold ID e número ISNI (Figura 21).

Figura 21 – Exemplo de exibição do Ringgold ID e ISNI no registro ORCID

Fonte: ORCID, 2021.

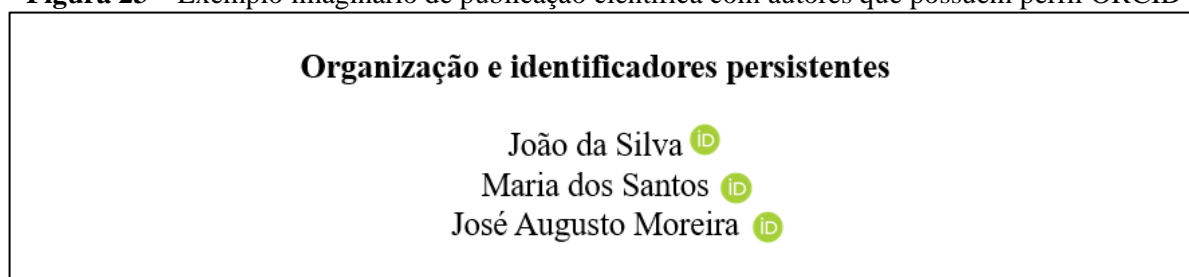
A interoperabilidade de informações entre os identificadores é assertiva uma vez que indica a possibilidade de conexões entre iniciativas, colaborando para otimizar a comunicação e possível benefício mútuo entre as partes e para o pesquisador.

A autorização de acesso tanto das instituições, quando realizam integrações de informações entre os registros ORCID e seus sistemas internos via API, quanto das integrações realizadas, conforme mencionado há pouco, se dá na própria plataforma. O pesquisador precisa aceitar a interoperabilidade de informações entre a plataforma externa e o perfil no identificador (Figura 22).

Figura 22 – Autorização de acesso à plataforma externa no ORCID

Fonte: ORCID, 2021.

Alguns grupos editoriais científicos e periódicos têm adotado iniciativas de inserir o identificador ORCID relacionado aos seus autores não apenas nos metadados das publicações, o que é fundamental, mas também de forma visual nos próprios artigos com o objetivo de identificá-los de maneira assertiva (Figura 23). Considerando que ao referenciar um material em trabalhos acadêmico-científicos é necessária a inserção do *link* da publicação, este *link* pode redirecionar para o periódico que por sua vez possui as informações dos autores, incluindo o perfil no ORCID.

Figura 23 – Exemplo imaginário de publicação científica com autores que possuem perfil ORCID

Fonte: elaborado pela autora

O ORCID, assim como todas as plataformas, possui suas vantagens e as suas limitações (Quadro 8). As vantagens do identificador se centram basicamente em seu formato aberto, sem vínculo direto com grupos editoriais, integrações e facilidade de criar o registro. Dentre as limitações, está sua estrutura, que não permite a inserção de itens audiovisuais ou textos completos dos artigos, ainda que essas opções possam esbarrar em problemas de direito autoral, além de permitir que nas configurações os pesquisadores deixem seus perfis privados ou que não sejam atualizados com frequência, sem nenhum tipo de alerta ou ação da plataforma para sugerir atualizações.

Quadro 7 – Vantagens e limitações do Open Researcher and Contributor ID (ORCID)

Vantagens	Limitações
Serviço não-proprietário para listar várias atividades acadêmicas individuais	Falta de imagens, vídeos e textos completos dos artigos
Facilidade de registro	Perfis fechados ou desatualizados de alguns usuários
Iniciativa aceita globalmente	
Desambigua os nomes dos autores	
Integração com diversas plataformas acadêmicas, bases de dados e redes sociais	
Transparência e validação das atividades acadêmicas	

Fonte: Zimba e Gasparian, 2020, p. 347, tradução nossa.

Considerando a atual exigência de muitos periódicos científicos pela inserção do identificador ORCID no momento de submissão da publicação para avaliação, é comum que pesquisadores se sintam tentados a simplesmente criar um perfil na plataforma para gerar seu ID, já que essa é a informação requisitada pelas revistas, e simplesmente não preencher mais nenhuma informação, sem que a plataforma tenha algum mecanismo que evite esse comportamento. O identificador não foi originalmente criado com a função de um “currículo” ou qualquer outro material do gênero, mas é fundamental que informações básicas do pesquisador sejam disponibilizadas, considerando que, por exemplo, este ID será inserido no artigo depois de publicado e se o perfil estiver sem informações, o usuário que está consultando continuará sem saber de quem se trata, ainda que o artigo em questão possa ser recuperado pelo

ID do autor em uma base de dados em que esteja indexado e que apresente essa opção de busca. O mesmo acontece com as opções de privacidade que, muitas vezes por desconhecimento, são ativadas e nenhuma informação do pesquisador fica disponível para exibição no identificador.

Silva (2021, p. 1, tradução nossa) menciona “a existência de contas ORCID ‘fantasmas’ na qual as informações estão ausentes, perdidas, incompletas ou não confiáveis, reduzindo assim a capacidade de verificar se o pesquisador é de fato esse pesquisador, ou não”.

Além das limitações já mencionadas, é fundamental destacar que a facilidade de criação do registro e falta de algum mecanismo de regulação por parte da plataforma, ocasiona a criação de perfis por qualquer pessoa, sem que este tenha um vínculo acadêmico-científico, por exemplo. Embora o cerne do problema não seja exatamente esse, afinal qualquer pessoa que queira realmente pode criar um perfil, Silva (2021) evidencia que existem casos em que perfis são criados em nomes de pessoas famosas, como Michael Jackson e John Travolta, e até perfis criados com palavras de baixo calão, que não serão aqui exemplificadas.

O ORCID é atualmente um identificador amplamente adotado na comunidade científica e se mostra de fácil utilização, ainda que possua lacunas sobretudo em relação a sua regulação de preenchimento e disponibilização de dados. Se comparado aos identificadores organizacionais, ele se diferencia pela ampla possibilidade de edição e inserção de conteúdos pelo pesquisador no registro.

5.1.5 ResearcherID

O identificador *ResearcherID* foi criado em 2008 e pertence à empresa *Clarivate Analytics*, que detém sobre seu poder a base de dados referencial *Web of Science*, que possui considerável prestígio como fonte para pesquisas acadêmicas e dados bibliométricos para avaliação de produção científica. Em 2019, o *ResearcherID* deixou de existir como uma ferramenta única e foi integrado ao *Publons*, uma ferramenta que também pertence à mesma empresa mencionada, que tem como objetivo reunir perfis de pesquisadores que realizam trabalhos como editores ou revisores de periódicos ou artigos científicos como um todo. As pessoas que eram cadastradas no *ResearcherID* tiveram seu perfil e informações automaticamente integradas ao *Publons*.

O *ResearcherID* é um identificador que permite aos pesquisadores inserirem suas informações acadêmico-científicas e corporativas, funcionando também como uma espécie de currículo, com histórico de pesquisa e atuação do cientista.

Diferentemente do ORCID, esse identificador possui um vínculo mais estreito com a base Web of Science, já que pertencem à mesma empresa, então é possível inserir suas publicações indexadas nessa base de forma mais fácil e obter dela o número de citações para seus trabalhos. Contudo, também é possível realizar a inclusão de publicações que não estão na Web of Science, conferindo autonomia ao pesquisador para construir seu perfil na rede. Se porventura o pesquisador possuir um perfil (Figura 24) nessa base de dados, ou seja, ter publicações científicas que pertencem a periódicos que lá estão indexados, é possível atrelá-lo ao seu identificador via Publons.

Figura 24 – Exemplo de perfil do pesquisador na base de dados Web of Science

The screenshot shows the Web of Science interface for a researcher profile. At the top, there is a navigation bar with 'Web of Science' logo, search options, and user actions like 'Sign In' and 'Register'. Below the navigation, there are buttons for '< Back' and 'Submit a correction'. The main profile section displays the name 'Maimone, Giovana Deliberati' with a note: 'This is an algorithmically generated author record'. Below the name, it lists 'Programa Posgrad Ciencia Informacao' and 'SAO PAULO, SP, BRAZIL'. To the right, there is a section titled 'Are you this Author?' with a 'Claim my record' button. Below that, the 'Author Metrics' section includes an 'Author Impact Beamplot Summary' chart. The chart shows a 'CITATION PERCENTILE' with a scale from 0 to 100, and a green circle indicating a score of approximately 25.

Fonte: Web of Science, 2021.

Além de exibir o número de publicações e citações, o identificador exibe outras informações como o *h-index*, métricas e as atividades de revisão, que estão alinhadas ao principal objetivo do Publons, que é conferir visibilidade também para este tipo de atividade de pesquisa (Figura 25).

Figura 25 – Exemplo de perfil do pesquisador no Publons

The image shows a screenshot of a researcher profile on the Publons website. The profile is for Thamyres Vieira dos Santos, a Librarian at Hospital Israelita Albert Einstein. The profile includes a circular profile picture with the letters 'TS', the name 'Thamyres Vieira dos Santos', and the alias '"Santos TV"'. It also displays the Web of Science ResearcherID (Q-3719-2017) and a table of statistics: PUBLICATIONS (4), TOTAL TIMES CITED (0), and H-INDEX (0). The interface features a navigation menu on the left with options for Summary, Metrics, Publications, and Peer review. Below the profile, there is a 'Research Fields' section with tags for Altmetrics, Bibliometrics, Information Science, Librarianship, Metrics, and Persistent Identifier. A button at the bottom of this section says '+ VIEW ALL INSTITUTIONS & AWARDS'.

Fonte: Publons, 2021.

Contudo, observa-se que, embora o ResearcherID tenha sido integrado ao Publons, o código alfa numérico não é utilizado como URI para o identificador, de forma que este pudesse acessar o perfil diretamente por este endereço em um navegador *web*. O perfil do pesquisador pode ser acessado mediante um endereço *web* sem necessidade de login prévio na plataforma, com o seguinte formato, se adequando de acordo com o nome do pesquisador: “<https://publons.com/researcher/1752506/thamyres-vieira-dos-santos/>”.

Um estudo realizado por Mikki, Zygmuntowska, Gjesdal e Al Ruwehy (2015) para verificar a presença digital de pesquisadores da Universidade de Bergen, na Noruega, nas plataformas ResearchGate, *Academia.edu*, *Google Scholar Citations*, *ResearcherID* e ORCID, concluiu que *ResearcherID* e ORCID possuíam a menor adesão por parte dos pesquisadores, 130 e 108 perfis, respectivamente. Este perfil de uso e adesão já pode ter sido alterado, considerando a evolução e mudanças nas plataformas ao longo do tempo, incluindo a mudança do *ResearcherID* para o Publons.

O *ResearcherID* é um identificador que, embora pertença a um grupo empresarial e que esteja de alguma forma ligado à base de dados Web of Science, o pesquisador possui liberdade para gerenciar seu perfil, incluir publicações de qualquer origem, preencher informações de sua trajetória acadêmica e profissional e, por fim, conferir visibilidade às suas atividades de revisão,

que são pouco conhecidas mas que também mostram a atuação do pesquisador nessa frente inerente ao desenvolvimento científico.

