JOSÉ AUGUSTO SALIM

Unifying biotic interactions data: terminology, data analysis, standardization, and proposal of a data schema for plant-pollinator interactions

São Paulo
2023

JOSÉ AUGUSTO SALIM

Unifying biotic interactions data: terminology, data analysis, standardization, and proposal of a data schema for plant-pollinator interactions

Corrected version

Ph. D. Thesis presented to the Polytechnic School, Universidade de São Paulo, Brazil to obtain the degree of Doctor of Science.

São Paulo
2023

JOSÉ AUGUSTO SALIM

UNIFYING BIOTIC INTERACTIONS DATA: TERMINOLOGY, DATA ANALYSIS, STANDARDIZATION, AND PROPOSAL OF A DATA SCHEMA FOR PLANT-POLLINATOR INTERACTIONS

Corrected version

Ph. D. Thesis presented to the Polytechnic School, Universidade de São Paulo, Brazil to obtain the degree of Doctor of Science.

Concentration area: Computer Engineering

Advisor:
Prof. Dr. Antonio Mauro Saraiva

São Paulo
2023

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 18 de dezembro de 2023

Assinatura do autor: José Augusto Salim

Assinatura do orientador: gov.br Documento assinado digitalmente
ANTONIO MAURO SARAIVA
Data: 18/12/2023 10:59:28-0300
Verifique em https://validar.iti.gov.br

Catalogação-na-publicação

Salim, José Augusto
Unifying biotic interactions data: terminology, data, analysis, standardization, and proposal of a data schema for plant-pollinator interactions / J.A. Salim -- versão corr. -- São Paulo, 2023.
220 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Informática para biodiversidade 2.Interações bióticas 3.Padronização de dados I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all those who have contributed to the completion of my PhD thesis.

First and foremost, I am thankful to my advisor, Dr. Antonio Mauro Saraiva, for his invaluable guidance and support throughout this journey, and I am deeply grateful for the opportunity to work with him.

I am also grateful to all members of the Brazilian Network on Plant-Pollinator Interactions (REBIPP) for the inestimable contributions to this thesis.

I am thankful to Dr. Paula Florencia Zermoglio, Dr. Debora Pignatari Drucker and M.S. Filipi Miranda Soares for their insightful feedback, constructive criticism, and valuable suggestions that have significantly improved the content of my work. I am indebted to them for all the time that they generously provided during the elaboration of this work.

Thanks also to Dr. Jorrit H. Poelen and Dr. Katja C. Seltmann for providing advice regarding exploration of biotic interactions data in Global Biotic Interactions (GloBI) and Global Biodiversity Information Facility (GBIF).

I would like to express my gratitude to my family, and in particular, my incredible partner Alexandra Guidelli Rosa, as well as my friends for their unwavering love, encouragement, patience, and belief in my abilities. Their constant support and understanding have been vital in keeping me motivated and focused.

Lastly, I want to acknowledge the funding agencies or scholarships that have provided financial support for my research. I am deeply thankful to the São Paulo Research Foundation (FAPESP) for funding related to the Safeguarding Pollination Services in a Changing World: theory into practice (SURPASS2) project, specifically under Award Number 18/14994-1, as well as for the fellowship granted under Award Number 2019/27179-7. Additionally, I also sincerely grateful to the Brazilian National Council for Scientific and Technological Development (CNPq) for the doctoral fellowship granted under Award Number 142028/2016-9. Their support has been instrumental in enabling me to pursue my research endeavors and has greatly contributed to my academic and professional development.

In conclusion, I am humbled and grateful to all those who have played a role, big or small, in shaping this thesis. Your contributions have been invaluable, and I am truly thankful for your support.

José Augusto Salim

ABSTRACT

Biotic interactions are essential for biodiversity, ecological communities, and ecosystem functioning, but primary data are limited in terms of accessibility and standardization. As the availability of biotic interaction data increases, there has been a growing demand to properly document data to enhance data aggregation, reuse and interoperability. While the adoption of the Darwin Core standard for biodiversity data is widespread, its application to biotic interactions data is limited due to the lack of specialized data schemas, vocabularies and guidelines. This study aims to develop a data model using the Darwin Core standard for sharing biotic interactions data, along with a specialized vocabulary of terms for the standardization of plant-pollinator interactions data. This was initially addressed by reviewing the concepts and terminology used in the research community, highlighting the key data elements that align with state-of-the-art knowledge in the field. The existing datasets were also explored for formats, metadata and data standardization. The review of scientific publications emphasizes the importance of aligning the existing terminology within the community of practice to ensure clarity and consistency in the interpretation and analysis of biotic interactions data. The review of existing datasets revealed the limitations in interoperability and data reuse, emphasizing the need for standardized formats and metadata annotation. Based on these findings, the novel data schema and vocabulary of terms aim to capture a contextualized and more realistic representation of biotic interactions and facilitate data sharing and analysis. The development process emphasizes a community-driven approach, prioritizing the engagement and input of the community of practice. The developed vocabulary represents a significant step towards data exchange and interoperability in the field of Pollination Biology. Overall, the findings contribute to the advancement of standardizing biotic interactions data.

Keywords – biotic interactions, Darwin Core, data standardization, vocabulary of terms, plant-pollinator interactions.

RESUMO

As interações bióticas são essenciais para a biodiversidade, comunidades ecológicas e o funcionamento dos ecossistemas, mas os dados primários encontram-se limitados em termos de acessibilidade e padronização. Com o aumento da disponibilidade de dados de interação biótica, há uma crescente demanda pela documentação adequada dos dados, a fim de aprimorar a agregação, reutilização e interoperabilidade dos dados. Embora a adoção do padrão Darwin Core para dados de biodiversidade seja amplamente difundida, sua aplicação em dados de interações bióticas é limitada devido à falta de esquemas de dados especializados, vocabulários e diretrizes. Este estudo tem como objetivo desenvolver um modelo de dados utilizando o padrão Darwin Core para compartilhamento de dados de interações bióticas, juntamente com um vocabulário de termos especializado para a padronização de dados de interações planta-polinizador. Inicialmente uma revisão dos conceitos e terminologia utilizados na comunidade de pesquisa foi realizada, destacando os principais elementos de dados que estão alinhados com o conhecimento estado-da-arte na área. Os conjuntos de dados existentes também foram explorados em relação aos formatos, metadados e padronização empregados. A revisão das publicações científicas enfatiza a importância de alinhar a terminologia utilizada pela comunidade, a fim de garantir clareza e consistência na interpretação e análise dos dados de interações bióticas. A revisão dos conjuntos de dados existentes revelou limitações na interoperabilidade e na reutilização dos dados, destacando a necessidade de formatos padronizados e inclusão de metadados. Com base nesses achados, o novo esquema de dados e o vocabulário de termos visam capturar uma representação contextualizada e mais realista das interações bióticas. O processo de desenvolvimento adotado enfatiza uma abordagem voltada para a comunidade, priorizando o envolvimento e a contribuição da comunidade. O vocabulário desenvolvido representa um passo significativo em direção à integração de dados e interoperabilidade no campo da Biologia da Polinização. No geral, os resultados contribuem para o avanço na padronização dos dados de interações bióticas.

Palavras-Chave – interações bióticas, Darwin Core, padronização de dados, vocabulário de termos, interações planta-polinizador.

LIST OF FIGURES

LIST OF TABLES

LIST OF CODES

LIST OF ABBREVIATIONS AND ACRONYMS

BioCASe Biological Collection Access Service. 79

BoW Bag of Words. 41, 43

CERN European Organization for Nuclear Research. 65

CoL Catalogue of Life. 81, 88

CSV Comma Separated Values. 68, 132, 133

CV Controlled Vocabularies. 128

DataONE Data Observation Network for Earth. 64, 65

DOI Digital Object Identifier. 65

DTM Document-Term Matrix. 42, 44, 47, 48, 51, 52

DwC Darwin Core. 28, 37–39, 71, 73–76, 78, 81, 84, 85, 97, 99–104, 106, 107, 110–113, 115, 116, 122, 124–126, 129, 130, 132, 133, 138

DwC-A Darwin Core Archive. 39, 72–74, 79–82, 84, 93, 99, 102–107, 111, 113, 114, 116, 121, 129, 130, 132, 133

DwC-SW Darwin Core Semantic Web. 112

EML Ecological Metadata Language. 65, 67, 104, 130

eMoF Extended Measurement or Fact. 102, 105, 106

FAIR Findable, Accessible, Interoperable, Reusable. 35, 63

FAO Food and Agriculture Organization of the United Nations. 116, 118

GBIF Global Biodiversity Information Facility. 28, 29, 36, 38, 63–65, 71, 73, 74, 79–81, 83, 84, 86–90, 93, 97, 113, 114, 129, 132, 137

GEF Global Environment Facility. 117–119

GloBI Global Biotic Interactions. 36, 38, 63, 71–74, 79, 81–84, 102, 137

HTTP HyperText Transfer Protocol. 128

IABIN-PTN InterAmerican Biodiversity Information Network–Pollinators Thematic Network. 99, 116, 117

ICTV International Committee on Taxonomy of Viruses. 88

IDF Inverse Document Frequency. 47

iDigBio Integrated Digitized Biocollections. 38, 79

IG Interest Group. 37, 114

IPT Integrated Publishing Toolkit. 102, 129, 130, 132

IRI Internationalized Resource Identifier. 38, 111, 112, 128

ITIS Integrated Taxonomic Information System. 81, 88

IUCN International Union for Conservation of Nature. 92, 96

IWDB Interaction Web DataBase. 70, 71

JSON JavaScript Object Notation. 76, 132, 133

LDA Latent Dirichlet Allocation. 44, 45, 53, 55, 60

MAGE-Tab MicroArray Gene Expression Tabular. 36

MG Maintenance Group. 37

MIAME Minimum Information About a Microarray Experiment. 36

NCBI National Center of Biotechnology Information. 82, 83, 88, 90

NCEAS National Center for Ecological Analysis and Synthesis. 70

NLP Natural Language Processing. 41, 42, 44

NoSQL non-relational. 132

NSF National Science Foundation. 63

OBIS Ocean Biodiversity Information System. 102

OTT Open Tree of Life. 82, 88

PCA Principal Component Analysis. 51, 52, 54, 55

PPI Plant-Pollinator Interactions. 122, 126, 128–130, 132, 133, 138, 139

RDA Research Data Alliance. 63

RDF Resource Description Framework. 39, 80, 98, 103, 110–114, 121, 130

REBIPP Brazilian Network on Plant-Pollinator Interactions. 29, 114, 115, 119, 122–124, 128–134, 139

RO OBO Relations Ontology. 38, 81, 87, 112, 139

TDM Term-Document Matrix. 42

TDWG Biodiversity Information Standards. 29, 37, 38, 97–100, 114, 116, 119

TF-IDF Term Frequency–Inverse Document Frequency. 42, 43, 47, 48, 51–53

TOP The Transparency and Openness Promotion. 63

tsv Tab-Separated Values. 68

UNEP United Nations Environment Programme. 117

URI Uniform Resource Identifier. 106, 111

VEM Variational Expectation-Maximization. 45

W3C World Wide Web Consortium. 110

WDPA World Database on Protected Areas. 92

Web World Wide Web. 110, 133

WFO World Flora Online. 82, 88

WGS84 World Geodetic System 1984. 91

XML Extensible Markup Language. 39, 99, 103, 104, 106, 107, 110, 113, 121

GLOSSARY

API Application Programming Interface is a way for two or more computer programs to communicate with each other. 71, 72, 79

blob store or binary large object store, is a storage system designed for handling binary data or large objects, such as images, videos, documents, or any other unstructured data. It is a specialized type of data storage that focuses on efficiently storing and retrieving large, immutable binary data. . 80

content negotiation is the mechanism that is used for serving different representations of a resource to the same URI to help the user agent specify which representation is best suited for the user (for example, which document language, which image format, or which content encoding). 128

data author the person who is responsible for producing and/or sharing data. 35, 36, 102, 129, 130, 132

data portal Is any online platform which supports users in accessing collections of data. 36

Linked Data is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. 103, 113

CONTENTS

[1] According to the Brazilian Association of Technical Standards (ABNT NBR 6023).