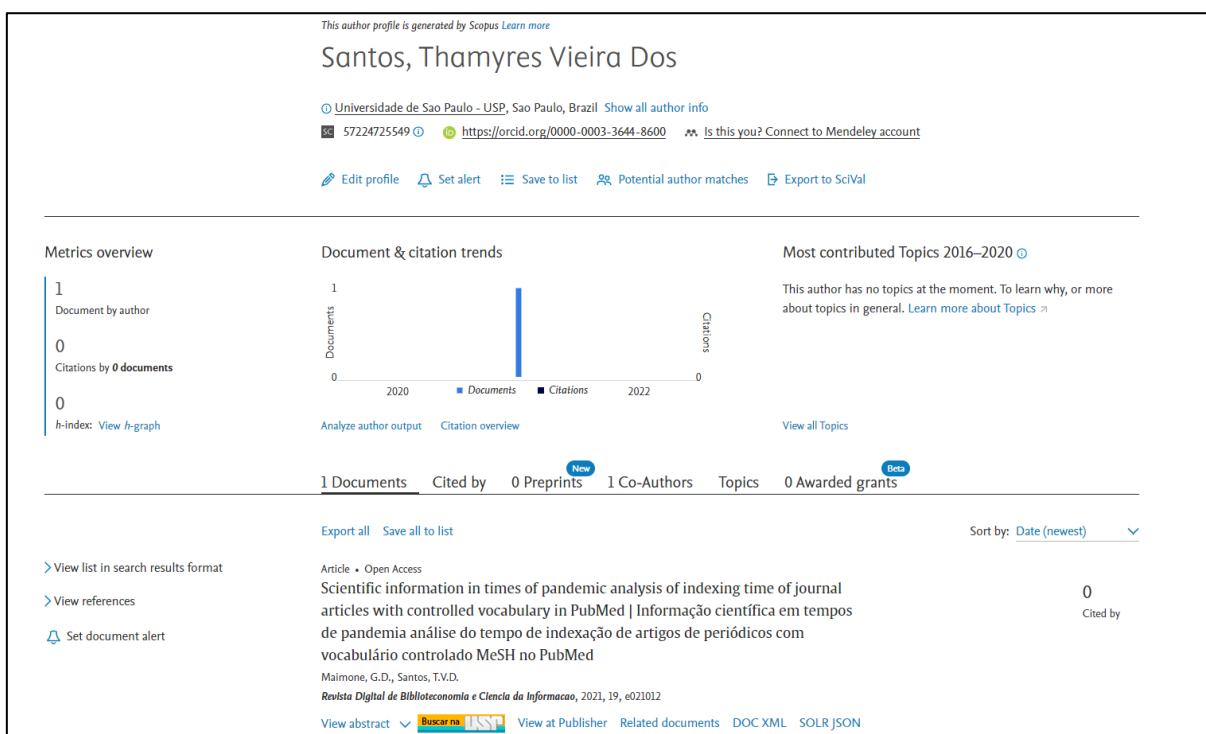
5.1.6 ScopusID

O Scopus ID é um identificador criado em 2006 e atrelado à base de dados Scopus, que pertence ao grupo editorial Elsevier, e diferentemente dos outros, que possibilitam um cadastro gratuito e alimentação com publicações e informações independentes de sua indexação, esse identificador exhibe somente publicações do autor que forem indexadas na própria Scopus, e o pesquisador não possui autonomia para gerenciar seu perfil. Dessa forma, o identificador é utilizado para buscar publicações apenas dentro dessa base de dados.

O identificador é automaticamente criado quando o pesquisador possui ao menos uma publicação indexada na Scopus. É comum que erros aconteçam na base e sejam criados mais de um perfil para o mesmo autor, mas devido à falta de autonomia já mencionada, todas as requisições de alteração ou correção são feitas via suporte da Scopus.

O perfil do pesquisador, como apresenta (Figura 26) atualmente informações como: nome completo do autor; Scopus ID; número do ORCID, quando o pesquisador faz essa vinculação; número da conta do Mendeley, gerenciador bibliográfico, com o qual o pesquisador pode fazer a vinculação; opção para editar o perfil, caso o autor precise alterar alguma de suas informações, incluindo sua afiliação preferida a ser exibida no perfil; criar um alerta para receber notificações de novas publicações ou citações recebidas por aquele autor; identificar possíveis autores para colaboração, com base em suas publicações disponíveis no perfil; exportar os dados de publicação para o SciVal, uma plataforma da Elsevier voltada para a análise de indicadores científicos; algumas métricas e gráfico com número de publicações, citações e *h-index*; a listagem de publicações do autor do perfil, exibindo algumas informações básicas sobre elas; e demais informações sobre documentos citantes, co-autores, *preprints*, tópicos e financiamentos recebidos.

Figura 26 – Exemplo de perfil do Scopus ID do pesquisador



Fonte: Scopus, 2021.

O Scopus ID permite a vinculação do ORCID, possibilitando que publicações que estejam vinculadas ao iD autor na base Scopus sejam imputadas no perfil ORCID. Não é possível integração com o ResearchID.

Como a base de dados Scopus não é de acesso livre, estando disponível apenas via assinatura, o identificador, que possui uma URI no formato “<https://www.scopus.com/authid/detail.uri?authorId=57224725549>”, pode ser acessado de um computador sem acesso a uma rede institucional autorizada, mas possui limitações como, por exemplo, não é possível acessar detalhes a respeito de publicações disponibilizadas no perfil.

Gasparyan, Nurmashev, Yessirkepov, Endovitskiy, Voronov e Kitas (2017) comentam que embora o Scopus ID esteja vinculado à base Scopus, que possui uma larga cobertura de documentos e também multidisciplinar, o período de cobertura de conteúdos pode distorcer o perfil de pesquisadores que possuem uma longa carreira acadêmica e com publicações muito citadas anterior a este período que a base indexa. Segundo o guia de cobertura da Scopus, 67,5% dos seus registros com data posterior ao ano 1995 e 32,5% com data anterior a 1995 (ELSEVIER, 2020). Sendo assim, é fundamental se atentar à fonte de informação do perfil do pesquisador, para que avaliações equivocadas não sejam realizadas.

O Scopus ID é um identificador que oferece um perfil interessante ao pesquisador, com visualizações e opções úteis ao usuário final da base de dados que o consulta. Contudo, a falta de autonomia do pesquisador em realizar as alterações em seu perfil e a restrição para exibição de publicações indexadas apenas em sua própria base de dados pode afastar pesquisadores de sua divulgação e manutenção de perfis, mediante contato com suporte, com mais afinco.

5.2 Opções de busca em fontes de informação científica

Ao considerar os identificadores e identificadores persistentes enquanto números que possibilitam a identificação e recuperação da informação científica, é pertinente verificar quais fontes de informação na comunidade científica utilizam esses identificadores.

O uso dos identificadores pode ser feito nos “bastidores” do sistema informacional, ou seja, ser utilizado para estruturação ou organização das informações na base, ou ser também utilizado enquanto filtro de busca por seus usuários.

Com o objetivo de realizar essa verificação, foram escolhidos três tipos de fontes de informação científica, que anteriormente foram discutidas, para análise: bases de dados, repositórios de dados de pesquisa e repositórios de *preprints*.

Sendo assim, foi realizada a análise da utilização de seis identificadores, entre organizacionais (ISNI, Ringgold e GRID) e autorais (ORCID, ResearcherID e ScopusID), em 39 bases de dados ou plataformas de conteúdos científicos, em três repositórios de dados de pesquisa e oito repositórios de *preprints* multidisciplinares.

A escolha apenas por análise em fontes multidisciplinares foi uma escolha metodológica que teve como objetivo se ater apenas a fontes que possuem maior potencial de serem utilizadas por pesquisadores de diferentes áreas de conhecimento, ou seja, se mostram potencialmente mais abrangentes e de maior utilização.

A análise realizada tem como principal objetivo a identificação de fontes de informação que possuem opções de busca ou filtros para pesquisa por identificadores persistentes, especificamente algum dos seis identificadores anteriormente analisados. Algumas das fontes podem atribuir ou utilizar identificadores em alguma parte de sua estruturação ou exibição de dados e inclusive recuperar as publicações pela caixa de busca mesmo que não tenha indicativo ou filtro disponível ao usuário, mas esse não é o objetivo da análise. Todas as fontes analisadas e os seus respectivos dados verificados constam nos Apêndices desse trabalho. A discussão

pontual dos resultados será realizada após cada tópico relativo às respectivas fontes informacionais e uma discussão geral será realizada em um novo tópico ao final do capítulo.

5.2.1 Bases de dados

Para a análise, foram escolhidas 39 fontes informacionais considerando uma listagem² de bases de dados disponibilizada pelo Portal de Busca Integrada da Agência USP de Gestão da Informação Acadêmica (AGUIA), com filtro apenas para fontes multidisciplinares. Originalmente, a listagem, após o filtro por fontes interdisciplinares, possuía 40 registros, contudo o item “Journal Citation Reports (JCR)” estava duplicado, resultando então no número final de itens analisados.

Dentre as fontes analisadas constam bases da própria USP e algumas fontes que se encontravam sem acesso disponível no momento de realização do trabalho e outras não apresentavam exatamente o escopo voltado para recuperação de informação científica enquanto publicações, portanto foram descritas com a nomenclatura “não se aplica”. Todas as bases de dados e os dados considerados na análise constam no Apêndice A.

A análise mostra (Tabela 1) que entre as bases de dados analisadas, cinco possuíam opção de busca pelo identificador ORCID, apenas uma pelo identificador ResearchID e apenas uma pelo identificador ScopusID. Já 16 delas apresentam opções de busca por outros identificadores, que não são os autorais e organizacionais analisados.

Os demais identificadores utilizados como opção de busca nas bases de dados analisadas e que apresentaram apenas: Affiliation ID, com uma ocorrência; CODEN, com uma ocorrência; DOI, com nove ocorrências; *Data Universal Numbering System* (DUNS) Number, com uma ocorrência; ISBN, com seis ocorrências; ISSN, com dez ocorrências; PMID, com duas ocorrências; e *North American Industry Classification System* (NAICS) Code, com uma ocorrência.

Tabela 1 – Quantidade de bases de dados com opção de busca por identificadores

Identificadores	Quantidade de bases de dados com ocorrências
ISNI	0
Ringgold	0

² <http://www.buscaintegrada.usp.br/>

GRID/ROR	0
ORCID	5
ResearcherID	1
ScopusID	1
Outros	16

Fonte: elaborada pela autora.

Os dados mostram que as bases de dados não utilizam a maioria dos identificadores autorais ou organizacionais, com destaque para a presença do identificador ORCID em um número significativo delas, se comparado aos demais. É interessante mencionar que a presença dos identificadores ResearcherID e ScopusID tiveram apenas uma ocorrência cada, pois foram apresentadas pelas fontes informacionais que pertencem ao mesmo grupo editorial dos dois identificadores, sendo elas, Web of Science e Scopus, respectivamente.

Quanto às 16 bases de dados que utilizam ao menos um outro identificador, percebe-se o destaque de ocorrências para a busca pelo identificador ISSN e DOI, que recuperam registros bibliográficos por periódicos científicos ou por objetos digitais, comumente publicações científicas. O identificador Affiliation ID que apresentou uma ocorrência pertence à base de dados Scopus e é utilizado para recuperar publicações de determinadas instituições que, ao possuírem um perfil nessa base de dados, recebem este identificador exclusivo. O identificador PMID é o ID utilizado para recuperar publicações na base de dados PubMed, referência na área de Ciências da Saúde. Enquanto aos demais identificadores: o CODEN, é um identificador utilizado para títulos de periódicos científicos; o DUNS Number, é um identificador universal de dados; e NAICS Code, é um padrão utilizado nos Estados Unidos da América para classificar estabelecimentos comerciais.

Ao adotar os identificadores persistentes como opções de busca ou filtros, as bases de dados podem, ao menos, mostrar à comunidade científica que eles podem ser utilizados como uma alternativa na recuperação informacional. Como evidenciam os dados da amostra selecionada, a presença dos identificadores autorais e organizacionais nesse sentido ainda não é uma realidade, exceto pela presença do identificador ORCID, que é mais difundido, e outros identificadores adotados e conhecidos há mais tempo, como o DOI ou ISSN.

5.2.2 Repositórios de dados de pesquisa

Para a análise, foram escolhidos 3 repositórios de dados de pesquisa considerando uma listagem³ disponibilizada pela AGUIA. Todos os repositórios e os dados considerados na análise constam no Apêndice B.

A análise mostra (Tabela 2) que entre os repositórios de dados de pesquisa analisados, nenhum deles apresenta opções de busca por identificadores organizacionais e autorais ou outros identificadores.

Tabela 2 – Quantidade de repositórios de dados de pesquisa com opção de busca por identificadores

Identificadores	Quantidade de repositórios de dados de pesquisa com ocorrências
ISNI	0
Ringgold	0
GRID/ROR	0
ORCID	0
ResearcherID	0
ScopusID	0
Outros	0

Fonte: elaborada pela autora.

Percebe-se que os repositórios de dados de pesquisa, embora sejam fundamentais no processo de disponibilizado e acesso aos dados brutos utilizados pelo pesquisador, ainda não apresentam uma estruturação de ferramentas de busca que possibilitem a pesquisa por identificadores. O uso desse tipo de fonte informacional deve ser cada vez mais incentivado na comunidade científica, por parte de instituições ou periódicos, contribuindo também com a otimização das plataformas utilizadas com essa finalidade.

5.2.3 Repositórios de *preprints*

Para a análise, foram escolhidos 8 repositórios de *preprints* listados⁴ pela *Accelerating Science and Publication in Biology* (ASAPbio), uma organização sem fins lucrativos que promove o uso de *preprints* para acelerar a disseminação da pesquisa. Todos os repositórios e os dados considerados na análise constam no Apêndice C.

³ <https://www.aguia.usp.br/apoio-pesquisador/dados-pesquisa/lista-repositorios-dados-pesquisa/>

⁴ <https://asapbio.org/preprint-servers>

A análise mostra (Tabela 3) que entre os repositórios de *preprints* analisados, apenas um possui opção de busca pelo identificador autoral ORCID e outro repositório que possui opções de busca por outros identificadores, que não são os organizacionais ou autorais analisados.

Tabela 3 – Quantidade de repositórios de *preprints* com opção de busca por identificadores

Identificadores	Quantidade de repositórios de <i>preprints</i> com ocorrências
ISNI	0
Ringgold	0
GRID/ROR	0
ORCID	1
ResearcherID	0
ScopusID	0
Outros	1

Fonte: elaborada pela autora.

O repositório que apresentou outras opções de busca por identificadores foi o arXiv, sendo elas: arXiv Author ID, utilizado como identificador autoral exclusivo do repositório para recuperar seus autores; arXiv ID, utilizado como identificador exclusivo do repositório para recuperar os *preprints*; e o DOI, também utilizado para recuperar os *preprints* do repositório.

Percebe-se que apenas dois repositórios, da amostra selecionada, apresentam opções de busca por identificadores, representando um número reduzido ao considerar que se tratam de repositórios multidisciplinares.

Os *preprints* têm sido cada vez mais discutidos e presentes nas rotinas da comunidade científica. Sua ampla utilização ainda é discutida, considerando o fato de que o documento ainda não passou pelo processo de avaliação por pares, mas é necessário que as discussões se façam cada vez mais presentes e possibilitem, conseqüentemente, a melhor utilização e estruturação dos repositórios desse tipo de material.

5.3 Discussão geral da análise realizada

Na análise da utilização dos seis identificadores nas fontes informacionais multidisciplinares ficou evidente que as iniciativas ainda não são amplamente adotadas e a maioria das fontes não apresentam opções de busca por identificadores.

Considerando todas as fontes analisadas, as bases de dados estão à frente quanto ao número de opções de busca por identificadores, destacando o já evidente papel e participação na rotina de pesquisa da comunidade científica. Seu destaque mostra o interesse de seus mantenedores, sejam eles grandes grupos empresariais com interesses comerciais ou não, em possuir essas opções de busca.

Enquanto as demais fontes informacionais, sendo elas os repositórios de dados de pesquisa, que não apresentam opções de busca dessa natureza, e os repositórios de *preprints*, que apresentam apenas duas opções, estão pouco inseridas nessa modalidade de pesquisa por códigos alfanuméricos. Esses ambientes, embora sejam amplamente discutidos pelo movimento de Ciência Aberta, ainda são pouco utilizados e disseminados pela comunidade acadêmica. Esse fato pode potencializar a pouca utilização dessas fontes devido à falta de inserção no contexto dos identificadores persistentes, dificultando a assertividade na busca por seus conteúdos.

Iniciativas que não possuem um vínculo direto com grupos comerciais se mostram, até o presente momento, mais adotadas pela comunidade científica, como é o caso do identificador autoral ORCID, que apresentou ocorrência de opção de busca em duas das três fontes analisadas. E à nível de identificadores organizacionais, o ROR, embora não tenha apresentado nenhuma ocorrência na análise, se mostra uma iniciativa que deve seguir caminho semelhante ao ORCID, partindo da mesma “filosofia” de identificador persistente que não está diretamente ligado a um grupo com interesses majoritariamente comerciais, ao menos em uma perspectiva inicial.

Dois exemplos de identificadores autorais que possuem opções de busca apenas em bases de dados de seus mesmos grupos empresariais de origem são o ResearcherID e o ScopusID, vinculados apenas à Web of Science e Scopus, respectivamente. Esse fato evidencia que, embora sejam identificadores conhecidos, estão limitados apenas a um universo limitado de busca.

A discussão dos resultados dessa análise é fundamental para evidenciar que embora diversos atores na comunidade científica, sejam eles periódicos científicos, instituições de pesquisa ou de fomento, incentivem a adoção de identificadores persistentes em seus processos de rotina, é fundamental que se tenha investimento e mudanças na estrutura das fontes informacionais para conseguir recuperar essas informações. É uma questão fácil de ser

compreendida: a informação pode ser representada por meio do código alfanumérico, mas este código precisa ser tratado como uma chave de busca onde quer que essa informação esteja armazenada. A recuperação informacional é a peça-chave no processo para que os identificadores persistentes consigam atingir seu objetivo: no início, por meio de opções de busca nas fontes de informação; e no final, por fazer com que a informação satisfaça a necessidade de seu usuário.

Nesse sentido, não é suficiente incentivar a utilização dos identificadores persistentes sem ter a consciência de que eles, sozinhos, não são capazes de resolver todos os problemas relacionados à recuperação informacional. É indispensável que a estrutura dessas fontes informacionais, e também de todos os metadados que são a elas fornecidas, consiga atender de maneira satisfatória à essas necessidades de busca informacional da comunidade científica.

6 CONSIDERAÇÕES FINAIS

A horizontalidade do acesso à informação científica pela comunidade acadêmica se potencializou em ambiente digital, de forma que é cada vez mais emergente a necessidade de planejamento e elaboração de estratégias que tenham como principal objetivo a recuperação desses conteúdos, para que sejam utilizados como insumo para o desenvolvimento de novas pesquisas e conseqüentemente contribuam com o desenvolvimento da ciência em diferentes áreas do conhecimento.

Contudo, para que seja possível a recuperação dessas informações, é necessário que essas passem por processos de organização e representação adequados, de maneira a proporcionar não somente sua simples recuperação, mas, sobretudo, uma recuperação que apresente resultados de pesquisa pertinentes às necessidades informacionais do pesquisador.

E essa recuperação é realizada, sobretudo, nas fontes de informação científica, presentes também em ambiente digital, sendo as mais consultadas pela comunidade científica no momento de buscar literatura produzida por seus pares para embasar suas pesquisas ou seguir do ponto em que essas publicações pararam.

Contudo, quando se discute a recuperação da informação científica, não se deve pensar apenas na recuperação da literatura por suas palavras-chave ou descritores de assunto. As publicações científicas também devem ser adequadamente recuperadas por seus autores e por suas respectivas instituições de origem, instituições de fomento ou instituições de realização da pesquisa. É interessante que a visibilidade da pesquisa na comunidade científica implique também na visibilidade de seus autores e de suas instituições, conferindo-os créditos pela sua realização e propiciando o contato entre pares e o estabelecimento de possíveis parcerias.

Nesse sentido, surgiram dessa necessidade de articulação, relacionamento e contribuição para além dos grupos de pesquisa, instituições e eventos científicos, as chamadas redes acadêmico-científicas. Essas redes são ambientes digitais, que muito lembram as redes sociais comumente utilizadas na sociedade, voltadas para pesquisadores e que contribuem de alguma forma com a manutenção dessa visibilidade de suas pesquisas e seus perfis ou currículos, considerando que é possível a ampla participação em discussões, o acesso a artigos legalmente disponibilizados pelos autores ou contribuição com o desenvolvimento das pesquisas de seus pares.

Contudo, a correta representação e a conseqüente identificação de autores e instituições na literatura científica não é uma tarefa centralizada e de fácil realização. É comum que existam:

pesquisadores homônimos, sejam eles de uma mesma área de pesquisa ou não; pesquisadores que mudam o nome ao longo de sua trajetória, pessoal e acadêmica, por diferentes motivos; ou pesquisadores que adotam diferentes formas de seu nome em diferentes publicações, seja por falta de atenção, conhecimento do impacto da ação ou falta de orientação ao longo de sua formação por suas instituições de origem. E quanto às instituições, muitas delas apresentam diferentes variações de seus nomes, muitas vezes sendo conhecidas por suas siglas, nomes traduzidos para o idioma inglês, entre outros.

Todas essas variações impactam diretamente nos processos de avaliação da ciência, que normalmente são realizados por departamentos de pesquisa, instituições ou *rankings* nacionais e internacionais, com o objetivo de verificar o nível de produtividade e alcance de suas pesquisas.

Muitas metodologias consideram a realização de busca em algumas fontes de informação pré-estabelecidas e que, por estas variações de nomenclatura, podem ocasionar a recuperação não eficaz de tudo o que foi produzido por aquele pesquisador ou instituição no período avaliado, gerando avaliações que não refletem a realidade e que implicam diretamente em contratações ou demissões de pesquisadores, realização de parcerias institucionais ou até o recebimento ou cancelamento de bolsas ou fomento à pesquisa.

São muitos os envolvidos no processo de publicação de um conteúdo científico e, dessa forma, o padrão utilizado nem sempre é o mesmo. A publicação, antes de ser disponibilizada para consulta e utilização da comunidade acadêmica, passa por seus autores, revisores e editores de periódicos e, finalmente, é indexada em fontes de informação. Nesse sentido, é evidente que um padrão estabelecido ao longo desse *workflow* poderia minimizar as variações de representação dos nomes de pesquisadores e suas instituições, embora, ainda assim, pudessem surgir variações decorrentes de erro humano, como o ortográfico ou simplesmente erros de digitação.

A preocupação com essas variações é recorrente na literatura e na prática profissional que, embora envolva diferentes áreas e profissionais, se trata de um trabalho feito essencialmente por profissionais e pesquisadores das áreas de Biblioteconomia e Ciência da Informação. O controle de autoridade é uma atividade adotada a partir de princípios e regras voltadas à padronização dos pontos de acesso nos catálogos bibliográficos, antes mesmo do surgimento e consolidação de fontes de informação em ambiente *web*.

Outro esforço no sentido de contribuir com a padronização e recuperação informacional está nas normas bibliográficas que, embora existam diversos padrões de utilização de acordo com áreas específicas do conhecimento, propõem a normalização para auxiliar no processo de troca de informações, objetivo intrínseco ao seu conceito. Contudo, muitas normas bibliográficas não

apresentam tanta rigidez, possibilitando adaptações por parte de seus usuários, e algumas lacunas por elas deixadas podem não possibilitar efetivamente uma recuperação informacional eficaz, sobretudo em ambiente digital.

Esses fatores são considerados complicadores quanto à correta identificação de autores e instituições, o que conseqüentemente suscita discussões a respeito da autenticidade de conteúdos acadêmico-científicos em ambiente digital, que por diversas vezes está diretamente envolvido com a disseminação das chamadas *fake news*, que também atingem a informação científica. Nesse sentido, garantir que a informação disseminada, seu pesquisador e sua instituição de origem sejam corretamente identificados é essencial para que as informações disseminadas tenham sua confiabilidade adequadamente certificada.

Diante de tantas implicações da representação inadequada de pesquisadores e suas instituições, potencializadas pela quantidade de informação disponível em ambiente digital, surgem iniciativas tecnologicamente nativas com o objetivo de auxiliar nos processos de desambiguação e identificação de nomes, despertando diversos projetos e metodologias discutidos na literatura e aplicados em diversas fontes informacionais, envolvendo codificações, linguagens de programação e utilização de algoritmos. Contudo, são alternativas que buscam consolidar o que já está disponível em ambiente digital e não antecipam a inserção de nomes fora do padrão na literatura científica.

Nesse sentido, ainda com o objetivo de oferecer uma solução ao problema em ambiente digital, encontram-se os identificadores persistentes que, enquanto códigos alfanuméricos, oferecem a possibilidade de reunir variações dos nomes de pessoas e instituições acadêmicas em uma única *string* de busca, culminando na exatidão comumente propiciada pela combinação entre letras e números ou, nesse caso, conhecidos como IDs.