1 INTRODUCTION

> *"All Nature is linked together by invisible bonds and every organic creature, however low, however feeble, however dependent, is necessary to the well-being of some other among the myriad forms of life"*
>
> -- George Perkins Marsh in Man and Nature
> 1864, 109

The concept of biotic interactions generally includes the actions performed by organisms on the one another (PRINGLE, 2016). It may focus on the species interacting (*species-level*) or on the organisms interacting (*organism-level*) (NAKAZAWA, 2020). Biotic interactions can be positive or negative, weak or strong, involve mutualism or competition (BEGON; TOWNSEND, 2021), be the result of a behavior (ECHELLE; ECHELLE; HILL, 1972; LIMA, 2002), be static (ODUM, 1959) or dynamic (MARON; BAER; ANGERT, 2014), be interspecific or intraspecific, and can scale to populations, communities or ecosystems. Biotic interactions play important roles in the evolution of biodiversity (MITTELBACH; MCGILL, 2019; THOMPSON, 1999; THOMPSON, 2005; TYLIANAKIS *et al.*, 2008; VELLEND, 2016), in the assembly and dynamics of communities (BEGON; TOWNSEND, 2021), and as drivers of ecosystems functioning (LOREAU *et al.*, 2001). They are central to the persistence of almost every form of life on Earth (THOMPSON, 1999). However, despite different efforts (FAYLE *et al.*, 2016; FORTUNA; ORTEGA; BASCOMPTE, 2014; GUIMARÃES; RAIMUNDO; CAGNOLO, 2012; POELEN; SIMONS; MUNGALL, 2014; POISOT *et al.*, 2016; THOMPSON *et al.*, 2012; VAZQUEZ; MELIAN, 2008), data on biotic interactions remains limited in terms of accessibility and reusability.

The growing interesting in information sharing in science demands not only on data sharing, but also proper preparation, annotation and consistency of data and metadata (KITA *et al.*, 2022). However, the majority of currently available biotic interactions datasets do not adopt any (meta)data standard (e.g., Allen-Perkins *et al.* (2022)). Even when standards are adopted, the absence of suitable data schema and vocabularies greatly contributes to the dispersion and heterogeneity of the data. This heterogeneity poses significant challenges to data aggregation, as it requires laborious and repetitive transformations of custom and nonstandardized datasets into a common structure. As a result, data integration and discovery become costly and time-consuming processes. In addition, the lack of comprehensive community guidelines for documenting data leads to variations in how sampling methods, protocols, and efforts are reported. These inconsistencies restrict the generalization of conclusions derived from biotic interactions data (BRIMACOMBE *et al.*, 2023; SIELEMANN; HAFNER; PUCKER, 2020).

In the biodiversity community, the adoption of the Darwin Core (DwC) standard (WIECZOREK *et al.*, 2014) has been widespread for documenting and sharing biodiversity data, particularly regarding occurrence data. The DwC standard has played a crucial role in the indexing and aggregation of billions of records by various information systems, such as the Global Biodiversity Information Facility (GBIF) (GBIF Secretariat, 2022). Despite the potential use of the DwC standard for documenting biotic interactions data, its adoption is limited and diverse across different communities due to the lack of specialized vocabularies of terms and guidelines for best practices in data preparation and annotation.

Therefore, this study aims to investigate various aspects related to the historical creation, management and sharing of biotic interactions data. The primary objectives are: to create a data model for the standardization of biotic interactions data, utilizing the DwC standard; to develop a specialized vocabulary of terms designed for plant-pollinator interactions data; and to elaborate common guidelines for documenting biotic interactions data. By creating a data model and a vocabulary of terms, this study ultimately aims to facilitate the efficient analysis and synthesis of information pertaining to biotic interactions, with an emphasis on plant-pollinator interactions.

This thesis is structured as follows:

a) chapter 2 provides a review of the study of biotic interactions in Ecology. Additionally, it introduces the field of Biodiversity Informatics, highlighting its significance in managing biodiversity data.

b) chapter 3 explores the relationships and differences among the most commonly used terms to refer to biotic interactions in the scientific literature;

c) chapter 4 offers an overview of the existing biotic interactions datasets. It investigates the repositories where data are deposited, the prevalent file formats and licenses, and limitations associated with current data;

d) chapter 5 outlines a methodology for extracting and standardizing biotic interactions data from the GBIF registry. It provides detailed information about the taxonomic, geographic and temporal coverage of the extracted data;

e) chapter 6 presents a community-driven data model for sharing biotic interactions data. This chapter builds upon the analysis conducted in previous chapters and in the synergies established with the Brazilian Network on Plant-Pollinator Interactions (REBIPP) and Biodiversity Information Standards (TDWG) Biological Interaction Data IG communities;

f) chapter 7 introduces a vocabulary of terms specifically developed for plant-pollinator interactions data. It also describes general guidelines for development of community-driven vocabularies based on insights from the REBIPP community;

g) chapter 8 presents the practical and technical outputs and approaches for standardizing and annotating plant-pollinator interactions data within the REBIPP community;

h) finally, Chapter 9 presents the overall conclusions of this study and provides directions for future work.

Summary of contributions

Below is the list of contributions of this thesis:

- Interactions data model and standardization: Salim, *et al.* Data standardization of plant–pollinator interactions. **GigaScience**, v. 11, p. giac043, 2022. ISSN 2047-217X. Disponível em: ⟨doi:10.1093/gigascience/giac043⟩

- Vocabulary of terms for plant-pollinator interactions data standardization: REBIPP. 2021. Plant-Pollinator Interactions Vocabulary of Terms. Brazilian Network on Plant-Pollinator Interactions (REBIPP). ⟨https://ppi.rebipp.org.br/terms/⟩,

- Controlled vocabulary for plant-pollinator interactions terms: REBIPP. 2021. Plant-Pollinator Interactions Controlled Vocabulary List of Terms. Brazilian Network on Plant-Pollinator Interactions (REBIPP). ⟨http://rs.rebipp.org.br/ppi/doc/cv/2021-12-03⟩,

- Review of the terminology adopted in the literature referring to the biotic interactions subject,

- SALIM, J. A. et al. Indexing Biotic Interactions in GBIF data. **Biodiversity Information Science and Standards**, v. 6, p. e93565, 23 ago. 2022,

- SALIM, J. A.; SARAIVA, A. A Google Sheet Add-on for Biodiversity Data Standardization and Sharing. **Biodiversity Information Science and Standards**, v. 4, p. e59228, 10 fev. 2020,

- DRUCKER, D. et al. Plant-pollinator Interaction Data: A case study of the World-FAIR project. **Biodiversity Information Science and Standards**, p. 643–683, set. 2022,

- SALIM, J. A. et al. Plant-pollinator Vocabulary - a Contribution to Interaction Data Standardization. **Biodiversity Information Science and Standards**, p. 4–12, set. 2021,

- SARAIVA, A. et al. Brazilian Plant-Pollinator Interactions Network: definition of a data standard for digitization, sharing, and aggregation of plant-pollinator interaction data. **Biodiversity Information Science and Standards**, p. 468–479, 14 ago. 2017.

2 BACKGROUND

"in nature, nothing exists alone"

-- Rachel Carson in Silent Spring (1962), 51

This chapter provides an overview of the background in the study of biotic interactions within the fields of Ecology and Biodiversity Informatics. It explores the concepts and methodologies that contribute to the understanding of biotic interactions. Additionally, it highlights the role of Biodiversity Informatics in facilitating the management and analysis of biodiversity data.

2.1 History and terminology

During the last century, biotic interactions have been widely studied, both theoretically and empirically, and a vast body of knowledge has been accumulated which provided the basis for formulation of general principles about biotic interactions. Nonetheless, significant knowledge gaps still exist, that must be addressed to provide a comprehensive and conclusive theory of the evolutionary and ecological consequences of biotic interactions (GÓMEZ; IRIONDO; TORRES, 2023).

Haskell (1947) was the first to present a comprehensive approach, proposing the "interaction grid" while studying human social behavior. The "interaction grid" serves as a summary of the net effects of different interactions. Two years later, Haskell proposed the categorization of "co-actions" (HASKELL, 1949) which was eventually embraced by biologists as "interactions". Furthermore, Burkholder (1952) had set the basis of the "effect-based" interactions terminology, which establish a functional rather than morphological classification of the interactions. Burkholder focused on microorganism interactions, where he utilized changes in activity rates as a proxy for measuring the effect of the interactions and to attribute a signal to each effect (i.e., no effect: 0, increase the rate: + and reduce the rate: −). Later, Odum (1959) introduced the "interaction grid" into ecology in his influential textbook, improving Burkholder's classification by distinguishing the effects of organisms when they are interacting from those when they are not. For example, he defined mutualism as a combination of negative effects on population growth and survival (-/-) when populations are not interacting, and positive effect (+/+) when they are interacting.

However, by the mid-1980s, the interaction grid had received much criticism regarding the ambiguity surrounding the interpretation of the interaction signals (i.e. +, −, 0) (TEMPLETON; GILBERT, 1985; ABRAMS, 1987), and the confusion between interaction mechanisms and their effects (ABRAMS, 1987; ARTHUR; MITCHELL, 1989; BRONSTEIN, 1994). Another problem with the interaction grid approach is its tendency to promote a static view of different forms of interactions. The effects of the interactions can substantial differ according to the space and time (MATTSON; ADDY, ; LEARY, 1985; POISOT; STOUFFER; GRAVEL,), and also, shift between one type to another over evolutionary time (BRADSHAW; SCHEMSKE, 2003; SACHS; WILCOX, 2006; KIERS *et al.*, 2010). However, it is widely assumed nowadays that most biotic interactions are *context-dependent* (HOEKSEMA *et al.*, 2010; BUTTERFIELD; CALLAWAY, 2013; CHAMBERLAIN; BRONSTEIN; RUDGERS, 2014; MARON; BAER; ANGERT, 2014; HOEKSEMA; BRUNA, 2015; FREDERICKSON, 2017), which poses a challenge to the practical application of the “interaction grid”.

Despite the existence of alternative classification schemes for biotic interactions (DINDAL, 1975; PRICE, 1984; BRONSTEIN, 1994), it is noteworthy that many of these are to some degree variations of the interaction grid. Consequently, they share the same underlying assumptions and challenges with the interaction grid. However, just two years after proposing the interaction grid, Haskell presented the “co-action compass” (HASKELL, 1949). The co-action compass is capable of representing not only the sign but the magnitude of an interaction’s net effect (Figure 1). Despite its utility, the co-action compass seems to remain unknown by biologists, as most ecology textbooks continue to rely on some form of the grid to describe biotic interactions. However, recent studies and ecology textbooks have explored the co-action compass (BRONSTEIN, 2001; PRINGLE, 2016; DAVISON, 2020; MATHIS; BRONSTEIN, 2020) and, also, the understating that biotic interactions vary on a continuum opposed to the static representation given by the grid (BERLOW *et al.*, 2004; WOOTTON; EMMERSON, 2005; BOLNICK *et al.*, 2011).

Another source of confusion in defining biotic interactions lies in the choice of the biological unit under consideration. While the observational unit is the interacting organisms or group of organisms (e.g., microorganisms), biotic interactions are often described at the population or species level - the biological unit of interest (LAZIC; CLARKE-WILLIAMS; MUNAFÒ, 2018). According to Lazic, Clarke-Williams & Munafò (2018), researchers should use measurements taken from the observational units to infer proper-

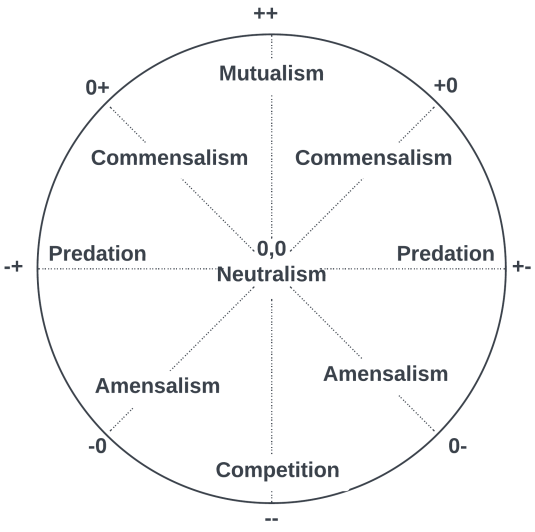


Figure 1: The “Co-action compass”

Source: image adapted from Bronstein (2001) by the author.

ties of the biological units of interest. It also applies to the study of biotic interactions, where the measurements are taken from the interacting individuals (observational units), but they are further used to infer the effects at the population or species level (the biological units of interest). Although many studies implicitly assume species homogeneity, it does not necessarily invalidate their findings. But caution should be exercised when generalizing the conclusions to broader contexts. Conversely, considering interactions at the individual level enables the incorporation of spatial-temporal information integrated with (functional) traits, outcomes and effects into a more realistic and context-dependent perspective of biotic interactions.

Biotic interactions, as pointed by many studies, are affected and exert direct (ALLGEIER; ADAM; BURKEPILE, 2017; BOLNICK *et al.*, 2011; BROSE *et al.*, 2019; BROUSSEAU; GRAVEL; HANDA, 2018; CIRTWILL; EKLÖF, 2018; COUX, 2016; LAIGLE *et al.*, 2018; MONTERO-CASTAÑO; VILÀ, 2017; OLIVAL *et al.*, 2017; RUMEU *et al.*, 2018; SEBASTIÁN-GONZÁLEZ *et al.*, 2017; SLETVOLD; TYE; AGREN, ; WATTS *et al.*, 2016) and indirect (WERNER; PEACOR, 2003) effects on individuals' traits and population dynamics (OLITO; FOX, 2015). Consequently, it is common for studies on biotic interaction to extend beyond the “tetranomials” (i.e., the concatenation of the two Latin binomials) (JORDANO, 2021) and incorporate the sampling of individuals or species' traits. Although, recent efforts to share and standardization of traits-data (KELLER *et al.*, 2023; SCHNEIDER *et al.*, 2019), publishing traits data alongside biotic interactions data have not been adequately addressed.

In a recent study, Gómez, Iriondo & Torres (2023) presented a sophisticated theoretical discussion and proposed the hypotheses of a *continua in interaction outcomes*. The fundamental basis of their definition is the idea of *interaction events*, which entails immediate effects on at least one of the interacting organisms. These events yield *immediate outcomes* for the individuals involved and exert an impact on their fitness, thereby influencing their (*individual outcomes*). Additionally, an interaction event may lead to a long-term effect on the demography, growth rate and dynamics of populations, referred to as *population outcomes*. In order to test their hypothesis, Gómez, Iriondo & Torres (2023) elaborated a mathematical model and conducted simulations based on real-world cases. This formulation enables the decomposition of the biotic interactions into multiple interdependent events and subevents, encompassing both short-term and long-term dynamics. The study conducted by Gómez, Iriondo & Torres (2023) supports the definition of biotic interactions adopted in this study as events and their associated outcomes and effects.

The lack of a convention of ecological nomenclature, as pointed out by Herrando-Pérez, Brook & Bradshaw (2014), also contributes to the persistence of conflicts and ambiguities in the interpretations of biotic interactions. Therefore, any study on biotic interactions should clearly specify the theoretical framework (e.g., *effect-based*, *mechanistic*, *continuum*, *organism vs. species level*) that forms the basis for its assumptions and supports its results and conclusions.

2.2 Biodiversity informatics developments: data digitization, standardization and aggregation

Technology advances of the last decades have enabled an innovative approach to manage biodiversity data and have promoted a significant transformation in how biodiversity information is shared (BISBY, 2000; HEIDORN, 2011). This transformation led to the emergence of Biodiversity Informatics. Contrasting with Bioinformatics and Computational Biology, which are universally applied to molecular biology applications (HUERTA *et al.*, 2000), Biodiversity Informatics is the application of information technologies to the management, discovery and analysis of biodiversity data (SOBERÓN; PETERSON, 2004; CANHOS *et al.*, 2004; BERENDSOHN *et al.*, 2011).

Biodiversity Informatics has transformed how biodiversity information is shared and led to the development of an infrastructure to allow biodiversity data to be used to address complex questions about life on Earth. Despite considerable progress, biodiversity science is still reliant on data (HARDISTY; ROBERTS; The Biodiversity Informatics Community, 2013), especially data adherent to the Findable, Accessible, Interoperable, Reusable (FAIR) guiding principles (WILKINSON *et al.*, 2016). Jones *et al.* (2006) commented when describing the goals of a new bioinformatics (i.e., *biodiversity informatics*): “it is undeniable that vast funds are expended on data creation and acquisition. It is false economy, and poor scientific practice, not to ensure that the data are present and useful to all users in the future”. However, data sharing is not a guarantee of data reuse, and despite improvements, data reuse is still an important issue in ecology, mostly because data authors and information system designers do not follow best practices for data structure, metadata annotation and data licensing (WHITE *et al.*, 2013).

Thus, sharing data is just one of the pillars of Biodiversity Informatics, and more broadly of the Open Science framework. Data reuse requires that sufficient metadata describing the contents of underlying data is available, preferably using a metadata standard. Metadata provides an abstraction level necessary to capture the information content of the underlying data, independent of the representational details. This is part of a fundamental premise, but not sufficient, for data interpretation. Additionally, data aggregation is dramatically simplified by the adoption of standards (ROCCA-SERRA *et al.*, 2015). Standards and tools are needed to structure and document data. Community-specific vocabularies and ontologies are the foundation for the definition of concepts within a domain and for a community of practices to share data unambiguously. Regarding data interoperability, standardization is the key. Standards ensure a uniform data representation, enabling data aggregation from different sources while reducing data losses and duplication (BERENDSOHN *et al.*, 2011).

Without a common data reporting schema, data exchange may require complex processing and transformation. Furthermore, the lack of a shared vocabularies and ontologies can lead to a phenomenon called “ontological drift” (THOMPSON, 2011; WANG; SCHLOBACH; KLEIN, 2011), in which data becomes distorted as it moves across semantic boundaries. Beyond *administrative* metadata (POMERANTZ, 2015), metadata should describe the contents of the data, explaining the measured attributes, their names, units, precision, data schema and provenance (i.e., *descriptive* and *structural* metadata). Metadata records enriched by adoption of domain-specific ontologies and vocabularies provide the semantic underpinning which enables data reusability (CRYSTAL-ORNELAS *et al.*, 2021; THANOS, 2017).

A prerequisite for data reusability is *exchangeability*. Data *exchangeability* is the of ability of two parties to exchange datasets, and three types of heterogeneity must be addressed: *syntactic*, *structural* and *semantic* (THANOS, 2017). A solution for data exchangeability is *mediation* (THANOS, 2014), and it is the solution adopted by many biodiversity data portals, such as GBIF and Global Biotic Interactions (GloBI). By implementing mediation systems, data authors are relieved of the need to individually provide standardized data. A domain-specific abstraction layer (ROBERTSON *et al.*, 2014; SALIM; SARAIVA, 2020) is provided to data authors, and the mediation system takes care of standardization and metadata annotation when importing or exchanging data.

The adoption of standards has boosted research in many scientific fields (ROCCA-SERRA *et al.*, 2015). In the life sciences, the Minimum Information About a Microarray Experiment (MIAME) was the first community metadata standard adopted (BRAZMA *et al.*, 2001). Later, the MicroArray Gene Expression Tabular (MAGE-Tab) was among the first machine-readable (meta)data exchange format created in the life sciences domain (RAYNER *et al.*, 2006). From that point, the use of ontologies, vocabularies of terms and standards has received great attention in the life sciences, especially with the creation of the Gene Ontology (The Gene Ontology Consortium, 2019), the BioPortal (NOY *et al.*, 2009), and more recently FAIRsharing (SANSONE *et al.*, 2019).

FAIRsharing is a service which provides curated, informative and educational resources on data and metadata standards (SANSONE *et al.*, 2019), it currently contains 693 active and ready to use standards from 899 standards in the “Biology” subject. FAIRsharing uses the FAIRsharing Subject Ontology (SRAO) for subjects annotation (FAIRsharing.org: SRAO, 2022). To classify the standards based on subjects, a search was conducted in the SRAO for each original subject annotated in the standard. The objective was to identify all parent terms and select the immediate children of the *Biology* (`http://purl.obolibrary.org/obo/NCIT_C16345`) as the new subjects for the respective standards. Figure 2 shows the distribution of the SRAO subjects across the standards. The analysis revealed that the majority of the standards are associated with subjects such as molecular biology and medical sciences (e.g., Biomedical Science, Omics, Genetics). In contrast, only 121 standards (17%) are related to the biodiversity (e.g., Ecology, Zoology, Botany). It does not imply that a standard classified in the Ecology subject, or any other related subject, is useful in the biotic interactions domain. Despite the number of available resources (e.g., vocabularies, thesauri, ontologies) there is currently a lack of general guidelines, formal standards, or even informal guides for the standardization of biotic interactions data. However, the biodiversity community can make use of the numerous available resources, which can be adopted or extended, to address the standardization of biotic interactions data.

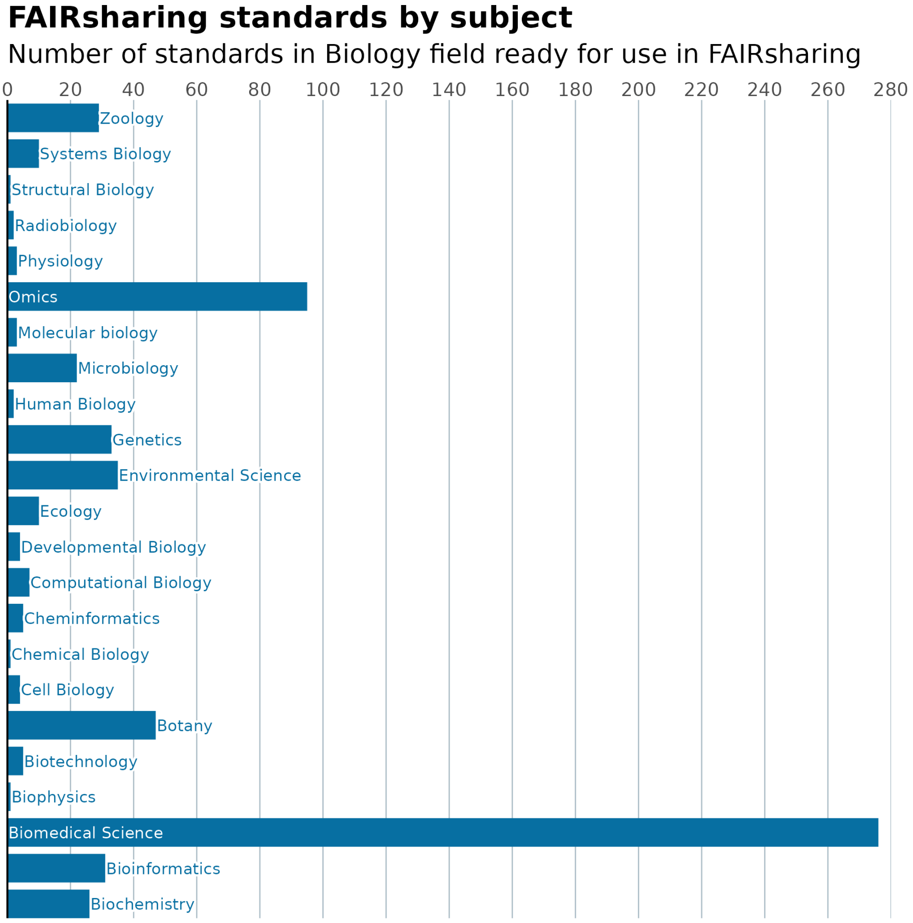


Figure 2: Number of standards for each Biology subfield ready for use in the FAIRsharing platform

Source: FAIRsharing (FAIRsharing.org: SRAO, 2022)

The main organization responsible for developing and promoting adoption of standards in biodiversity is the Biodiversity Information Standards (TDWG). The TDWG community is organized in different Interest Group (IG) (e.g Biological Interactions Data[1], Citizen Science[2], Species Information[3]), which in turn may contain multiple task groups with a dedicated output. The ratified and currently in use standards (i.e., status *current standard*) have dedicated Maintenance Group (MG), responsible for managing suggested changes to standards, providing usage guidelines and examples, and ensuring the preservation and stability of metadata related to components of the standard. Currently, there are 15 IG's and 5 MG organized under TDWG authority. The Darwin Core (DwC) standard (WIECZOREK *et al.*, 2012) is mature and among the most widely used biodiversity standards ratified by TDWG. DwC is predicated on occurrence (physical or observa-

[1] https://www.tdwg.org/community/interaction/
[2] https://www.tdwg.org/community/citizen-science/
[3] https://www.tdwg.org/community/species/

tional) data, but it has been extended to incorporate other types of data (PEARSON *et al.*, 2021; NILSSON *et al.*, 2022; ENDRESEN; GAIJI; ROBERTSON, 2009; FINSTAD *et al.*, 2020). It has been used by several platforms like Barcode of Life[4], Integrated Digitized Biocollections (iDigBio)[5], and Global Biodiversity Information Facility. In addition to the DwC, there are other resources which can be used to document and standardize biotic interaction data. OBO Relations Ontology (RO) "is a collection of relations intended primarily for standardization across ontologies" (MUNGALL *et al.*, 2023), and it is in adopted by GloBI to standardize the interaction types. Beyond recording the "what, where, when, how and whom" of an organism's occurrence, when recording biotic interactions it is common to include other information components which are context-dependent (e.g., the organism's traits, interaction effects and outcomes). For plants, ontologies such as the Plant Ontology (CONSORTIUM, 2002), the Plant Phenology Ontology (STUCKY *et al.*, 2018), the Plant Trait Ontology (ARNAUD *et al.*, 2012) and the Plant Phenotype Ontology (HOEHNDORF *et al.*, 2016) provide valuable resources for plant data standardization and annotations. There are also ontologies specific to insects, such as the Hymenoptera Anatomy Ontology (YODER *et al.*, 2010) and the Lepidoptera Anatomy Ontology (KOCH *et al.*, 2018), as well as ontologies for vertebrates, such as the Vertebrate Trait Ontology (PARK *et al.*, 2013) and the Mammalian Phenotype Ontology (SMITH; EPPIG, 2009). Additionally, there are more general ontologies like Uberon (multi-species anatomy ontology) (HAENDEL *et al.*, 2009) and the Environment Ontology (BUTTIGIEG *et al.*, 2013).