Na prática, os identificadores persistentes podem ser utilizados no momento da busca em fontes de informação científica, fazendo com que os resultados de pesquisa sejam exatos, ou seja, não sejam menores ou maiores devido à possíveis variações de nomenclatura da chave de busca, seja ela o nome de uma publicação, autor ou instituição.

Os identificadores persistentes, sendo que os bibliográficos são os mais comumente utilizados e conhecidos na comunidade acadêmica, possuem objetivos claros em sua conceituação que proporcionaram o surgimento de iniciativas voltadas para a identificação também de pesquisadores e instituições. O principal objetivo dos identificadores é auxiliar e otimizar o tempo do pesquisador na busca por informações científicas. Contudo, sua compreensão por parte de todos os atores envolvidos no processo de pesquisa e a conseqüente utilização ainda é um desafio.

Esse desafio é potencializado pela falta de padrão em relação aos próprios identificadores, por mais contraditório que pareça ser. Se existem diferentes identificadores autorais e organizacionais a serem utilizados e adotados como padrão, no fim, menos padronizada e compreendida se torna a iniciativa.

A crítica às diversas iniciativas não pode deixar de considerar que existem de fato alguns identificadores que não possuem todas as características técnicas que são determinadas pela literatura científica da área como essenciais para que atinjam seus objetivos de recuperação informacional e que automaticamente suscita a criação de novos identificadores com o objetivo de solucionar essas lacunas. A ausência de algumas dessas características impacta na sua ampla e eficaz utilização.

É compreensível que quando existem muitas iniciativas com o mesmo objetivo e ao considerar que parte significativa delas supõe uma participação e manutenção ativa dos pesquisadores de seus perfis nos identificadores, exista um abandono ou desinteresse pela adoção. O tempo dedicado à realização da pesquisa deve ser prioridade para o pesquisador e demais atividades que despertem a sensação de trabalho duplicado, ou seja, informações semelhantes sendo fornecidas a identificadores com objetivos semelhantes, a precisão dos identificadores se perde por falta de insumos fornecidos.

Muitos identificadores possuem essa característica de “currículo” do pesquisador, pressupondo essa manutenção ativa por sua parte. Contudo, um dos principais objetivos dos identificadores é proporcionar a interoperabilidade informacional. Nesse sentido, caso autores, instituições de pesquisa, universidades, editoras e fontes de informação científica adotem um mesmo padrão de identificação persistente, esses conteúdos seriam constantemente retroalimentados e dispensariam uma ação direta do pesquisador, de fato integrando a comunidade científica e seus produtos. Essa interoperabilidade pressupõe que o identificador seja submetido juntamente a todas as atividades de pesquisa realizadas, como um pré-requisito.

A existência de apenas um padrão de identificação persistente é uma questão que esbarra no constante interesse na criação de diferentes iniciativas, ao invés de contribuir para o desenvolvimento e readequação das já existentes, envolvendo também grupos editoriais com interesses majoritariamente corporativos.

É perceptível, ao realizar a análise mais detalhada dos identificadores organizacionais e autorais, que esses últimos possuem uma interface e utilização muito mais voltada para o conceito de “currículo” acadêmico ou profissional, considerando o objetivo de propiciar visibilidade ao pesquisador. Já os identificadores organizacionais, possuem uma característica mais voltada para

a identificação das instituições fornecendo suas principais informações como nome, variações do nome, localização e relação com outros identificadores, mas não possuem um perfil, por exemplo, com publicações da organização, como acontece nos autorais. As iniciativas organizacionais se mostram mais alinhadas ao conceito original de identificação, possibilitando, através de serviços oferecidos, que os identificadores e a padronização do nome institucional sejam utilizados por meio da integração de sistemas como um índice de autoridades, aproximando-se do conceito trabalhado pelo controle de autoridade realizado nos catálogos bibliográficos.

Normalmente os identificadores são solicitados ao pesquisador pelo periódico científico no momento de submissão da publicação. Contudo, não há uma relação automática direta entre o nome do pesquisador que se encontra manualmente preenchido por ele no perfil do identificador e o nome que será exibido no momento de publicação do artigo. Em contrapartida, o identificador autorial é passível de localização e exibição atualmente em diversas publicações científicas, considerando que tem sido adotado por diversas editoras e periódicos, mas pouco se fala ou se vê nas publicações científicas o identificador das organizações nelas envolvidas.

Considerando que a principal proposta dos identificadores persistentes organizacionais e autorais é proporcionar a identificação adequada de instituições e pesquisadores, é fundamental que exista um alinhamento entre as informações que são disponibilizadas nos metadados das publicações, com esses identificadores, a disponibilização em fontes informacionais e as opções de busca para que seja possível recuperá-las.

Investir em iniciativas únicas e consistentes de identificadores persistentes se mostra o caminho mais adequado quando se fala em seu principal objetivo: padronização para identificação e recuperação.

Não é possível colocar a culpa da não-utilização ou adoção dos identificadores persistentes nos pesquisadores ou nos demais atores envolvidos no processo de comunicação científica. Os identificadores só serão capazes de cumprir seu propósito com assertividade se a cobrança pela alimentação de múltiplos perfis der lugar à utilização de um único identificador persistente adequado à cada necessidade (bibliográfico, autorial, organizacional, etc.), contribuindo para o fortalecimento de um único padrão que então poderá ser trabalhado de forma a propiciar uma interoperabilidade informacional efetiva entre diferentes sistemas e processos acadêmico-científicos.

Embora os identificadores persistentes não sejam uma iniciativa tão recente, é perceptível a lacuna de discussões e estudos mais aprofundados na temática, sobretudo no Brasil. Durante a realização da pesquisa foi possível perceber a presença majoritária, à nível nacional, de artigos

científicos que falam de forma breve e superficial a respeito da utilização dos identificadores, geralmente publicações como editoriais, cartas ao editor ou comentários.

É fundamental que a discussão e a pesquisa sobre identificadores persistentes suscitem o desenvolvimento e, preferencialmente, a aplicação de um único padrão que propicie de fato uma interoperabilidade informacional na comunidade acadêmica, consolidando os identificadores como uma estratégia eficaz de representação, organização e acesso da informação científica em ambiente digital.

REFERÊNCIAS⁵

AERTS, Raf. Digital identifiers work for articles, so why not for authors? **Nature**, London, v. 453, p. 979, jun. 2008. Disponível em: <https://doi.org/10.1038/453979b>. Acesso em: 11 jul. 2021.

AGUIA. **Qual é a diferença entre pre-print, post-print e artigo publicado?** Disponível em: <https://www.aguia.usp.br/noticias/qual-e-a-diferenca-entre-pre-print-post-print-e-artigo-publicado/>. Acesso em: 28 jun. 2020.

ALVARENGA, Lídia. 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, jan. 2003. ISSN 1518-2924. Disponível em: <https://doi.org/10.5007/1518-2924.2003v8n15p18>. Acesso em: 02 maio 2020.

ÁLVAREZ DE TOLEDO, María Luisa. Control de publicaciones em las universidades: abor de las bibliotecas. **Anuario ThinkEPI**, [s. l.], v. 7, p. 58-60, 2013. Disponível em: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/30331/15927>. Acesso em: 11 jul. 2021.

ANANTHAKRISHNAN, Rachana; CHARD, Kyle; D'ARCY, Mike; FOSTER, Ian; KESSELMAN, Carl; MCCOLLAM, Brendan; PRUYNE, Jim; ROCCA-SERRA, Philippe; SCHULER, Robert; WAGNER, Rick. An Open Ecosystem for Pervasive Use of Persistent Identifiers. *In: 2020 Conference on Practice and Experience in Advanced Research Computing: Catch the Wave, 2020. ACM International Conference Proceeding Series*, p. 99-105, 2020. Disponível: <https://dl.acm.org/doi/pdf/10.1145/3311790.3396660>. Acesso em: 09 jul. 2021.

ARAÚJO JÚNIOR, Rogério Henrique de. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT/CB-014 - Comitê Brasileiro de Informação e Documentação. São Paulo: ABNT, c2014. Disponível em: <http://www.abnt.org.br/cb-14#:~:text=ABNT%2FCB%2D014%20%2D%20Comit%C3%AA%20Brasileiro%20de%20Informa%C3%A7%C3%A3o%20e%20Documenta%C3%A7%C3%A3o,-ABNT%2FCB%2D014&text=%C3%82MBITO%20DE%20ATUA%C3%87%C3%83O%3A%20Normaliza%C3%A7%C3%A3o%20no,ci%C3%A4ncia%20da%20informa%C3%A7%C3%A3o%20e%20publica%C3%A7%C3%A3o>. Acesso em: 18 jun. 2021

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 10520**: Citações em documentos – Apresentação. Rio de Janeiro: ABNT, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 12225**: Lombada – Apresentação. Rio de Janeiro: ABNT, 2004.

⁵ De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023)

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 6022**: Artigo em publicação periódica técnica e/ou científica – Apresentação. Rio de Janeiro: ABNT, 2018.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 6023**: Referências – Elaboração. Rio de Janeiro: ABNT, 2018.

ASSUMPCÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ZAFALON, Zaira Regina. O controle de autoridade no domínio bibliográfico: os catálogos digitais. **Biblios**, Pittsburgh, v. 68, p. 21-33, 2017a. Disponível em: <http://dx.doi.org/10.5195/biblios.2017.342>. Acesso em: 30 jul. 2021.

ASSUMPCÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ZAFALON, Zaira Regina. O controle de autoridade no domínio bibliográfico: os catálogos em livros e em fichas. **Biblios**, Pittsburgh, v. 67, p. 84-98, 2017b. Disponível em: <http://dx.doi.org/10.5195/biblios.2017.341>. Acesso em: 17 jul. 2021.

AVENTURIER, Pascal; ALENCAR, Maria Cleofas de. Os desafios dos dados de pesquisas abertos. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 10, n. 3, 2016. Disponível em: <https://doi.org/10.29397/reciis.v10i3.1069>. Acesso em: 27 maio 2020.

BACHELET, Vivienne C.; URIBE, Francisco A.; DÍAZ, Rubén A.; VERGARA, Alonso F.; BRAVO-CÓRDOVA, Fabiana; CARRASCO, Victor A.; LIZANA, Francisca J.; MEZA-DUCAUD, Nicolás; NAVARRETE, María S. Misrepresentation of institutional affiliations: The results from an exploratory case study of Chilean authors. **Learned Publishing**, [s. l.], v. 32, n. 4, p. 335-344, oct. 2019. Disponível em: <https://doi.org/10.1002/leap.1257>. Acesso em: 03 jul. 2021.

BAILEY JUNIOR, C. W. The role of reference librarian in institucional repositories. **References Services Review**, California, v. 33, n. 3, 2005. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00907320510611294/full/pdf>. Acesso em: 17 jul. 2021.

BAUWENS, Michel. A economia política da produção entre pares. **P2PFoundationWiki**. 2010. Disponível em: https://wiki.p2pfoundation.net/A_Economia_Pol%c3%adica_da_Produ%c3%a7%c3%a3o_e_nre_Pares. Acesso em: 01 jun. 2020.

BECK, Kathrin; RITZ, Raphael; WITTENBURG, Peter. **Towards a global digital object cloud** – Report from the views on PID Systems training course and workshop. *In*: RDA Europe Workshop, Mas Planck Compute and Data Facility (MPCDF), Garching-Munich, Germany, 2016. Disponível em: https://www.rd-alliance.org/sites/default/files/attachment/20160901_RDA_PID_event_Garching_report_final.pdf. Acesso em: 17 jul. 2021.