Therefore, the creation of a common schema and adoption of available ontologies and vocabularies have the potential to increase not only the access to biotic interactions data, but equally important, the increase in data reuse and interoperability.

2.2.1 The Darwin Core standard

The Darwin Core standard was ratified by TDWG in 2009, since then, it has been the main standard used by different communities to document and share occurrence data. The last version of DwC list of terms (2021-07-15) contains 179 terms divided into 11 classes. The DwC reuses terms from the Dublin Core Metadata Initiative (DCMI) (DCMI Usage Board, 2020) defined in two namespaces (`dc:` and `dcterms:`) and provides definitions of terms for literal objects (string, `dwc:`) and non-literal objects (Internationalized Resource Identifier reference, `dwciri:`).

[4] https://ibol.org/
[5] https://www.idigbio.org

In its simplest form, called Simple Darwin Core, the DwC terms are used to standardization of flat data (e.g., spreadsheet, database table). In Simple Darwin Core, no structure assumption beyond the concept of rows and columns is assumed, so, it is not possible to document complex data relationships (e.g., one-to-many relationships). For complex data, the DwC can be used with Resource Description Framework (RDF), Extensible Markup Language (XML) or Darwin Core Archive (DwC-A)). As part of the standard, the DwC provides implementation guides for each one of these reporting schemas.

Since biotic interactions are naturally relationships between two (or more) organisms, their representation in Simple Darwin Core is not possible ("a field name MUST NOT be repeated in a [DwC] record"). For that reason, documenting biotic interactions data requires a reporting schema where the relationships between interacting organisms can be represented without ambiguity. Despite DwC standard already defines terms to represent general aspects of biotic interactions (the "tetranomials"), there is no guide or consensus in the community how it should be documented. As will be discussed in Section 4.4.1, this diversity of schemas to represent biotic interactions data leads to confusion, loss of information and requires complex data interpretation and transformation.

The DwC is a standard "for sharing data about biodiversity – the occurrence of life on earth and its associations with the environment" (WIECZOREK *et al.*, 2012), but it inevitably lacks terms to cover details of specific sub-disciplines (e.g., genetic resources, biotic interactions, biological inventories). To overcome these gaps, the standard was designed to allow the creation of extensions. A Darwin Core Extension consists of additional terms or guidance for data documentation of specific subdomains of biodiversity (WIECZOREK *et al.*, 2012). Since its ratification, many extensions have been created by different communities to improve access to standardized data (BRENSKELLE *et al.*, ; ENDRESEN; GAIJI; ROBERTSON, 2009; POOTER *et al.*, 2017; SCHNEIDER *et al.*, 2019).

3 BIOTIC INTERACTIONS TERMINOLOGY AND ITS USE IN THE SCIENTIFIC LITERATURE

The “biotic interactions” term and other related terms such as “species interactions”, “inter(-)specific interactions”, “ecological interactions”, “community interactions” and “biological interactions” (hereinafter collectively referred to as *interaction terms*) are frequently adopted in the ecological scientific literature. Nakazawa (2020) has conducted a comprehensive examination of the demerits of using these different terms to describe (possibly) the same concept. This study conducted a text analysis to examine the concepts associated with the aforementioned terms and the specific contexts in which they are used by the research community. Moreover, a text-mining approach was employed to assess whether these terms are used interchangeably as synonyms in the field of Ecology and to determine if they represent similar or equivalent concepts.

3.1 Introduction

3.1.1 Bag of words and numerical representation of textual documents

Natural Language Processing (NLP) is a field in artificial intelligence which combines knowledge from linguistics and computer science, especially machine learning, to create models and methods for processing and understanding human language (CHOWDHARY, 2020; NADKARNI; OHNO-MACHADO; CHAPMAN, 2011). It involves a range of techniques and methods to process and analyze natural language data, supporting the identification of contextual nuances within documents. In NLP, a document refers to any unit of text which is considered for analysis. The information obtained in the analysis can be used to classify and organize the documents according to their similarities.

A common approach to representing textual data in a numerical format is the Bag of Words (BoW) model. The BoW is a simple and effective method, which treats each document as a "bag" of unrelated and unordered words. It consists in the elaboration of a dictionary with all unique words present in a collection of documents (i.e., corpus) and represents each document as a vector of word frequencies (BEHESHTI *et al.*, 2022). This numerical representation of the documents is used in Machine Learning for classification and information retrieval.

The numerical representation usually involves preprocessing steps for tokenization (breaking down sequences of text into smaller units, such as words, called tokens), stemming (reducing words to their root by removing affixes) and lemmatization (determination of the canonical form of the words or lemma). Since some words are likely to appear more than others in natural language data, removal of *stop words* (common words, e.g. "the", "and", "is") is commonly included in the preprocessing. Despite variations in tokenization approaches (e.g. word tokenization, subword tokenization, character tokenization), in this chapter a "token" and "word" are treated as synonyms in the context of NLP.

There are mainly two different two-dimensional matrix representations of a "bag of words". The numerical data can be represented in a Document-Term Matrix (DTM), where the rows represent documents and the columns represent words, or in its transpose form, a Term-Document Matrix (TDM), where the rows represent words and the columns represent documents. The DTM focuses on documents and their associated words, while TDM focuses on words and their occurrence in different documents.

Despite word frequencies in each document can be used to create DTM or TDM representations, it may lead to limitations and potential drawbacks. The word frequencies have a bias towards words that are common across the documents, but which do not carry significant meaning to the understanding of the document (SALTON; BUCKLEY, 1988). Additionally, it is incapable of capturing the importance of a word in the whole collection of documents, and it is sensitive to the length of documents. The word frequencies alone may introduce noise and overestimate words that appear frequently within a document but have limited relevance in the corpus. An alternative is the Term Frequency–Inverse Document Frequency (TF-IDF) approach, where a weighting factor is used to calculate the importance of words within a corpus (SALTON; BUCKLEY, 1988). There are several variations of TF-IDF calculation, but the basis is the multiplication of the term frequencies within documents (TF, $tf(t,d)$) by the inverse document frequencies (IDF, $idf(t,D)$), the word frequencies in the whole corpus (Equation 3.1). While the $tf(t,d)$ captures the word frequencies within documents (local importance), the $idf(t,D)$ captures the significance of the terms across the corpus (global importance). The inverse document frequency is used such that it penalizes words that appear in many documents and assigns higher weights to words that are less frequent in the corpus. The main idea is that rare terms are more informative than common terms. Thus, the higher the TF-IDF for a term, the more relevant it is to the document within the corpus.

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \tag{3.1}$$

3.1.2 A brief introduction to clustering

Clustering is a simple and effective method for discovering meaningful patterns and similarities in textual data. *Clustering* may reveal how the documents and words within a corpus are inter-related and support the identification of groupings in the data. Documents and words which are more or less related tend to be grouped depending on their similarities calculated over the numerical representation of the textual data in the BoW model.

Among the clustering algorithms, *k-means* (MACQUEEN, 1967; LLOYD, 1957) is widely used to find *clusters* in textual data (HOTHO; NÜRNBERGER; PAASS, 2005; SHUKLA; NAGANNA, 2014). Hierarchical clustering methods and density-based clustering are also applied (ALIGULIYEV, 2009; LOMAKINA; RODIONOV; SURKOVA, 2014; MURTAGH; CONTRERAS, 2012).

The *K-means* algorithm aims to partition a dataset into a predefined number of clusters (k) while it searches for centroids (i.e., clusters centers) that minimize the distance between data points within each cluster and maximize the distance between different clusters (LLOYD, 1957; MACQUEEN, 1967). The determination of the best number of clusters involves multiple executions of the algorithm with different values for k and the selection of the value which employing different methods and evaluation metrics, such as, silhouette score (ROUSSEEUW, 1987), calinski-harabasz index (CALIŃSKI; HARABASZ, 1974), davies-bouldin index (DAVIES; BOULDIN, 1979) and domain knowledge.

In hierarchical clustering, clusters are iteratively merged or divided based on their similarities to create a tree-like structure, called a dendogram, that represents the relationships between clusters (LOMAKINA; RODIONOV; SURKOVA, 2014). In Biology, dendrograms are frequently used to depict a phylogenetic hypothesis. There are two main approaches to hierarchical clustering: an agglomerative approach and a divisive approach (OAKES; JI, 2012). Agglomerative clustering starts with each data point as an individual cluster, and the algorithm progressively merges similar clusters until a stopping criterion is met. In divisive clustering, the opposite approach is considered. It starts with a single clutter containing all data points and recursively splits clusters into smaller, more similar, subclusters until a stopping criterion is met. There are various methods to calculate similarities between clusters, and different method can lead to different cluster structures. Thus, it is common to apply more than one linkage criteria and compare the results according to a combination of goals, interpretability and evaluation metrics.

A crucial aspect of clustering is the distance metric or similarity measure used to calculate the proximity of the data points in multidimensional space. The choice of the distance metric involves various aspects of the data and objectives of the clustering. Despite the Euclidean distance being a suitable metric for continuous and numeric features, in NLP the cosine similarity is commonly used (AGGARWAL, 2015; DEERWESTER *et al.*, 1990; MIKOLOV *et al.*, 2013). Cosine similarity measures the cosine of the angle between two non-zero vectors (e.g. two documents in a DTM) to determine the similarity among them.

The Euclidean distance is sensitive to the "curse of dimensionality", meaning that as the number of dimensions (e.g., tokens in a DTM) increases, the distance between data points tends to become less meaningful. In contrast, the cosine similarity measures the angle between the vectors representing the data points, which is unaffected by the increase in dimensionality (SCHUTZE; MANNING; RAGHAVAN, 2008). Textual data are often sparse (i.e., most documents contain only a subset of all possible tokens), and Euclidean distance tends to be dominated by the large number of zero values, leading to less meaningful distances. On the other hand, the cosine similarity is more robust to sparse data, since it ignores zero values. Instead of using the cosine similarity directly with clustering algorithms, it can be transformed to the complement of cosine similarity (cosine distance: $1 - S_C(d_i, d_j)$). Since the token frequencies are always non-negative, the cosine similarity between two documents will fall within the interval of $[0, 1]$. A value closer to 1 indicates a higher degree of similarity between the documents in terms of their content. Conversely, when using the cosine distance, two identical documents will have a distance of 0, while the maximum distance will be 1.

3.1.3 Topic modelling

Topic modeling is a statistical and computational technique used to uncover latent topics within a corpus. In topic modeling, the generative model Latent Dirichlet Allocation (LDA) (BLEI; NG; JORDAN, 2003) is widely used in various fields such has conservation (WESTGATE *et al.*, 2015), ecology (VALLE *et al.*, 2014; VALLE *et al.*, 2018), bioinformatics (KONIETZNY; DIETZ; McHardy, 2011; PRATANWANICH; LIO, 2014; ROGERS *et al.*, 2005; SHIVASHANKAR *et al.*, 2011) and biomedicine (HOSODA *et al.*, 2020; ZHAO; ZOU; CHEN, 2014). LDA is an unsupervised machine learning model which assumes that each document in the corpus can be represented as a probabilistic distribution over latent topics, and each topic is a probabilistic distribution over words. In LDA, Dirichlet priors are assumed for both the document-topic and topic-word probability distributions.

There are two main groups of algorithms for LDA: Gibbs sampling (GRIFFITHS; STEYVERS, 2004) and variational inference (BLEI; NG; JORDAN, 2003). Sampling-based algorithms, as Gibbs sampling, are based on Markov Chain Monte Carlo (MCMC) methods to iteratively sample from the joint posterior distribution and gradually converging to a stationary distribution. In the Variational Expectation-Maximization (VEM) algorithm, rather than approximating the joint posterior distribution with samples, the inference problem is transformed into an optimization problem to approximate the log-likelihood. Despite VEM being faster and computationally less intensive than Gibbs sampling, it may underestimate the variance of the posterior (BLEI; KUCUKELBIR; MCAULIFFE, 2017).

Similar to *k-means* clustering, the number of topics (k) must be set a priori, but choosing the best number of topics for a given corpus is not trivial. In order to assist the selection of k many evaluation metrics have been proposed, such as *perplexity* and coherence score (RÖDER; BOTH; HINNEBURG, 2015). Similar to choosing the number of clusters, the number of topics usually involves the combination of multiple metrics and contextual knowledge.

3.2 Material and Methods

This section outlines the methodology employed for text-mining scientific publications in Ecology. The focus is the evaluation of the usage of the *interaction terms*, specifically in terms of the contexts in which they are employed, to identify whether these terms are used interchangeably as synonyms or if they are employed to represent different concepts.

3.2.1 Data used

The scientific publications were obtained from Digital Science's Dimensions platform (Digital Science, 2018; HOOK; PORTER; HERZOG, 2018). The Dimensions platform was selected because it is recognized as the "largest and broadest indexer of scientific documents" (BASSON *et al.*, 2023). The platform was searched, individually, for exact matches of each *interaction term* on publications' title and abstract. The obtained results were subsequently filtered to include only scientific articles within Ecology and written in English language, published between 1970 and 2022.

To mitigate any potential bias resulting from the occurrence of multiple *interaction terms* within the same publication, duplicated publications obtained from different *interaction terms* queries were removed and not included in the analysis.

The publications were divided into six classes according to the *interaction term* used to retrieve them from Dimensions platform. The classification of publications serves as "ground truth" for the analysis of the use of the *interaction terms* across different classes. Hereafter, the set of publications returned by the same *interaction term* is referred as *class* of publications. Indeed, this approach enables a focused analysis that facilitates the examination of similarities and differences among publications based on their associated *interaction terms*.

3.2.2 Preprocessing data

The title and abstract of the publications were combined to create a collection of documents (corpus). As part of the preprocessing phase, the text was tokenized to break it down into individual words (tokens). Following tokenization, stop-words were removed. Subsequently, lemmatization was applied to the remaining tokens to reduce them to their lemmas. To refine the preprocessing, a customized list of common words, including all the *interaction terms*, was used to remove frequent words such as, "species", "community", "result". The lemmas were used to generate bi-grams and tri-grams, capturing two and three consecutive words occurring together in a text, respectively. The combination of adjacent words can facilitate the interpretation of results, since combined words can provide meaningful information about the underline concept otherwise not captured by single words (e.g., 'species' and 'species richness').

Previous studies have pointed out the benefits of utilizing only nouns in topic modeling (MARTIN; JOHNSON, 2015; WELBERS; KASPER, 2019). Thus, a Part-of-Speech

(POS) tagging process was implemented to assign a syntactic category for each token. Tokens that were not tagged as nouns or adjectives were subsequently removed. Since the objective is to access the similarities and dissimilarities between the documents regarding their biological contexts, nouns and adjectives are most likely to provide insights about their contents. Additionally, tokens (lemmas, bi-grams and tri-grams) that occurred in fewer than five documents or in over 90% of the documents were also removed.

Two DTM were generated for analysis. The first DTM used in the clustering analysis utilized the TF-IDF weighting scheme with a logarithmically scaled IDF, as described in Equation 3.2. The second DTM used in the topic modeling employed just the (normalized) word frequencies, as shown in Equation 3.3.

$$idf(t, D) = log_2 \frac{|D|}{|d \in D : t \in d|} \tag{3.2}$$

where:

$|D|$: total number of documents in corpus
$|d \in D : t \in d|$: number of documents where the term t appears

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{3.3}$$

where:

$f_{t,d}$: number of occurrences of the term t in the document d
$\sum_{t' \in d} f_{t',d}$: total number of terms in the document d

3.2.3 Clustering publications based on cosine similarity

In the clustering analysis, the DTM of all documents and tokens was summarized by grouping the documents according to their classes (i.e., *interaction terms*). Since the objective is to evaluate the structure and similarities of the different results returned by querying Dimensions for each *interaction term*, this approach treats each set of publications as a single document composed by the union of publications contents in the set. This “short” version of the DTM has six rows, one for each *interaction term* and the same number of columns as the original DTM.

Both the DTMs were used with *k-means* and hierarchical clustering (available in R package `stats`, R Core Team (2023)). The number of clusters (k) in the *k-means* was evaluated from 2 to 50 and the value which resulted in the best Silhouette score was taken.

In the case of the clustering the original DTM, the returned clusters are used to evaluate if documents from different queries rely on the same cluster or not. If documents returned by the same query tend to form uniform clusters (clusters with documents from the same class), even if the number of clusters is greater than the number of *interaction terms*, it is possible to infer that there is separation between the context of the publications returned by the different queries. In other words, the *interaction terms* are used in different contexts with possibly different meanings. On the other hand, by clustering the "short" version of the DTM is possible to evaluate the organization and structure of the classes. In both cases, the cosine distance was used as the distance metric for the algorithms.

3.2.4 Topic modeling of publications

The topic modeling of the publications retrieved from Dimensions platform considered the DTM with normalized word frequencies, since TF-IDF is not necessary for estimating the probability distribution for topics in documents (BLEI; NG; JORDAN, 2003). The objective was to identify the number of topics in the corpus, but mainly, evaluate the distribution of the documents from different classes over the topics. Complementary to the clustering analysis, the topic modeling allows the identification of common topics between the publications of different classes. Ultimately, it offers an evaluation of the extent to which documents from different classes cover the same topics and provides insights into the nature of those topics. In the analysis, the number of topics (k) was evaluated between 2 and 20 topics, and the value of k which returned the best coherence score was set as the final number of topics. The topic modeling was implemented in Python using the `gensim` package (REHUREK; SOJKA, 2010).

3.3 Results and Discussion

A total of 14,907 articles were retrieved from the Dimensions platform when searching for the *interaction terms*. After filtering the results to contain only articles written in English and removing articles with different *interaction terms* within the same article (duplicated articles), the remain 13,808 publications were used in the subsequent analysis. Table 1 shows the number of publications for each *interaction term* (class of publications) and Figure 3 presents the usage of the terms along the years (publication date). During the period of 1970 to 2000, there was a relatively consistent usage of the *interaction terms*. However, starting around the 2000s, a notable divergence in their usage became more evident, specially for the terms “species interaction”, “biotic interaction”, “ecological interaction” and “inter-specific interaction”. The usage of remaining terms practically have not changed. The term “species interaction” is the most used and occurs in 36% of publications. Nevertheless, there is no clear trend indicating the fixation of this particular term.

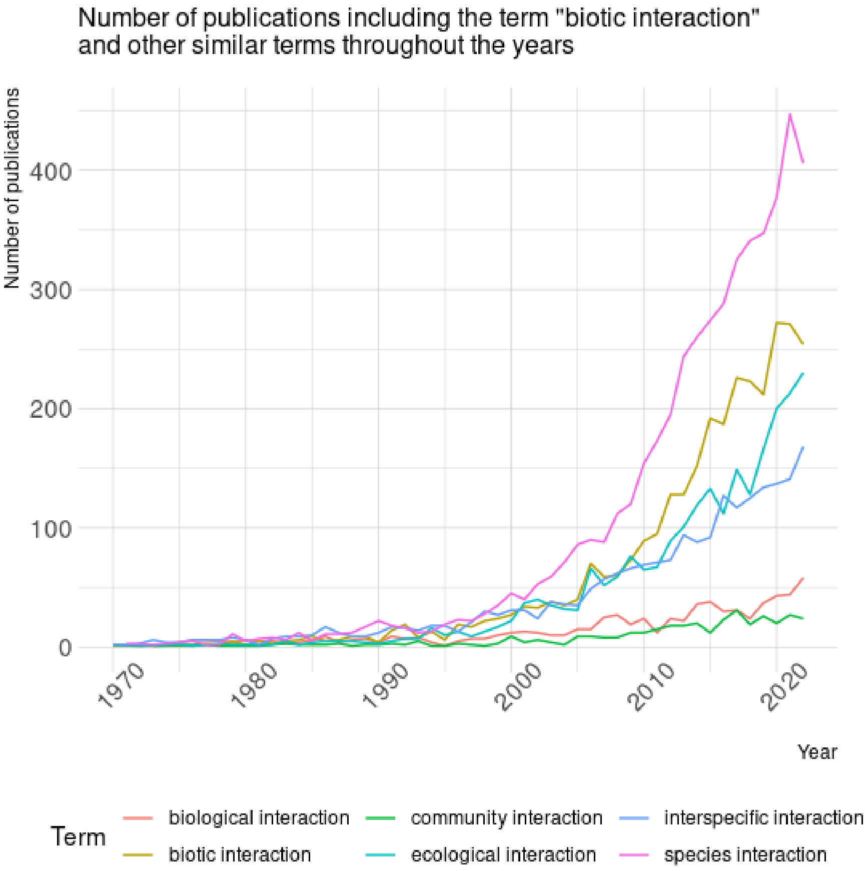


Figure 3: Number of publications including the term “biotic interaction” and related terms from 1970 to 2022.

The terms “species interaction(s)”, “biotic interaction(s)”, “ecological interaction(s)” and “inter(-)specific interaction(s)” appear in 91.6% of the publications, while “biological interaction(s)” and “community interaction(s)” appear only in 8.4% of the publications. Given the limited usage of the terms “biological interaction” and “community interaction” in scientific publications in Ecology, it is uncertain whether they are synonyms or if they denote distinct concepts associated with less studied subjects in Ecology.

Despite the lack of convention of nomenclature in Ecology (HERRANDO-PÉREZ; BROOK; BRADSHAW, 2014), an attempt for interpreting the *interaction terms* can be derived by considering the meanings of their constituent prefixes (e.g “biotic”, “species”, “ecological”). The term ‘’biotic interaction” can be understood as a broader concept that encompasses interactions among organisms of the same taxon (intraspecific interactions) and different taxa (interspecific interactions). It is closely related to the term “ecological interaction”. However, the term “ecological interaction” might be understood to also include interactions between organisms and their abiotic environment. “Species interaction” and “interspecific interaction” appear to be synonyms, given the similar etymology of the words “species“ (from Latin speciēs) and “specific” (from Latin speciēs + faciō), both terms are related to the interaction between taxonomically homogeneous groups of organisms. On the other hand, the term “community interaction” suggests that the interactions are limited to specific communities, which is, interactions between individuals of populations of different species coexisting in a particular area or habitat. Finally, the term “biological interaction” appears to be the most general of the terms, encompassing all kinds of interactions studied in Biology, including, for example, interactions involving biomolecules such as DNA and proteins. However, the usage of these terms spans from various subjects in Biology(NAKAZAWA, 2020).

Term	**Number of publications**
`species interaction`	4954 (35.9%)
`biotic interaction`	3154 (22.9%)
`ecological interaction`	2391 (17.3%)
`interspecific interaction`	2201 (15.9%)
`biological interaction`	731 (5.3%)
`community interaction`	377 (2.7%)
Total	**13,808 (100%)**

Table 1: Number of publications retrieved from Digital Science’s Dimensions platform for each of the *interaction terms*

3.3.1 Clustering the “interaction terms”

The hierarchical clustering considering the weighted word frequencies within each class of the publications partially corroborates with the interpretation of the meanings of the interaction terms by their prefix words (Figure 4). The terms “biotic interaction” and “species interactions” are more close related to each other than to the other terms, which suggests that the publications within these classes share a similar vocabulary compared to the other classes. Together with the term “ecological interaction”, they form a more general group which is related to the term “interspecific interaction”. The terms “biological interaction” and “community interaction” are in distinct branches of the dendogram of Figure 4, which indicates that these two terms present more distinct word frequencies and vocabulary from the other terms. However, this hierarchical clustering is based on the aggregated word frequencies extracted from publications within each class. In this context, classes with more publications inherently have a broader vocabulary. Consequently, the similarities observed between classes may be attributed to differences in the vocabulary sizes, where larger vocabularies are more likely to share the same words with similar frequencies solely due to their larger size. However, the TF-IDF normalization mitigates the impact of the word frequencies across different classes, the influence of vocabulary size on clustering results can not be eliminated.