BELLINI, Angela de Brito; CECCANTINI, João Luis Cardoso Tápias; ANTUNES, Margaret Alves; FUJITA, Mariângela Spotti Lopes; FAGUNDES, Silvana Aparecida; NOVELLI, Valéria Aparecida Moreira. Normalização documentária: o caminho de acesso à produção científica. *In*: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS... 2006,

- Salvador. **Anais** [...] Salvador: UFBA, 2006. Disponível em: <http://repositorio.febab.org.br/items/show/5428>. Acesso em: 20 jun. 2021.
- BENKLER, Yochai. Saber comum: produção de materiais educacionais entre pares. **Revista FACED**, Salvador, n.15, jan./jul. 2009. Disponível em: <http://dx.doi.org/10.9771/2317-1219rf.v14i15.3649>. Acesso em: 26 maio. 2020.
- BERNERS-LEE, Tim. **Cool URIs don't change**. 1998. Disponível em: <http://www.w3.org/Provider/Style/URI>. Acesso em: 10 jul. 2021.
- BERRY, D. (Ed). **Understanding digital humanities**. Gurgaon: Palgrave Macmillan, 2012.
- BILDER, Geoffrey; BROWN, Josh; DEMERANVILLE, Tom. Organisation identifiers: current provider survey. 2016. Disponível em: <https://doi.org/10.5438/4716>. Acesso em: 17 jul. 2021.
- BORDIEU, Pierre. O campo científico. *In*: BORDIEU, Pierre. **Sociologia**. São Paulo: Ática, 1983. p. 122-155.
- BOUDRY, Christophe; DURAND-BARTHEZ, Manuel. Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): a case study. **Plos One**, San Francisco, v. 15, n. 9, e0238583, set. 2020. Disponível em: <https://doi.org/10.1371/journal.pone.0238583>. Acesso em: 27 jun. 2021.
- BOURNE, Charles P. Frequency and impact of spelling errors in bibliographic data bases. **Information Processing & Management**, New York, v. 13, n. 1, p. 1-12, 1977.
- BRANDT, Scott. Authority files for microcomputer databases. **Special Libraries**, New York, v. 79, n. 4, p 296-301, fall 1988.
- BUCKLAND, Michel. Information as thing. **Journal of the American Society of Information Science**, New York, v. 42, n. 5, p. 229-251, 1991. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3). Acesso em: 01 maio 2020.
- BURGUER, R. H. **Authority work**: the creation, use, maintenancce, and evaluation of authority records and files. Littleton: Libraries Unlimited, 1985.
- BURKE, Susan K.; SHORTEN, Jay. Name authority work today: a comparison of types of academic libraries. **Library Resources & Technical Services**, Chicago, v. 54, n. 1, p. 4-20, 2011. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5050/6115>. Acesso em: 11 jul. 2021.
- CÂMARA BRASILEIRA DO LIVRO. **Estrutura do ISBN**. c2021. Disponível em: <https://www.cbbservicos.org.br/isbn/estrutura/>. Acesso em: 17 jul. 2021.
- CAMPOS, Luiz Fernando de Barros. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli**: revista eletrônica de

biblioteconomia e ciência da informação Florianópolis, n. 23, 2007. Disponível em: <https://doi.org/10.5007/1518-2924.2007v12n23p16>. Acesso em: 12 jun. 2021.

CAMPELLO, Bernadete Santos. **Introdução ao controle bibliográfico**. 3. ed. Belo Horizonte: Autêntica Editora, 2019.

CAPES. **COVID-19**: editores internacionais liberam acesso gratuito. Disponível em: <https://www.capes.gov.br/36-noticias/10224-covid-19-editores-internacionais-liberam-acesso-gratuito>. Acesso em: 28 jun. 2020.

CAR, Nicholas J.; GOLODONIUC, Pavel; KLUMP, Jens. The challenge of ensuring persistency of identifier systems in the world of ever-changing technology. **Data Science Journal**, London, v. 16, n. 13, p. 1-18, 2017. Disponível em: <https://doi.org/10.5334/dsj-2017-013>. Acesso em: 17 jul. 2021.

CERRAO, Natalia Gallo; CASTRO, Fabiano Ferreira de. Repositórios institucionais das Universidades Federais brasileiras: análise da representação da informação. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 5, n. 1, jan./jun. 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/110392>. Acesso em: 07 jun. 2020.

CHURCHES, Tim; CHRISTEN, Peter; LIM, Kim; XI ZHU, Justin. Preparation of name and address data for record linkage using hidden Markov models. **BMC Medical Informatics and Decision Making**, London, v. 2, e9, p. 1-16, dez. 2002. Disponível em: <https://dx.doi.org/10.1186%2F1472-6947-2-9>. Acesso em: 26 jun. 2021.

CINTRA, Ana Maria; TÁLAMO, Maria de Fátima Gonçalves Moreira; LARA, Marilda Lopes Ginez de Lara; KOBASHI, Nair Yumiko. **Para entender as linguagens documentárias**. São Paulo: Polis, 1994.

CLACK, Doris H. Authority control and linked bibliographic databases. **Cataloging & Classification Quarterly**, New York, v. 8, n. 3-4, p. 35-46, out. 1988.

COUSIJN, Helena; BRAUKMANN, Ricarda; FENNER, Martin; FERGUSON, Christiane; VAN HORIK, Reneé; LAMMEY, Rachael; MEADOWS, Alice; LAMBERT, Simon. Connected Research: The Potential of the PID Graph. **Patterns**, New York, v. 2, n. 2, 100180, jan. 2021. Disponível em: <https://doi.org/10.1016/j.patter.2020.100180>. Acesso em: 15 jul. 2021.

CREDIT Where credit is due. **Nature**, London, v. 462, p. 825, dez. 2009. Disponível em: <https://doi.org/10.1038/462825a>. Acesso em: 11 jul. 2021.

CRESPO, Isabel Merlo; RODRIGUES, Ana Vera Finardi. Normas técnicas e comunicação científica. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 9, n. 1, p. 36-55, jul./dez. 2011. Disponível em: <https://doi.org/10.20396/rdbci.v9i1.1918>. Acesso em: 17 jun. 2021.

CRESS, Phaedra E. Why do academic authors need an ORCID ID? **Aesthetic Surgery Journal**, Oxford, v. 39, n. 6, p. 696-697, 2019.

CROSSREF. **Posted content (includes preprints)**. Disponível em:

<https://www.crossref.org/education/content-registration/recommendations-by-content-type/posted-content-includes-preprints/>. Acesso em: 30 jun. 2020.

CRUSE, Trisha; HAAK, Laure; PENTZ, Ed. Organization Identifier Project: A Way Forward. 2016. Disponível em: <https://doi.org/10.5438/2906>. Acesso em: 17 jul. 2021.

CUNHA, Lélia Galvão Caldas da. Normalização de originais. **Ciência da Informação**, Rio de Janeiro, v. 2, n. 1, p. 59-63, 1973. Disponível em: <http://revista.ibict.br/ciinf/article/view/24/24>. Acesso em: 17 jun. 2021.

CUTTER, Charles A. **Rules for a dictionary catalog**. 4th ed. Washington: Government Printing Office, 1904.

DAMASIO, Edilson. Preprints na comunicação científica: uma introdução. **BIBLOS**: Revista do Instituto de Ciências Humanas e da Informação, Rio Grande, v. 32, n. 2, p. 155-169, jul./dez. 2018. Disponível em: <https://doi.org/10.14295/biblos.v32i2.8635>. Acesso em: 30 jun. 2020.

DAPPER, Angela; FARQUHAR, Adam; KOTARSKI, Rachael; HEWLETT, Kirstie. Connecting the persistente identifier ecosystem: Building the technical and human infrastructure for open research. **Data Science Journal**, London, v. 16, n. 28, p. 1-16, 2017. Disponível em: <https://doi.org/10.5334/dsj-2017-028>. Acesso em: 17 jul. 2021.

DE BRUIN, R. E.; MOED, H. F. The unification of adresses in scientific publications. **Informetrics**, Amsterdam, v. 89/90, p. 65-78, 1990.

DELLAVALLE, Robert P.; HESTER, Eric J.; HEILIG, Lauren F.; DRAKE, Amanda L.; KUNTZMAN, Jeff W.; GRABER, Marla; SCHILLING, Lisa M. Going, Going, Gone: Lost Internet References. **Science**, New York, v. 302, n. 5646, p. 787-788, out. 2003. Disponível em: <https://doi.org/10.1126/science.1088234>. Acesso em: 16 jul. 2021.

DERVOS, Dimitris A.; SAMARAS, Nikolaos; EVANGELIDIS, Georgios; HYVÄRINEN, Jaakko P.; ASMANIDIS1, Ypatios. The Universal Author Identifier System (UAI_Sys). *In*: Proceedings of the First International Scientific Conference, eRA: The contribution of information technology in Science, Economy, Society and Education. 2006. Disponível em: https://repository.arizona.edu/bitstream/handle/10150/105755/C-CAP_UAI.pdf?sequence=1&isAllowed=y. Acesso em: 11 jul. 2021.

DIAS, Maria Matilde Kronka. Normas técnicas. *In*: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Orgs). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Ed. UFMG, 2000. p. 137-152.

ELLIOTT, Sarah. Survey of author name disambiguation: 2004 to 2010. **Library Philosophy and Practice**, Lincoln, v. 473, s. n., p. 1-11, 2010. Disponível em: <https://digitalcommons.unl.edu/libphilprac/473/>. Acesso em: 13 jun. 2021.

ELSEVIER. **Scopus Content Coverage Guide**. Amsterdam: Elsevier, 2020. Disponível em: https://www.elsevier.com/_data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf. Acesso em: 04 jul. 2021.

ENSERINK, Martin. Are you ready to become a number? **Science**, New York, v. 323, p. 1662, mar. 2009. Disponível em: <https://doi.org/10.1126/science.323.5922.1662>. Acesso em: 11 jul. 2021.

FALAGAS, Matthew E. Unique Author Identification Number in Scientific Databases: a suggestion. **PLoS Medicine**, San Francisco, v. 3, n. 5, p. e248-e249, maio 2006. Disponível em: <https://doi.org/10.1371/journal.pmed.0030249>. Acesso em: 11 jul. 2021.

FEDERAÇÃO BRASILEIRA DE ASSOCIAÇÕES DE BIBLIOTECÁRIOS – FEBAB. **Código de Catalogação Anglo-Americano**. 2. ed. rev. 2002. Preparado sob a direção do Joint Steering Committee for Revision of AACR. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004.

FENNER, Martin. Author identifier overview. **LIBREAS Library Ideias**, Berlin, v. 18, 2011. Disponível em: <https://libreas.eu/ausgabe18/texte/03fenner.htm>. Acesso em: 10 jul. 2021.

FENNEY, Patricia. **Proposed schema changes: have your say**. Oxford: Crossref, 2019. Disponível em: <https://www.crossref.org/blog/proposed-schema-changes-have-your-say/>. Acesso em: 25 jul. 2021.

FERREIRA, Anderson A.; GONÇALVES, Marcos André; LAENDER, Alberto H. F. A brief survey of automatic methods for author name disambiguation. **SIGMOD Record**, [s. l.], v. 41, n. 2, p. 15-26, jun. 2012. Disponível em: <https://sigmodrecord.org/publications/sigmodRecord/1206/pdfs/04.surveys.ferreira.pdf>. Acesso em: 26 jun. 2021.

FERREIRA, Sueli Mara Soares Pinto; MARCHIORI, Patrícia Zeni; CRISTOFOLI, Fulvio. *In*: FERREIRA, Sueli Mara Soares Pinto; TARGINO, Maria das Graças (Orgs). **Acessibilidade e visibilidade de revistas científicas eletrônicas**. São Paulo: Editora Senac São Paulo: Cengage Learning, 2010.

FONSECA, Edson Nery da. **Problemas de comunicação da informação científica**. São Paulo: Thesaurus Editora, 1973.

FOX, Robert. From strings to things. **Digital Libraries Perspective**, Bingley, v. 32, n. 1, p. 02-06, 2016. Disponível em: <https://doi.org/10.1108/DLP-10-2015-0020>. Acesso em: 07 jul 2021.

FOUREZ, Gérard. **A construção das ciências: introdução à filosofia e à ética das ciências**. São Paulo: Editora da Universidade Estadual Paulista, 1995.

GARCÍA-GÓMEZ, Consol. ORCID: um sistema global para la identificación de investigadores. **El profesional de la información**, Barcelona, v. 21, n. 2, p. 210-212,

mar./abr. 2012. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=3883673>. Acesso em: 11 jul. 2021.

GARFIELD, Eugene. British quest for uniqueness versus American egocentrism. **Nature**, London, v. 223, p. 763, 1969. Disponível em: <https://doi.org/10.1038/223763b0>. Acesso em: 11 jul. 2021.

GARFIELD, Eugene. Citation Indexes for Science: A New Dimension in Documentation Through Association of Ideas. **Science**, New York, v. 122, p. 108-111, 1955. Disponível em: <https://doi.org/10.1126/science.122.3159.108>. Disponível em: 11 jul. 2021.