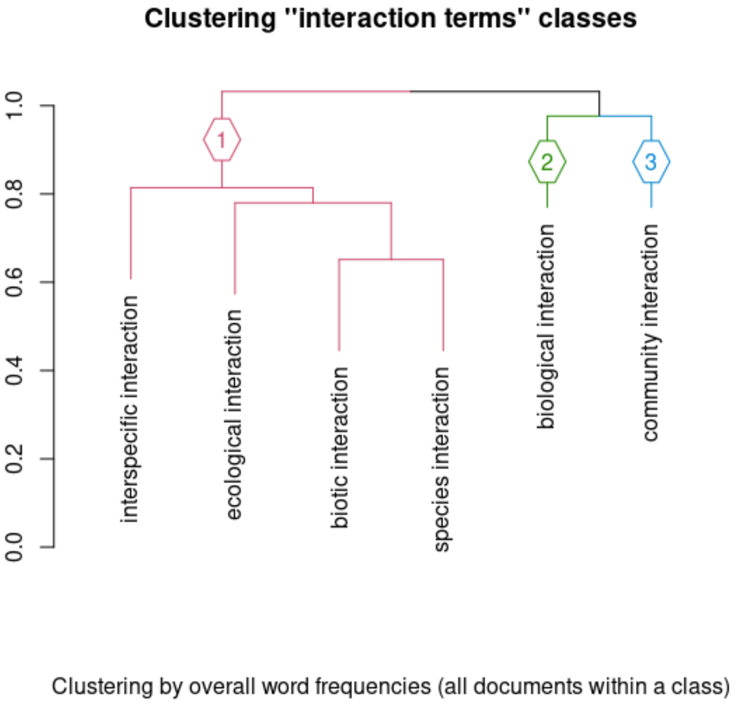


Figure 4: Dendogram from hierarchical clustering the “interaction terms” based on the word frequencies of each class.

The high-dimensionality of the DTM can be reduced to a 2-dimensional space by taking the first two principal components (PC) of the orthogonal projection of the Principal Component Analysis (PCA) (JOLLIFFE, 2014; PEARSON, 1901). The first PC represents the direction of maximum variance in the data, and the second PC captures the orthogonal direction of the second-highest variance. These two PCs were used to create the plot of Figure 5, which preserves the pairwise dissimilarities between the classes of *interaction terms*. Applying the PCA on the same DTM used for hierarchical clustering, a similar pattern is observed. However, in the 2-dimensional space the terms “biological interaction” and “community interaction” appear far from each other, whereas the remaining terms are relatively close. The distance between the terms “biological interaction” and “community interaction” in the PCA plot suggests that the publications associated with each class tend to possess distinct vocabularies, which are most dissimilar from the other terms.

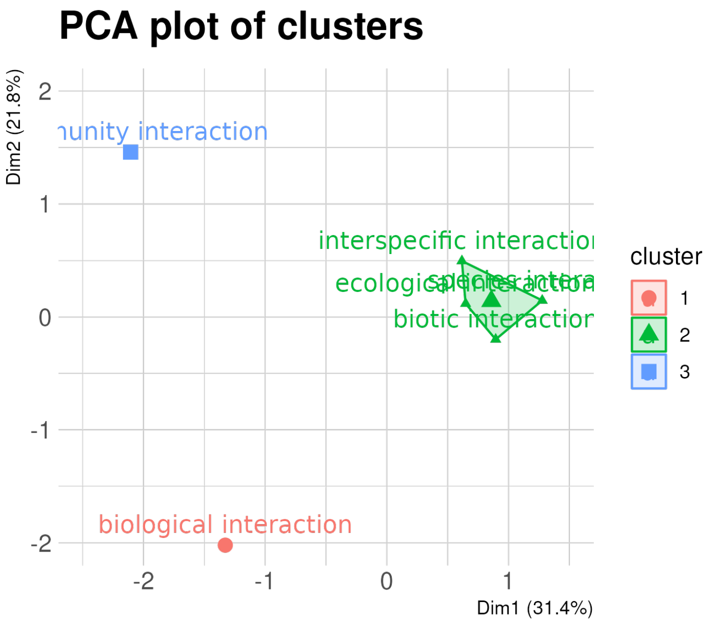


Figure 5: Principal Component Analysis (PCA) plot: the two principal components showing the relationships between the "interaction terms" word frequencies.

When considering the DTM of weighted word frequencies (TF-IDF), it was observed that the highest silhouette score was achieved with 8 clusters (Figure 6). However, it is important to note that the resulting silhouette score was 0.007, indicating the presence of overlapping or ambiguous clusters within data as illustrated in Figure 7. Other cluster configurations (2, 3, and 4 clusters) also displayed scores that were slight positive but lower than the configuration with 8 clusters.

The configuration of 8 clusters did not show a clear separation of publications within the classes (Figure 7), since there is great overlap between the clusters (left plot in Figure 7). As illustrated in the right plot of Figure 7 and, separately, by each class in Figure 8, publications from all classes are assigned indistinctly to all clusters, indicating that there is no one-to-one relationship between the clusters found and the document classes. In addition, none of the classes are dominated by publications from a particular cluster, indicating that no single cluster is exclusively representative of any specific class (Figure 9).

By computing the average TF-IDF values across all documents within the same cluster, it is possible to extract the most important words for each cluster. Table 2 presents

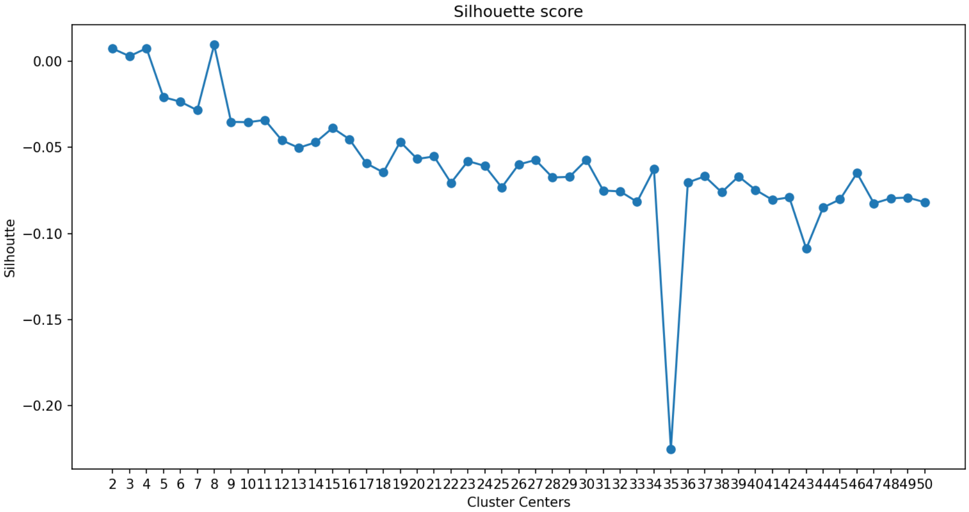


Figure 6: Silhouette score for each number of clusters used in the selection of the best number of centers. The maximum silhouette score is returned by 8 clusters.

the top five words with the highest average TF-IDF, providing insights into the vocabularies associated with each cluster. Despite the lack of clear separation between clusters, the analysis of the most important words can still provide insights into the specific topics captured by each cluster. However, while there may be distinct topics according to the most frequent words associated with each cluster, the similarities between documents from different clusters make it difficult to establish strict boundaries between them. It explains the high overlap between the clusters in Figure 7.

The application of *k-means* clustering indicates that there is not a clear difference in the vocabularies used across the publications. This implies that the *interaction terms* are being used within similar contexts. In addition, the hierarchical clustering reveals that some classes are more related than others, what can be explained by the high number of the documents in the overlapping region of Figure 7 and the smaller, but relevant, number of documents distant from the overlapping region. It is also emphasized by the individual plots in Figure 8, where the overall structure of the clusters in the 2-dimensional space shows similar structure to the dendogram of Figure 4.

K-means clustering creates partitions on a dataset and treats each document as a vector of a high-dimensional space represented by its TF-IDF scores. The algorithm relies on minimizing the sum of squared distances between documents and their assigned cluster centroids. It may produce similar results to the topic LDA but they employ different techniques to groups documents into groups or clusters.

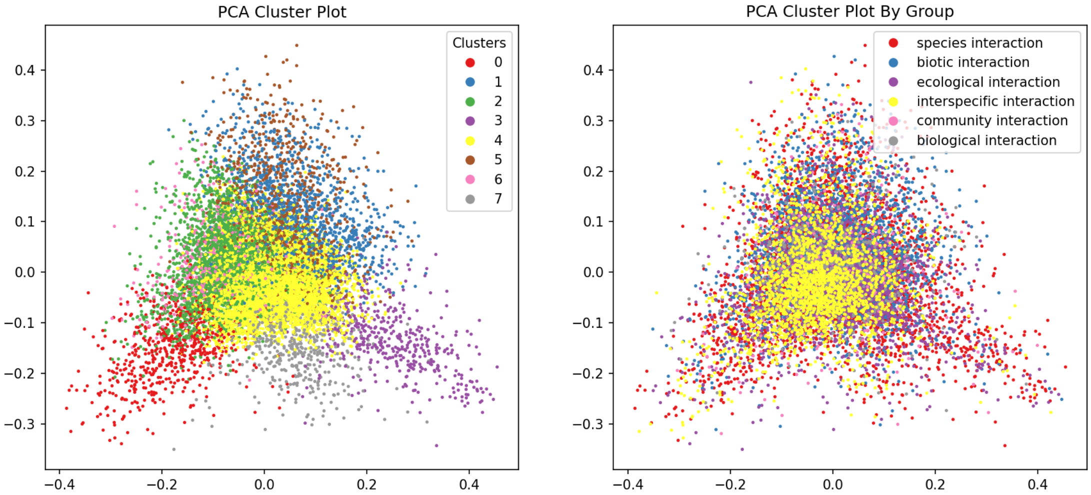


Figure 7: Principal Component Analysis (PCA) plot of two main PCs showing the distribution of publications, cluster assignments and distribution over classes

Left: Principal Component Analysis (PCA) plot of two main PC's showing the distribution of publications and clusters assignments (colors represent a cluster from 0 to 7). *Right*: Principal Component Analysis (PCA) plot of two main PC's showing the distribution of publications and the classes of "interaction terms" (colors represent a specific "interaction term").

Cluster	Top 5 words
Cluster 0	specialization, structure, extinction, island, network
Cluster 1	pollinator, parasite, native, fish, host
Cluster 2	forage, predation_risk, predation, prey, predator
Cluster 3	ecology, coexistence, trait, competition, dynamic
Cluster 4	pattern, assemblage, structure, distribution, diversity
Cluster 5	treatment, root, growth, tree, soil

Table 2: Most frequent (top five) words within each cluster.

3.3.2 Topic modeling

The hierarchical clustering analysis revealed that the *interaction terms* can be divided into three distinct clusters: one for the "biological interaction", another for the "community interaction", and a third cluster encompassing all the remaining terms. During the topic modeling, the grid search for the optimal number of topics and the hyperparameters (alpha and beta) was conducted. The alpha hyperparameter controls the distribution of topics in documents, where high values make the distribution of document-topics more uniform (documents contain a mixture of many topics). On the other hand, the hyperparameter beta controls the distribution of words in topics, where high values of beta result in a more uniform topic-word distribution (topics contain a mixture of many words). The

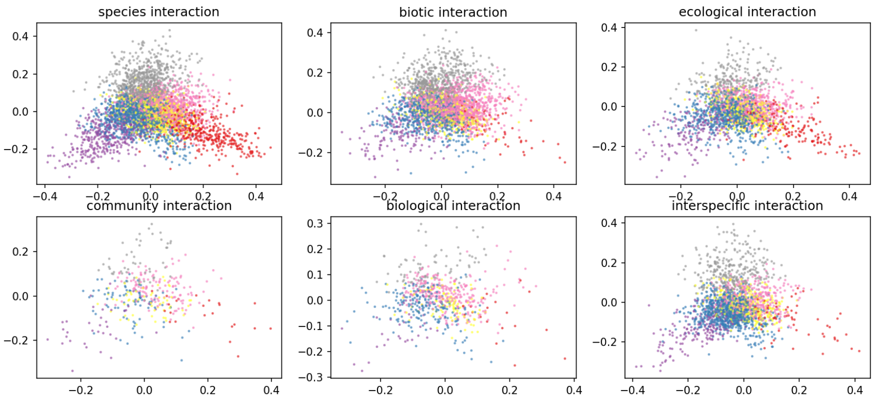


Figure 8: Principal Component Analysis (PCA) plot of two main PCs showing the distribution of publications and cluster assignments divided by "interaction term"

Colors represent a clusters from 0 to 7.

number of topics explored was from 2 to 20, with an increment of 2 topics at each execution. The hyperparameter values for alpha and beta were set to 0.05, 0.1, 0.5, 1.5, 10 and the default heuristic applied by the LDA algorithm of the Python *scikit learn* package (PEDREGOSA *et al.*, 2011). A total of 360 executions of the LDA were performed, and the execution with the highest Coherence score was selected (Figure 10). The best number of topics in the evaluated interval was 6 topics with an alpha of 1.5 and beta of 0.05.