GARVEY, William D. **Communication, the essence of science**: facilitating information among librarians, scientists, enginners and students. Oxford: Pergamon, 1979.

GASPARYAN, Armen Yuri; AKAZHANOV, Nurbek A.; VORONOV, Alexander A.; KITAS, George D. Systematic and Open Identification of Researchers and Authors: focus on Open Researcher and Contributor ID. **Journal of Korean Medical Science**, Seoul, v. 29, n. 11, p. 1453-456, nov. 2014. Disponível em: <https://dx.doi.org/10.3346/jkms.2014.29.11.1453>. Acesso em: 11 jul. 2021.

GASPARYAN, Armen Yuri; NURMASHEV, Bekaidar; YESSIRKEPOV, Marlen; ENDOVITSKIY, Dimitry A.; VORONOV, Alexander A.; KITAS, George D. Researcher and author profiles: opportunities, advantages, and limitations. **Journal of Korean Medical Science**, Seoul, v. 32, n. 11, p. 1749-1756, 2017. Disponível em: <https://dx.doi.org/10.3346/jkms.2017.32.11.1749>. Acesso em: 02 jul. 2021.

GASPARYAN, Armen Yuri; YESSIRKEPOV, Marlen; DUISENOVA, Akmaral; TRUKHACHEV, Vladimir I.; KOSTYUKOVA, Elena I.; KITAS, George D. Researcher and Author Impact Metrics: Variety, Value, and Context. **Journal of Korean Medical Science**, Seoul, v. 33, n. 18, e139, apr. 2018. Disponível em: <https://doi.org/10.3346/jkms.2018.33.e139>. Acesso em: 27 jun. 2021.

GASPARYAN, Armen Yuri; YESSIRKEPOV, Marlen; GERASIMOV, Alexey N.; KOSTYUKOVA, Elena I.; KITAS, George D. Scientific author names: errors, corrections, and identity profiles. **Biochemia Medica**, Zagreb, v. 26, n. 2, p. 169-173, 2016. Disponível em: <http://dx.doi.org/10.11613/BM.2016.017>. Acesso em: 27 jun. 2021.

GATENBY, Janifer; MACEWAN, Andrew. ISNI: a new system for name identification. **Information Standards Quarterly**, Baltimore, v. 23, n. 3, p. 4-9, 2011. Disponível em: https://www.niso.org/sites/default/files/stories/201712/FE_Gatenby_Macewan_ISNI_isqv23no3.pdf. Acesso em: 24 jul. 2021.

GONZÁLEZ DE GÓMEZ, Maria Nélide. A representação do conhecimento e o conhecimento da representação: algumas questões epistemológicas. **Ciência da Informação**, Brasília, v. 22, n. 3, p. 217-222, set./dez. 1993.

GONZÁLEZ, Paula Regina Ventura Amorim; SIMIONATO, Ana Carolina; ARAKAKI, Felipe Augusto; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Repositórios de dados na América Latina: a ISSO 16363: 2012 e a representação da informação. *In: XIX*

Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), 2018, Londrina. **Anais eletrônicos...** Londrina: UEL, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103008>. Acesso em: 07 jun. 2020.

GRID. **Policies**. 2021. Disponível em: <https://www.grid.ac/pages/policies>. Acesso em: 25 jul. 2021.

GUEDÓN, Jean-Claude. Acesso aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, Sueli Mara Soares Pinto; TARGINO, Maria das Graças (Orgs). **Acessibilidade e visibilidade de revistas científicas eletrônicas**. São Paulo: Editora Senac São Paulo: Cengage Learning, 2010.

HAAK, Laurel L.; MEADOWS, Alice; BROWN, Josh. Using ORCID, DOI, and Other open identifiers in research evaluation. **Frontiers in Research Metrics and Analysis**, Lausanne, v. 3, n. 2, out. 2018. Disponível em: <https://doi.org/10.3389/frma.2018.00028>. Acesso em: 09 jul. 2021.

HAAK, Laurel L. Persistent identifier can improve provenance and attribution and encourage sharing of research results. **Information Services & Use**, [s. l.], v. 34, p. 93-96, 2014. Disponível em: <https://doi.org/10.3233/ISU-140736>. Acesso em: 17 jul. 2021.

HABERMANN, Ted. Metadata and Reuse: Antidotes to Information Entropy. **Patterns**, New York, v. 10, n. 1, 100004, apr. 2021. Disponível em: <https://doi.org/10.1016/j.patter.2020.100004>. Acesso em: 15 jul. 2021.

HOFMAN, Hans. Review: Some Comments on Preservation Metadata and the OAIS Model. **DigiCULT.Info - A Newsletter on Digital Culture**, n. 2, 2002.

HOPPEN, Natacha Helena Franz; MAGNUS, Ana Paula Medeiros; VANZ, Samile Andréa de Souza. Referências de atos normativos brasileiros na comunicação científica internacional: estilos Vancouver e APA. **Prisma.com**, Porto, v. 26, p. 41-58, 2018. Disponível em: <https://brapci.inf.br/index.php/res/v/73975>. Acesso em: 20 jun. 2021.

HUANG, Yongwen; LI, Jiao; SUN, Tan; XIAN, Guojian. Institution information specification and correlation based on institutional PIDs and IND tool. **Scientometrics**, Amsterdam, v. 122, p. 381-396. Disponível em: <https://doi.org/10.1007/s11192-019-03268-9>. Acesso em: 18 jul. 2021.

HUBER, R.; KLUMP, J. **How dead is dead in the PID Zombie Zoo?** In? RDA Europe Workshop, Max Planck Compute and Data Facility, Garching-Munich, Germany, 2016. Disponível em: https://www.rd-alliance.org/sites/default/files/attachment/20160902-RDA_EU_View_on_PID_Systems_Garching-Robert_Huber-Jens_Klump-How_dead_is_dead_in_the_PID_Zombie_zoo.pdf. Acesso em: 17 jul. 2021.

HUSSAIN, Ijaz; ASGHAR, Sohail. A survey of author name disambiguation techniques: 2010-2016. **The Knowledge Engineering Review**, Cambridge, v. 32, e22, p. 1-24, dec. 2017. Disponível em: <https://doi.org/10.1017/S0269888917000182>. Acesso em: 26 jun. 2021.

IFLA WORKING GROUP ON FLANAR. **Functional Requirements for Authority Data: a conceptual model**. Den Haag: IFLA, 2013. Disponível em: https://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf. Acesso em 12 jun. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Declaração dos Princípios Internacionais de Catalogação (PIC)**. Holanda, 2016. Tradução de Marcelo Votto Teixeira. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf. Acesso em: 10 jun. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Name of persons**. [s. l.]: IFLA, c2020. Disponível em: <https://www.ifla.org/node/4953#:~:text=According%20to%20the%20International%20Cataloguing,conventions%20of%20the%20country%20and>. Acesso em: 10 jun. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Statement of International Cataloguing Principles**. [s. l.]: IFLA, 2009. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf. Acesso em: 16 maio 2021.

INTERNATIONAL STANDARD BOOK NUMBER SYSTEM. **ISBN Users' Manual**. London: International ISBN Agency, 2017. Disponível em: <https://www.isbn-international.org/content/isbn-users-manual>. Acesso em: 17 jul. 2021.

ISNI INTERNATIONAL AGENCY. **Get an ISNI**. 2021a. Disponível em: <https://isni.org/page/get-an-isni/>. Acesso em: 24 jul. 2021.

ISNI INTERNATIONAL AGENCY. **Joining ISNI**. 2021b. Disponível em: <https://isni.org/page/joining-isni/>. Acesso em: 23 jul. 2021.

ISNI INTERNATIONAL AGENCY. **Key statistics**. 2021c. Disponível em: <https://isni.org/>. Acesso em: 23 jul. 2021.

JAMALI, Hamid R.; NICHOLAS, David; HERMAN, Eti. Scholarly reputation in the digital age and the role of emerging platforms and mechanisms. **Research Evaluation**, Surrey, v. 25, n. 1, p. 37-49, 2016. Disponível em: <https://doi.org/10.1093/reseval/rvv032>. Acesso em: 04 jul. 2021.

JARDIM, Hallini Izabel Ruberto; ZAIDAN, Phillipe Derwich Silva. Controle de informação: uma análise sobre o papel da censura e da Fake News na história brasileira. **Múltiplos Olhares em Ciência da Informação**, [s. l.], v. 8, n. 2, 2018. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/16883>. Acesso em: 03 jul. 2021.

JINHA, Arif E. Article 50 million: an estimate of the number of scholarly articles in existence. **Learned Publishing**, [s. l.], v. 23, p. 258-263, jul. 2010. Disponível em: <https://doi.org/10.1087/20100308>. Acesso em: 27 jun. 2021.

JOLY, Etienne. Further Advantages of a Unique Author Identification Number. **PLoS Medicine**, San Francisco, v. 3, n. 8, p. e367-e368, aug. 2006. Disponível em: <https://doi.org/10.1371/journal.pmed.0030368>. Acesso em: 11 jul. 2021.

JUTY, Nick; WIMALARATNE, Sarala M.; SOILAND-REYES, Stian; KUNZE, John; GOBLE, Carole A.; CLARK, Tim. Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR. **Data Intelligence**, Cambridge, v. 2, n. 1-2, p. 30-39, jan. 2020. Disponível em: https://doi.org/10.1162/dint_a_00025. Acesso em: 17 jul. 2021.

KAPLAN, F. A map for big data research in digital humanities. **Frontiers in Digital Humanities**, [s. l.], v. 2, p. 1-7, 2015. Disponível em: <https://doi.org/10.3389/fdigh.2015.00001>. Acesso em: 20 ago. 2019.

KARAULOVA, Maria; GÖK, Adbullah; SHAPIRA, Philip. Identifying Author Heritage Using Surname Data: An Application for Russian Surnames. **Journal of the Association for Information Science and Technology**, Hoboken, v. 70, n. 5, p. 488-498, 2019. Disponível em: <https://dx.doi.org/10.1002%2Fasi.24104>. Acesso em: 27 jun. 2021.

KENDALL, Graham; YEE, Angelina; HARDY, Steven. We should be just a number and we should embrace it. **The Electronic Library**, [s. l.], v. 35, n. 2, p. 348-357, 2017. Disponível em: <https://doi.org/10.1108/EL-04-2016-0090>. Acesso em: 11 jul. 2021.

KURIEN, Biji T. Name variations can hit citation rankings. **Nature**, London, v. 453, p. 450, may 2008. Disponível em: <https://doi.org/10.1038/453450a>. Acesso em: 27 jun. 2021.

LAMMEY, Rachael. Solutions for identification problems: a look at the Research Organization Registry. **Science Editing**, Seoul, v. 7, n. 1, p. 65-69, 2020. Disponível em: <https://doi.org/10.6087/kcse.192>. Acesso em: 25 jul. 2021.

LAWRENCE, Steve; COETZEE, Frans M.; GLOVER, Eric; PENNOCK, David M.; FLAKE, Gary William; NIELSEN, Finn Arup; KROVETZ, Robert; KRUGER, Andries; GILES, C. Lee. Persistence of Web references in scientific research. **Computer**, [s. l.], v. 34, n. 2, p. 26-31, feb. 2001. Disponível em: <https://doi.org/10.1109/2.901164>. Acesso em: 16 jul. 2021.

LE COADIC, Yves-François. **A ciência da informação**. Brasília (DF): Briquet de Lemos, 2004.

LECARDELLI, Jane; MONTEIRO, Gabriela; LUCAS, Elaine Rosangela de Oliveira. Gestão de dados científicos no contexto dos paradigmas de Capurro: algumas reflexões. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 15, n. 3, set./dez. 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1284>. Acesso em: 31 maio 2020.

LEE, Dong; STVILIA, Besiki. Developing a data identifier taxonomy. **Cataloging & Classification Quarterly**, New York, v. 52, p. 303-336, 2014. Disponível em: <https://doi.org/10.1080/01639374.2014.880166>. Acesso em: 17 jul. 2021.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília: IBICT, 2009. Disponível em: <http://livroaberto.ibict.br/handle/1/775>. Acesso em: 14 jun. 2020.