Topic modeling shows an overlapping between the topics, which is the result of documents having similar probabilities for different topics. LDA aims to estimate the topic distribution for each document. Typically, a document covers multiple topics (e.g., conservation, restoration, climate change), and the LDA model will assign different probabilities to each of these topics. When a document is dominated by a specific topic, its probability will be greater than of the other topics. However, when the same documents are equally likely to address multiple topics, it leads to the overlapping observed in of Figure 11. In Figure 11, the best topic (i.e., the topic with the highest probability) for each document is chosen to represent the main topic of a document. The documents at the center (the overlapping region) are documents that, despite being assigned to the topic represented by its color, could potentially be assigned to other topics which are in this overlapping region. Conversely, the documents that are far from the center are those for which the probability of the best topics is sufficiently high to distinguish them from other documents belonging to different topics. In other words, the documents that are located at the center primarily cover multiple topics, whereas the documents located far from the center tend

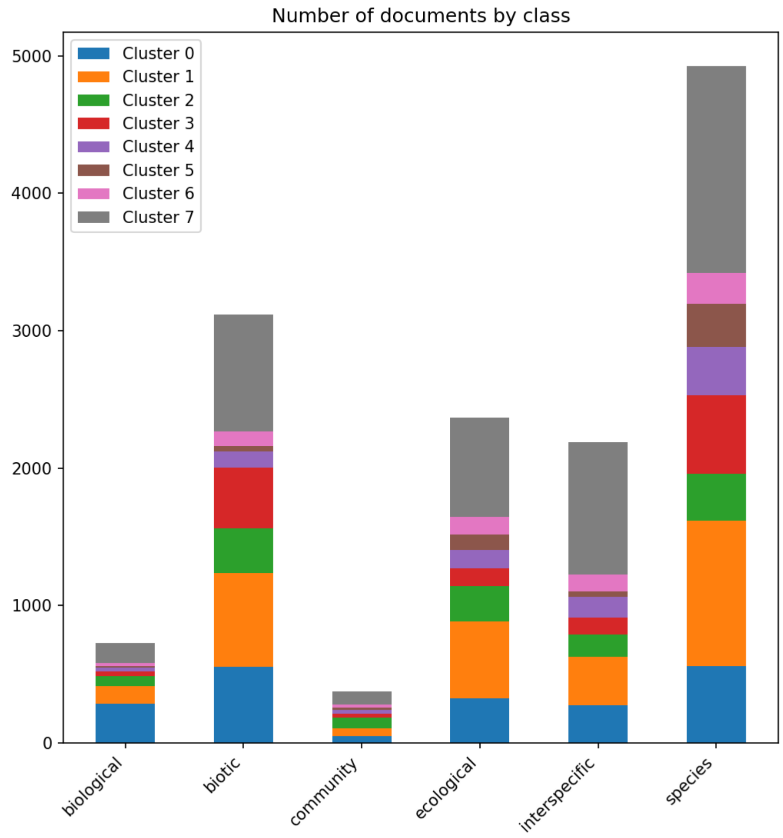


Figure 9: Distribution of clustered documents for each class

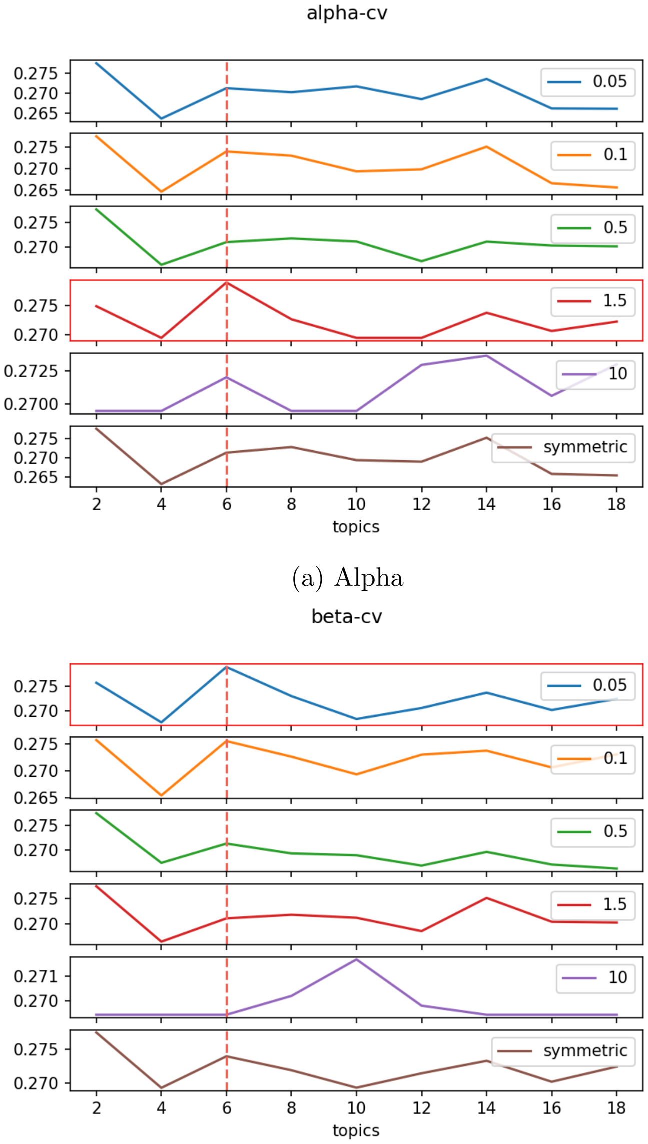
to focus on specific, singular topics.

Despite differences in the topics, some words appear in almost all the topics (e.g., diversity, dynamic, pattern) but with different contributions. This reveals that the topics involve similar subjects, but some specificity is given by their distinct words (Figure 12). Since the documents are from the same scientific field (i.e., Ecology) these common words across topics are expected.

Documents from the same class are assigned to different topics (Figure 13). It indicates that besides the existence of latent topics providing a relative separation of the documents into 6 topics, it does not reflect into the classes. The usage of *interaction terms* across the publications indicates that these terms are employed in publications with different, yet related, topics. It suggests that the *interaction terms* have been used interchangeably, regardless of the specific main topic of the publications.

Both the clustering analysis and the topic modeling did not provide evidence indicating contextual variations in the utilization of the *interaction terms* within scientific publications. This reinforces the notion that these terms are frequently employed interchangeably, despite potential distinctions that may exist between them. However, without a nomenclature that establishes formal definitions for these terms and their conceptual relationships, the true biological meanings underlying them may remain elusive. Although not explicitly explored here, the publications removed from the analysis, because of their use of different *interaction terms* in the same publication, are another evidence for the interchangeable usage of these terms.

Given the lack of a clear distinction between the *interaction terms*, the term “biotic interactions” will be used in this study as a general term without implying any differentiation unless explicitly specified.



(a) Alpha

(b) Beta

Figure 10: Coherence score for values of the hyperparameter alpha and number of topics

a) The plot highlighted in red was the alpha value which resulted in the maximal value for the coherence score (alpha = 1.5). b) Coherence score for values of the hyperparameter beta and number of topics. The plot highlighted in red was the beta value which resulted in the maximal value for the coherence score (beta = 0.05).

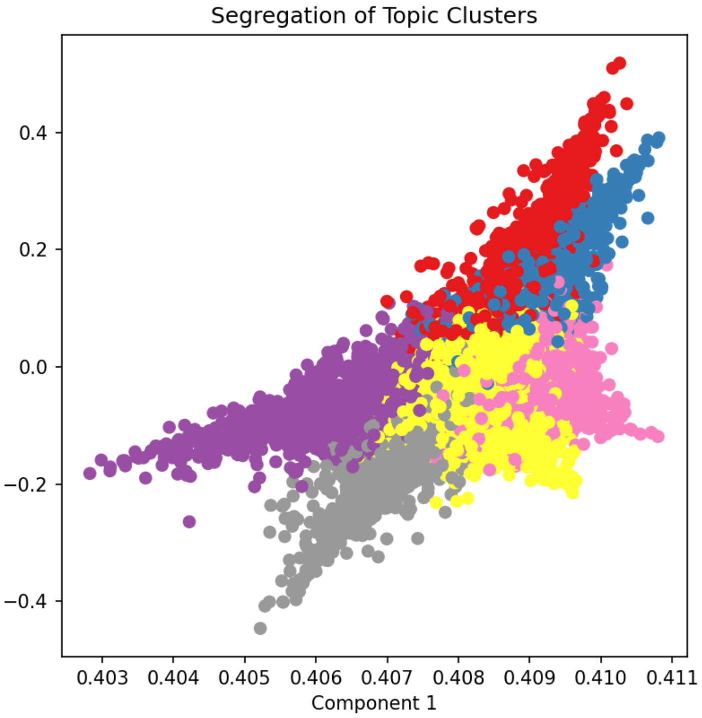


Figure 11: Distribution of the publications across the topics

(Each color represents a topic) after dimensionality reduction by truncated singular value decomposition (SVD).

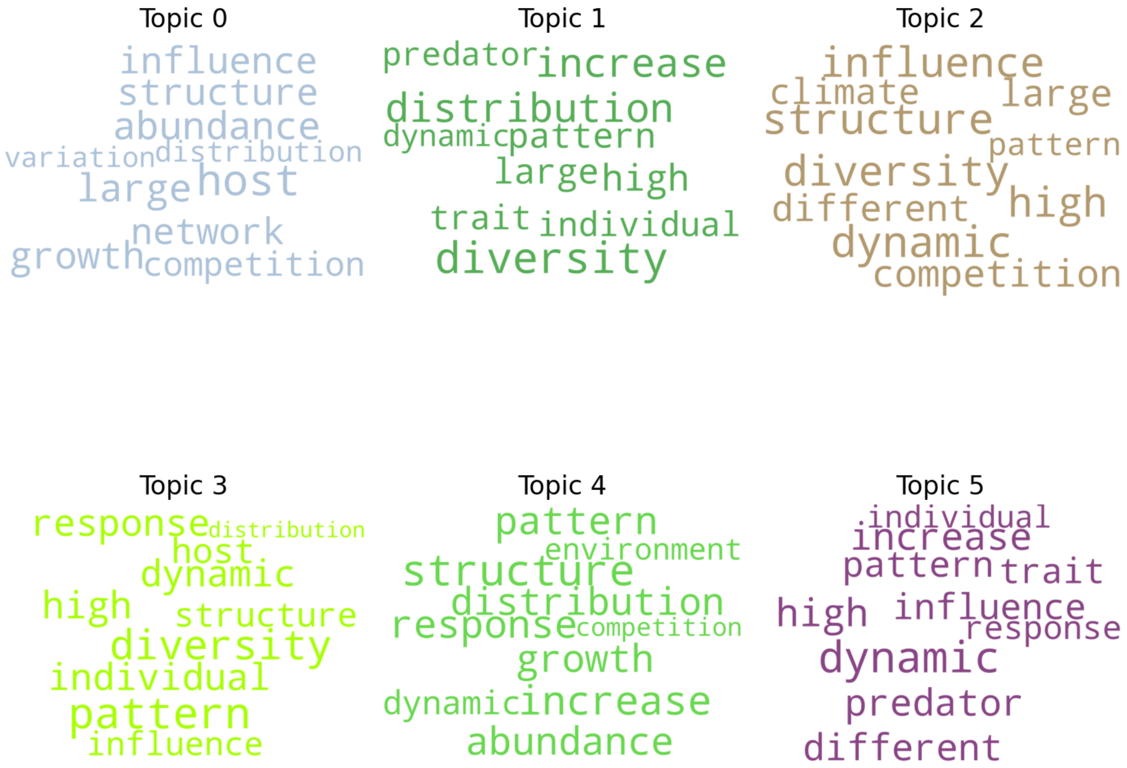


Figure 12: Word-clouds with the 10 most frequent words for each topic found in the LDA

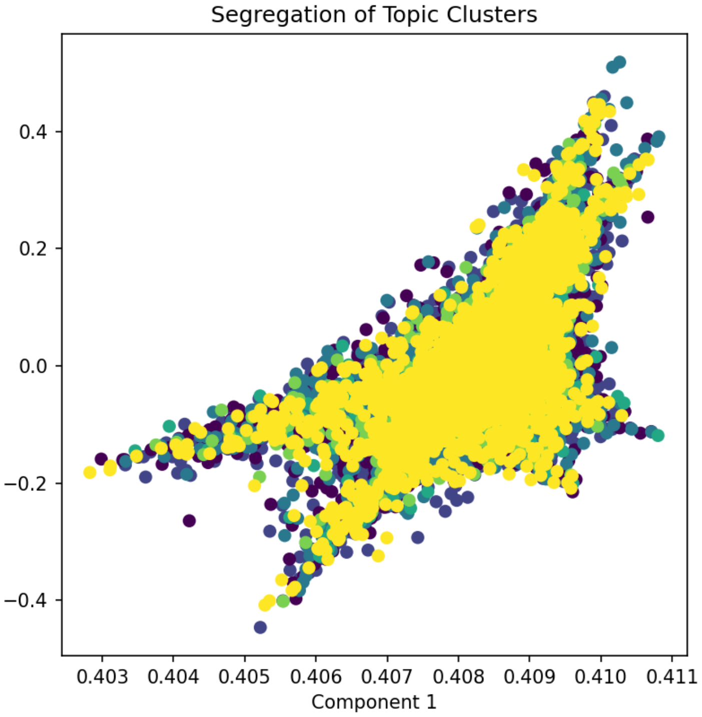


Figure 13: Distribution of the publications across classes of “interaction terms”

Each color represents a “interaction term” after dimensionality reduction by truncated singular value decomposition (SVD).