LIMA, José Leonardo Oliveira; ALVARES, Lilian. Organização e representação da informação e do conhecimento. In: ALVARES, Lilian (Org.). **Organização da informação e**

do conhecimento: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Editores, 2012.

LOPES, I. L. Estratégia de busca e recuperação da informação: revisão de literatura. **Ciência da Informação**, Brasília, DF, v. 31, n. 2, p. 60-71, ago. 2002.

MARTELETO, Regina Maria. Análise de redes sociais – aplicação nos estudos de transferência da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 71-81, jan./abr. 2001. Disponível em: <https://www.scielo.br/pdf/ci/v30n1/a09v30n1>. Acesso em: 23 abr. 2021

MCGARRY, Kevin. **O contexto dinâmico da informação**. Brasília: Briquet de Lemos, 2004.

MCMURRY, Julie A.; JUTY, Nick; BLOMBERG, Niklas; BURDETT, Tony; CONLIN, Tom; CONTE, Nathalie; COURTOT, Mélanie; DECK, John; DUMONTIER, Michel; FELLOWS, Donal K.; GONZALEZ-BELTRAN, Alejandra; GORMANNS, Phillip; GRETHE, Jeffrey; HASTINGS, Janna; HÉRICHÉ, Jean-Karim; HERMJAKOB, Henning; ISON, Jon C.; JIMENEZ, Rafael C.; JUPP, Simon; KUNZE, John; LAIBE, Camille; LE NOVÈRE, Nicolas; MALONE, James; MARTIN, Maria Jesus; MCENTYRE, Johanna R.; MORRIS, Chris; MUILU, Juha; MÜLLER, Wolfgang; ROCCA-SERRA, Phillippe; SANSONE, Susanna Assunta; SARIYAR, Murat; SNOEP, Jackie L.; SOILAND-REYES, Stian; STANFORD, Natalie J.; SWAINSTON, Neil; WASHINGTON, Nicole; WILLIAMS, Alan R.; WIMALARATNE, Saraia M.; WINFREE, Lilly M.; WOLSTENCROFT, Katherine; GOBLE, Carole; MUNGALL, Christopher J.; HAENDEL, Melissa A.; PARKINSON, Helen. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. **PLoS Biology**, San Francisco, v. 15, n. 6, e2001414. Disponível em: <https://doi.org/10.1371/journal.pbio.2001414>. Acesso em: 10 jul. 2021.

MEADOWS, Alice; HAAK, Laurel L.; BROWN, Josh. Persistent identifiers: the building blocks of the research information infrastructure. **Insights**, [s. l.], v. 32, n. 1, p. 1-6. Disponível em: <http://doi.org/10.1629/uksg.457>. Acesso em: 16 jul. 2021.

MEADOWS, Alice; HAAK, Laurel L. How persistent identifiers can save scientists time. **FEMS Microbiology Letter**, Amsterdam, v. 365, fny143, 2018. Disponível em: <https://doi.org/10.1093/femsle/fny143>. Acesso em: 11 jul. 2021.

MEADOWS, Alice. PIDapalooza – the open festival for persistente identifiers. **Insights**, [s. l.], v. 30, n. 3, p. 1-4, nov. 2017. Disponível em: <https://doi.org/10.1629/uksg.393>. Acesso em: 17 jul. 2021.

MEADOWS, Arthur Jack. **A comunicação científica**. Brasília (DF): Briquet de Lemos, 1999.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. **Catálogo no plural**. Brasília, DF: Briquet de Lemos, 2009.

MICHENER, William K.; BRUNT, James W.; HELLY, John J.; KIRCHNER, Thomas B.; STAFFORD, Susan G. Nongeospatial metadata for the ecological sciences. **Ecological Applications**, Washington, v. 7, n. 1, p. 330-342, 1997. Disponível em:

[https://doi.org/10.1890/1051-0761\(1997\)007\[0330:NMFTEJ\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTEJ]2.0.CO;2). Acesso em: 15 jul. 2021.

MIKKI, Susanne; ZYGMUNTOWSKA, Marta; GJESDAL, Øyvind Liland; AL RUWEHY, Hemed Ali. Digital Presence of Norwegian Scholars on Academic Network Sites—Where and Who Are They? **Plos One**, San Francisco, v. 10, n. 11, e0142709, nov. 2015. Disponível em: <https://dx.doi.org/10.1371/journal.pone.0142709>. Acesso em: 27 jun. 2021.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT'ANA, Ricardo Cesar Gonçalves. Repositórios de dados científicos nas universidades brasileiras e portuguesas. *In: Seminário em Ciência da Informação (SECIN)*, 6, 2016, Londrina. **Anais eletrônicos...** Londrina: UEL, 2016. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/viewFile/338/166>. Acesso em: 07 jun. 2020.

MOREIRO GONZÁLEZ, José Antonio. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2011.

MUELLER, Suzana Pinheiro Machado. A ciência, o sistema de comunicação científica e a literatura científica. *In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Orgs). Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: Ed. UFMG, 2000.

MUGNAINI, Rogério; DIGIAMPIETRI, Luciano Antonio; OLIVEIRA, Laucivaldo Cardoso de; FERREIRA, Sueli Mara Soares Pinto. Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. **Em Questão**, Porto Alegre, v. 18, n. especial, p. 263-279, dez. 2012. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/33265>. Acesso em: 12 jun. 2021.

NASCIMENTO, Andréa Gonçalves do. **Almetria para bibliotecários**: guia prático de métricas alternativas para avaliação da produção científica. São Paulo: Scortecci, 2017.

NASSI-CALÒ, Lilian. Critérios de autoria preservam a integridade na comunicação científica. **SciELO em Perspectiva**, mar. 2018. Disponível em: <https://blog.scielo.org/blog/2018/03/14/criterios-de-autoria-preservam-a-integridade-na-comunicacao-cientifica/#.YOD4FuhKiUk>. Acesso em: 03 jul. 2021.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. Arlington: National Science Foundation, 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 31 maio 2020.

OECD. Organization for Economic Co-operation and Development. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. 2007. Disponível em: <https://www.oecd.org/sti/inn/38500813.pdf>. Acesso em: 27 maio 2020.

OLIVEIRA, Marlene de. Origens e evolução da Ciência da Informação. *In: OLIVEIRA, Marlene de (Org.). Ciência da Informação e biblioteconomia*: novos conteúdos e espaços de atuação. 2 ed. Belo Horizonte: Editora UFMG, 2011. p. 9-28.

OLIVER, Chris. **Introdução à RDA**: um guia básico. Brasília, DF: Briquet de Lemos, 2011.

ORTEGA, José Luis. Relationship between altmetric and bibliometric indicators across academic social sites: the case of CSIC's members. **Journal of Informetrics**, Amsterdam, v. 9, n. 1, p. 39-49, jan. 2015. Disponível em: <https://doi.org/10.1016/j.joi.2014.11.004>. Acesso em: 02 jul. 2021.

PARSONS, Mark A.; DUERR, Ruth E.; JONES, Matthew B. The History and Future of Data Citation in Practice. **Data Science Journal**, London, v. 18, n. 52, p. 1-10, 2019. Disponível em: <http://doi.org/10.5334/dsj-2019-052>. Acesso em: 10 jul. 2021.

PASCHUAL, Cristina Herrerb. El control de autoridades. **Anales de Documentación**, Murcia, n. 2, p. 121-136, 1999. Disponível em: <https://revistas.um.es/analesdoc/article/view/2621>. Acesso em: 30 maio 2021.

PATRIAS, Karen; WENDLING, Dan (ed.). **Citing Medicine**: The NLM Style Guide for Authors, Editors, and Publishers. 2a ed. Bethesda (MD): National Library of Medicine: 2007. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK7256/>. Acesso em: 20 jun. 2021.

POBLACIÓN, Dinah Aguiar. À guisa de prefácio. *In*: POBLACIÓN, Dinah Aguiar; WITTER, Geraldina Porto; RAMOS, Lúcia Maria S. V. Costa; FUNARO, Vânia M. B. de Oliveira (Org.). **Revistas científicas**: dos processos tradicionais às perspectivas alternativas de comunicação. Cotia: Ateliê Editorial, 2011. p. 9-10.

PRINCIPE, Eloisa. Comunicação científica e redes sociais. ALBAGLI, Sarita (Org.). **Fronteiras da Ciência da Informação**. Brasília: IBICT, 2013. p. 196-216.

PYLARINOU, Stamatoula; KAPIDAKIS, Sarantos. Tracking scholarly publishing of hospitals using MEDLINE, Scopus, WoS e Google Scholar. **Journal of Hospital Librarianship**, New York, v. 17, n. 3, p. 209-216, 2017. Disponível em: <https://doi.org/10.1080/15323269.2017.1332934>. Acesso em: 04 jul. 2021.

QIU, Jane. Identity crisis. **Nature**, London, v. 451, n. 14, p. 766-767, 2008. Disponível em: <https://doi.org/10.1038/451766a>. Acesso em: 27 jun. 2021.

REYNOLDS, Regina Romano. Everything old is new again: ISSN in the digital environment. **Ciência da Informação**, Brasília, v. 44, n. 1, p. 96-11, jan./abr. 2015. Disponível em: <http://revista.ibict.br/ciinf/article/view/1435>. Acesso em: 17 jul. 2021.

RESNIK, David B.; TYLER, Ana M.; BLACK, Jennifer R.; KISSLING, Grace. Authorship policies of scientific journals. **Journal of Medical Ethics**, London, v. 42, n. 3, p. 199-202, mar. 2016. Disponível em: <https://dx.doi.org/10.1136%2Fmedethics-2015-103171>. Acesso em: 03 jul. 2021.

RINGGOLD. **Identify Database**. 2021a. Disponível em: <https://www.ringgold.com/identify/>. Acesso em: 14 jul. 2021.

RINGGOLD. **Identify Online free lookup service**. 2021b. Disponível em: <https://www.ringgold.com/identify-online-guests/>. Acesso em: 14 jul. 2021.

RINGGOLD. **Samples: Identify Database Hierarchy & Records**. 2021c. Disponível em: <https://www.ringgold.com/identify-sample/>. Acesso em: 14 jul. 2021.

RODRIGUES, Mara Eliane Fonseca; LIMA, Marcia H. T. de Figueredo; GARCIA, Marcia Japor de Oliveira. A normalização no contexto da comunicação científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 3, n. 2, p. 147-156, jul./dez. 1998. Disponível em: <https://brapci.inf.br/index.php/res/v/33139>. Acesso em: 17 jun. 2021.

ROR. **About**. 2021. Disponível em: <https://ror.org/about/>. Acesso em: 25 jul. 2021.

ROSADO, Kelia Mara Lara; DIAS, Célia da Consolação. Controle de autoridade de nomes de pessoas: uma proposta a partir da análise da Biblioteca Digital do Ministério Público Federal. **Múltiplos Olhares em Ciência da Informação**, [s. l.], v. 9, n 2, p. 1-19, 2019.

ROSADO, Kelia Mara Lara; DIAS, Célia da Consolação. Representação e recuperação de nomes de pessoas em catálogos de autoridades. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 25, p. 1-25, 2020. Disponível em: <https://doi.org/10.5007/1518-2924.2020.e70349>. Acesso em: 12 jun. 2021.

RUIZ-PÉREZ, R.; LÓPEZ-CÓZAR, E. Delgado; JIMÉNEZ-CONTRERAS, E. Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies. **Journal of the Medical Library Association**, Chicago, v. 90, n. 4, p. 411-430, oct. 2002. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC128958/>. Acesso em: 27 jun. 2021.

TORRES-SALINAS, Daniel; MARTÍN-MARTÍN, Alberto; FUENTE-GUTIÉRREZ, Enrique. Analysis of the coverage of the Data Citation Index – Thomson Reuters: disciplines, document types and repositories. **Revista Española de Documentación Científica**, Madrid, v. 37, n. 1, p. 9-6, 2014. Disponível em: <http://dx.doi.org/10.3989/redc.2014.1.1114>. Acesso em: 07 jun. 2020.

SANDBERG, Jane; JIN, Qiang. How should catalogers provide authority control for journal article authors? Name identifiers in the linked data world. **Cataloguing & Classification Quarterly**, New York, v. 54, n. 8, p. 537-552, 2016. Disponível em: <https://doi.org/10.1080/01639374.2016.1238429>. Acesso em: 10 jul. 2021.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <https://doi.org/10.22478/ufpb.1981-0695.2017v12n1.34194>. Acesso em: 20 jun. 2020.

SANTARÉM SEGUNDO, J. E. **Representação Iterativa: um modelo para repositórios digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <https://repositorio.unesp.br/handle/11449/103346>. Acesso em: 07 jun. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em:

http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 20 jun. 2020.

SCHROTER, Sara; MONTAGNI, Ilaria; LODER, Elizabeth; EIKERMANN, M.; SCHÄFFNER, Elke; KURTH, Tobias. Awareness, usage and perceptions of authorship guidelines: an international survey of biomedical authors. **BMJ Open**, London, v. 10, e036899, 2020. Disponível em: <https://doi.org/10.1136/bmjopen-2020-036899>. Acesso em: 03 jul. 2021.

SCIELO. **Crítérios, política e procedimentos para a admissão e a permanência de periódicos científicos na coleção SciELO Brasil**. 2017. Disponível em: https://www.scielo.br/avaliacao/Criterios_SciELO_Brasil_versao_revisada_atualizada_outubro_20171206.pdf. Acesso em: 07 jun. 2020.

SCIELO. **SciELO Preprints (pilot)**. Disponível em: <https://preprints.scielo.org/index.php/scielo>. Acesso em: 30 jun. 2020.

SERRA, Liliana Giusti; SILVA, Luciana Candida da; SANTARÉM SEGUNDO, José Eduardo. O ORCID como aplicação de *Linked Data* no Catálogo de Bibliotecas. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 4, n. 2, p. 185-205, jul./dez. 2017. Disponível em: <https://doi.org/10.22478/ufpb.2358-3908.2017v4n2.40186>. Acesso em: 11 jul. 2021.

SILVA, Hemerson Soares da; LAZZARIN, Fabiana Aparecida. Preprints na comunicação científica: proposta de modelo para aceleração do sistema de comunicação científica. **Informação@Profissões**, Londrina, v. 8, n. 2, p. 150-170, jul./dez. 2019. Disponível em: <http://dx.doi.org/10.5433/2317-4390.2019v8n2p150>. Acesso em: 26 jun. 2020.

SILVA, J. A. Teixeira da. Non-compliance with ethical rules caused by misuse of ORCID accounts: Implications for medical publications in the COVID-19 era. **Ethics, Medicine and Public Health**, France, v. 18, 100692, 2021. Disponível em: <https://doi.org/10.1016/j.jemep.2021.100692>. Acesso em: 18 jul. 2021.

SILVA, Jonathas Luiz Carvalho. Normatividade, tecnicidade e/ou científicidade da Biblioteconomia. **Transinformação**, Campinas, v. 25, n. 1, p. 5-17, jan./abr. 2020. Disponível em: <https://doi.org/10.22478/ufpb.1981-0695.2013v8n2.18635>. Acesso em: 17 jun. 2021.

SILVA, José Antônio Dias da; GOUVEIA, Fábio Castro. Uso de identificadores persistentes e research blogging por blogues científicos em língua portuguesa e sua cobertura pela altmetric. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 25, p. 01-19, 2020. Disponível em: <https://doi.org/10.5007/1518-2924.2020.e71277>. Acesso em: 09 jul. 2021.

SIMIONATO, A. C. Mapeamento dos metadados para dados científicos. *In*: Encontro Nacional de Pesquisa em Ciência da Informação, 18, 2017, Marília. **Anais...** Marília: UNESP, 2017. Disponível em: <https://brapci.inf.br/index.php/article/view/0000027279/de0369e0f94a8669cb0852f73053935c>. Acesso em: 14 jun. 2020.

SMALHEISER, Neil R.; TORVIK, Vetle I. Author name disambiguation. **Annual Review of Information Science and Technology**, Medford, v. 43, p. 287-313, 2009. Disponível em: http://abel.lis.illinois.edu/tutorial/ARIST_preprint.pdf. Acesso em: 13 jun. 2021.

STARR, Joan; CASTRO, Eleni; CROSAS, Mercè; DUMONTIER, Michel; DOWNS, Robert R.; DUERR, Ruth; HAAK, Laurel L.; HAENDEL, Melissa; HERMAN, Ivan; HODSON, Simon; HOURCLÉ, Joe; KRATZ, John Ernest; LIN, Jennifer; NIELSEN, Lars Holm; NURNBERGER, Amy; PROELL, Stefan; RAUBER, Andreas; SACCHI, Simone; SMITH, Arthur; TAYLOR, Mike; CLARK, Tim. Achieving human and machine accessibility of cited data in scholarly publications. **PeerJ Computer Science**, San Diego, v. 1, e1, 2015. Disponível em: <https://doi.org/10.7717/peerj-cs.1>. Acesso em: 10 jul. 2021.

STROTMANN, Andreas; ZHAO, Dangzhi. Author name disambiguation: what difference does it make in author-based citation analysis. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 63, n. 9, p. 1820-1833, 2012. Disponível em: <https://doi.org/10.1002/asi.22695>. Acesso em: 09 jul. 2021.

SUGIMOTO, Cassidy R.; LARIVIÈRE, Vincent. **Measuring research**: what everyone needs to know. New York: Oxford University Press, 2018.

TANENBAUM, Theresa Jean. Publishers: let transgender scholars correct their names. **Nature**, London, v. 583, p. 493, jul. 2020. Disponível em: <https://doi.org/10.1038/d41586-020-02145-3>. Acesso em: 10 jul. 2021.

TARGINO, Maria das Graças. Comunicação científica: uma revisão de seus elementos básicos. **Revista Informação & Sociedade**: estudos, João Pessoa, v. 10, n. 2, 2000, p. 37-85.

TARGINO, Maria das Graças. O óbvio da informação científica: acesso e uso. **TransInformação**, Campinas, v. 19, n. 2, p. 95-105, 2007. Disponível em: <https://www.scielo.br/j/tinf/a/qRRhvBw8yjJH3D7rqdMJtVN/abstract/?lang=pt>. Acesso em: 12 jun. 2019.

TARTAROTTI, Roberta Cristina Dal'Evedove; DAL'EVEDOVE, Paula Regina; FUJITA, Mariângela Spotti Lopes. Biblioteconomia de dados em repositórios de pesquisa: perspectivas para a atuação bibliotecária. **Informação & Informação**, Londrina, v. 24, n. 3, p. 207-226, set./dez. 2019. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2019v24n3p207>. Acesso em: 21 jun. 2020.

THIBODEAU, K. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. **The State of Digital Preservation: An International Perspective**. Anais... Washington: CLIR and Library of Congress, 2002. Disponível em: https://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_37e.pdf. Acesso em: 12 jun. 2021.

THOMAS, Joseph; CHIEN, Barbara; CLEMENT, Gail. ORCID Identifiers: Planned and Potential Uses by Associations, Publishers, and Librarians. **The Serials Librarian**, New York, v. 68, p. 332-341, 2015. Disponível em: <https://doi.org/10.1080/0361526X.2015.1017713>. Acesso em: 11 jul. 2021.

TORVIK, Vetle J.; SMALHEISER, Neil R. Author name disambiguation. **ACM Transactions on Knowledge Discovery from Data**, New York, v. 3, n. 3, p. 1-29, jul. 2009. Disponível em: <https://dl.acm.org/doi/10.1145/1552303.1552304>. Acesso em: 26 jun. 2021.

VEIGA, José Eli da. Revista científica mais antiga do mundo completa 354 anos. **Jornal da USP**. 29 ago. 2019. Disponível em: <https://jornal.usp.br/atualidades/revista-cientifica-mais-antiga-do-mundo-completa-354-anos/>. Acesso em: 10 maio 2021.

VELLUCCI, Sherry L. Metadata. **Annual Review of Information Science and Technology**, Medford, v. 33, p. 187-222, 1998.

VEVE, Marielle. Supporting name authority control in XML metadata: a practical approach at the University of Tennessee. **Library Resources & Technical Services**, Chicago, v. 53, p. 41-52, 2009. Disponível em: <https://doi.org/10.5860/lrts.53n1.41>. Acesso em: 13 jun. 2021.

ZIMBA, Olena; GASPARYAN, Armen Yuri. Scientific authorship: a primer for researchers. **Reumatologia**, Poland, v. 58, n 6, p. 345-349, 2020. Disponível em: <https://doi.org/10.5114/reum.2020.101999>. Acesso em: 04 jul. 2021.

WARNER, Simeon. Author identifiers in scholarly repositories. **Journal of Digital Information**, Texas, v.11, p. 1-10, 2010. Disponível em: <https://journals.tdl.org/jodi/index.php/jodi/article/download/752/641/0>. Acesso em: 10 jul. 2021.

WEITZEL, Simone da Rocha. **Os repositórios de e-prints como nova forma de organização da produção científica**: o caso da área das Ciências da Comunicação no Brasil. 2006. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2011. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-14052009-133509/pt-br.php>. Acesso em: 19 jun. 2019.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle; AXTON, Myles; BAAK, Arie; BLOMBERG, Niklas; BOITEN, Jan-Willem; SANTOS, Luiz Bonino da Silva Santos; BOURNE, Philip E.; BOUWMAN, Jildau; BROOKES, Anthony J.; CLARK, Tim; CROSAS, Mercè; DILLO, Ingrid; DUMON, Olivier; EDMUNDS, Scott; EVELO, Chris T.; FINKERS, Richard; GONZALEZ-BELTRAN, Alejandra; GRAY, Alasdair J.G.; GROTH, Paul; GOBLE, Carole; GRETHE, Jeffrey S.; HERINGA, Jaap; HOEN, Peter A.C.; HOOFT, Rob; KUHN, Tobias; KOK, Ruben; KOK, Joost; LUSHER, Scott J.; MARTONE, Maryann E.; MONS, Albert; PACKER, Abel L.; PERSSON, Bengt; ROCCA-SERRA, Philippe; ROOS, Marco.; VAN SCHAİK, Rene; SANSONE, Susanna-Assunta; SCHULTES, Erik; SENGSTAG, Thierry; SLATER, Ted; STRAWN, George; SWEERTZ, Morris A.; THOMPSON, Mark; VAN DER LEI, Johan; VAN MULLIGEN, Erik; VELTEROP, Jan; WAAGMEESTER, Andra; WITTENBURG, Peter; WOLSTENCROFT, Katherine; ZHAO, Jun; MONS, Barend. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, New York, v. 3, n. 1, p. 1-9, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 09 jul. 2021.

WITTER, Geraldina Porto. Apresentação: revistas científicas: da incubação à visibilidade. *In*: POBLACIÓN, Dinah Aguiar; WITTER, Geraldina Porto; RAMOS, Lúcia Maria S. V. Costa; FUNARO, Vânia M. B. de Oliveira (Org.). **Revistas científicas**: dos processos tradicionais às perspectivas alternativas de comunicação. Cotia: Ateliê Editorial, 2011. p. 9-10.

YAMAOKA, Eloi Juniti; GAUTHIER, Fernando Ostuni. Objetos digitais: em busca da precisão conceitual. **Informação & Informação**, Londrina, v. 18, n. 2, p. 77-97, maio/ago. 2013. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2013v18n2p77>. Acesso em: 12 jun. 2021.

SciELO Livros	Não	Não	Não	Não	Não	Não	ISBN
ScienceDirect	Não	Não	Não	Não	Não	Não	ISBN; ISSN
Scimago Journal & Country Rank	Não se aplica	Não se aplica	Não se aplica	Não se aplica	Não se aplica	Não se aplica	ISSN
SciVal	Não se aplica	Não se aplica	Não se aplica	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Scopus	Não	Não	Não	Sim	Não	Sim	ISSN; CODEN; ORCID; Affiliation ID; DOI
Web of Science	Não	Não	Não	Sim	Sim	Não	DOI; PubMed ID

APÊNDICE B – Análise de repositórios de dados de pesquisa

Fontes de informação	ISNI	Ringgold	GRID/ROR	ORCID	ResearcherID	ScopusID	Outros
Figshare	Não	Não	Não	Não	Não	Não	Não
Dryad	Não	Não	Não	Não	Não	Não	Não
Zenodo	Não	Não	Não	Não	Não	Não	Não