4 OVERVIEW OF BIOTIC INTERACTIONS DATA

This chapter aims to provide (i) an overview of available data on biotic interactions, (ii) an introduction to the Global Biotic Interactions, the main resource to access biotic interactions data, (iii) a preliminary analysis of biotic interactions in the Global Biodiversity Information Facility (GBIF), and (iv) a discussion of the major biotic interactions data formats and representations currently used by the biodiversity community.

4.1 Available data

Biotic interaction data have been collected and made available for decades in many formats and sources. Among these sources, the scientific literature is a prominent resource of biotic interactions data. Typically, data are commonly found as appendices or supplementary materials in publications, or either summarized in tables within the main text of the publications. However, there has been a notable shift recently and currently it is more common that scientific journals and funding agencies require, or even obligate, researchers to provide access to raw data as a prerequisite for publication or funding approval (e.g., Nature Announcement (Nature, 2016), National Science Foundation (NSF): Biological Sciences Guidance on Data Management Plans[1]). In addition, the establishment of international data sharing guidelines, such as the Findable, Accessible, Interoperable, Reusable (FAIR) (WILKINSON *et al.*, 2016), The Transparency and Openness Promotion (TOP) (NOSEK *et al.*, 2015) and the Beijing Declaration (CODATA, Committee on Data of the International Science Council CODATA *et al.*, 2019), reflects efforts to promote data sharing and accessibility. Furthermore, the creation of the Research Data Alliance (RDA) (BERMAN; WILKINSON; WOOD, 2014), a worldwide initiative involving over 13,000 members from 148 countries, plays an important role in "building the social and technical infrastructure to enable open sharing and re-use of data".

[1] https://www.nsf.gov/bio/biodmp.jsp

These combined efforts have contributed to increasing the availability of biotic interactions data in various data repositories. Despite the published datasets have facilitated the reproducibility of studies and experiments, data reusability and interoperability remains a challenge, primarily due to the lack of standardization and common practices for sharing data. Researchers often face laborious tasks of data compilation and transformation to make it compatible with their analyses and research objectives. Although, the biotic interactions data publishing is increasing (Figure 14), it is still difficult to reuse and aggregate data from different sources.

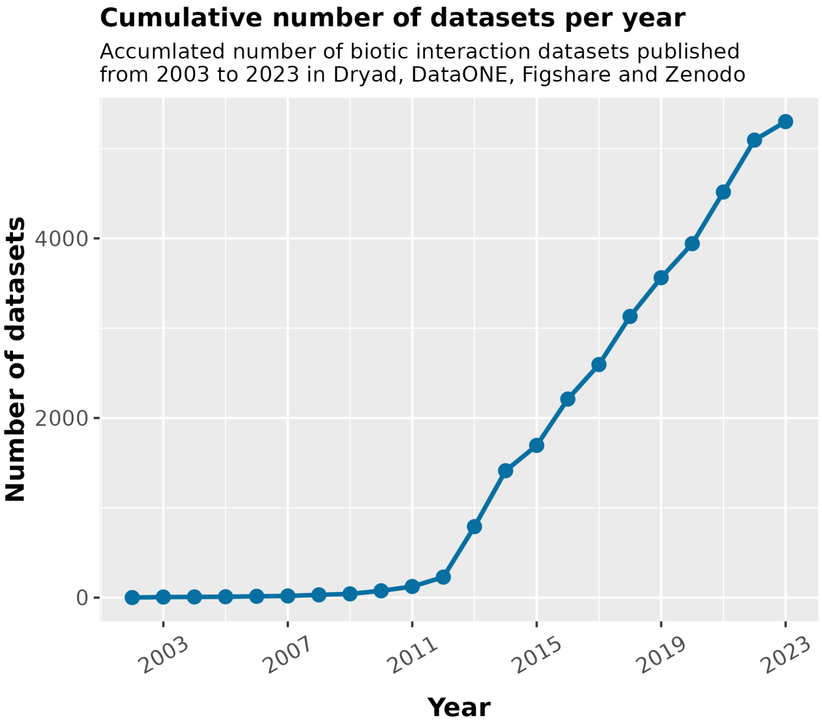


Figure 14: Accumulated number of biotic interaction datasets published from 2003 to 2023 in Dryad, DataONE, Figshare and Zenodo

Source: Dryad, DataONE, Figshare and Zenodo

Given the requirement for an increasing amount of data by meta-analysis and large-scale studies and the emergence of biodiversity databases, many initiatives, such as the Global Biodiversity Information Facility (GBIF), were created to centralize the access to biodiversity data. In the case of biotic interactions, the same phenomenon is observed, but on a smaller scale and in more decentralized coordination.

In order to understand the characteristics of biotic interactions data currently available, an exploratory analysis was conducted considering four data repositories commonly used to share research data. Metadata about datasets deposited in Dryad, Zenodo, Figshare and Data Observation Network for Earth (DataONE) were investigated for biotic interactions data.

DataONE (MICHENER *et al.*, 2011) is a network of interoperable data repositories for preserving, access, use, and reuse of multidiscipline ecological and environmental data. Currently, (June 2, 2023) there are 841,251 datasets published in the DataONE network, a number almost ten times greater than the total number of datasets in GBIF. It requires the member nodes to adopt a metadata standard, which must be registered in the DataONE. Member nodes of the DataONE network may opt to use Metacat (JONES *et al.*, 2001), a metadata and data management software which uses Ecological Metadata Language (EML) (JONES *et al.*, 2019) as the default metadata standard, to share and replicate metadata with the DataONE coordinate nodes.

Dryad is a manually curated open-access repository of research data, mainly of biology. It promotes data citations by assigning a unique, persistent, and resolvable Digital Object Identifier (DOI) for published datasets in the form of a DataCite DOI. The Creative Commons Zero waiver[2] is applied to all datasets, which makes the terms of reuse both clear and nonrestrictive. Dryad is a former member node of the DataONE networking, meaning that datasets metadata deposited in Dryad are shared with DataONE.

Zenodo is a general-purpose and multidisciplinary open repository developed under the European OpenAIRE program and hosted at European Organization for Nuclear Research (CERN). It has been paving the way in data citation and publishing by being the first data repository in promoting assignment of DOI to datasets. Although, Zenodo recommends the usage of open licenses, such as Creative Commons, it allows the attribution of different licenses and access levels. Dryad and Zenodo have formed a data sharing partnership, and Dryad stores a copy of all datasets on Zenodo to enhance data preservation.

Figshare is another general-purpose open-access repository which is widely used by many universities and scientific journals to share research and publications supplementary data[3]. All datasets are released under a Creative Commons license, CC-BY or CC0 (public domain).

Datasets were searched in each of the four repositories using the same *interaction terms* used in Section 3. The metadata of all datasets in the results were collected and organized in a common data structure. A total of 9,101 datasets were found containing at least one of the *interaction terms* in the title, description and keywords. The number

[2] http://creativecommons.org/publicdomain/zero/1.0
[3] ⟨https://knowledge.figshare.com/institutions#features⟩

of datasets for each repository is shown in Table 3. After eliminating duplicate datasets with identical titles and publication years, excluding versioned copies and datasets from Dryad int the results of DataONE and Zenodo, a total of 5,326 unique datasets remained. These datasets were then examined regarding of file types, licenses attribution and the occurrences of the *interaction terms*. Interestingly, the combined copies of Dryad datasets within the results of DataONE and Zenodo (1,748 datasets) surpasses the total number of datasets directly retrieved from Dryad. It could be attributed to the possibility that Dryad datasets replicated in DataONE and Zenodo contain the *interaction terms* in their metadata records, but the same dataset in Dryad has undergone modifications and no longer includes the *interaction terms*. As an example, see the dataset ⟨https://doi.org/10.5061/dryad.35rt5⟩ in Dryad repository and the respective dataset in DataONE[4]. In this example, the DataONE metadata record contains the keyword "species interaction" while the Dryad version contains only the keyword "amphibian decline". However, it may not be the only cause of the differences between the results.

DataONE (including Dryad datasets) and Figshare returned the highest number of datasets among the repositories. However, after eliminating duplicated datasets from Dryad, Figshare accounted for 59.3% of the results, followed by Dryad (Figure 15). Excluding the replicated Dryad datasets from DataONE and Zenodo becomes evident that they are not the primary choices among researchers for sharing biotic interaction data. Additionally, considering that Figshare is commonly used for publishing supplementary data by various scientific journals, it explains why a significant number of datasets were retrieved from this repository. In contrast, Dryad provided the most reliable results and apparently is the most likely repository for directly accessing biotic interaction data.

Repository	**#Datasets**
DataONE	3,183
Figshare	2,947
Dryad	1,588
Zenodo	1,383
Total	**9,101**

Table 3: Total number of biotic interaction datasets retrieved from each data repository

The term "species interaction" resulted in the highest number of datasets among all the *interaction terms*. It resulted in more datasets than the combined total of datasets returned by all the other *interaction terms* (Figure 16). In contrast, the "biological

[4] ⟨https://search.dataone.org/view/https%3A%2F%2Fdoi.org%2F10.5061%2Fdryad.35rt5%3Fver%3D2018-05-22T10%3A41%3A00.177%2B00%3A00⟩

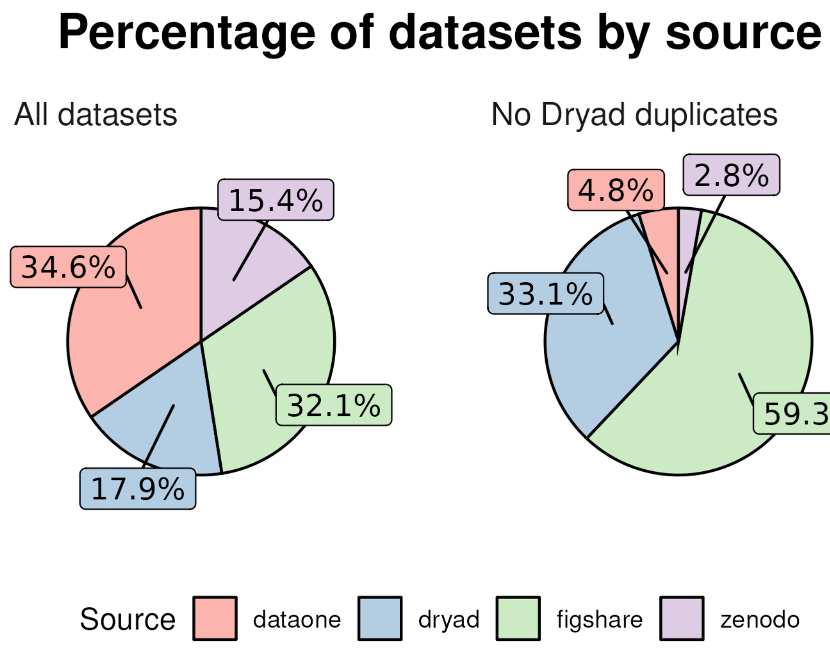


Figure 15: Distribution of biotic interaction datasets across data repositories

The plot illustrates the percentages of datasets retrieved by different repositories: *left* including all datasets; *right* after removing Dryad datasets retrieved from DataONE and Zenodo. Source: Dryad, DataONE, Figshare and Zenodo.

interaction" returned the smallest number of datasets, with a total of 226 datasets. The significant number of datasets retrieved by the "species interaction" term can be partially attributed to its high frequency as a keyword (see Figure 17). The keywords used in datasets extend beyond the *interaction terms*, spanning from more specific terms, such as "plant-plant", "trophic" and "predator-prey", to broader terms like "interaction network" and "species interaction network". It emphasizes the inclusion of specific and broader terms when searching for biotic interactions data in those repositories.

When examining the dataset licenses, the only datasets lacking proper metadata regarding license statements and attributions are exclusively from DataONE. Despite DataONE's adoption of metadata standardization, such as EML, for data exchange between member nodes and the coordinate nodes, the metadata records retrieved from DataONE, excluding the Dryad replicas, did not contain information about license attribution. The field `intelectualRights` in DataONE metadata records were frequently

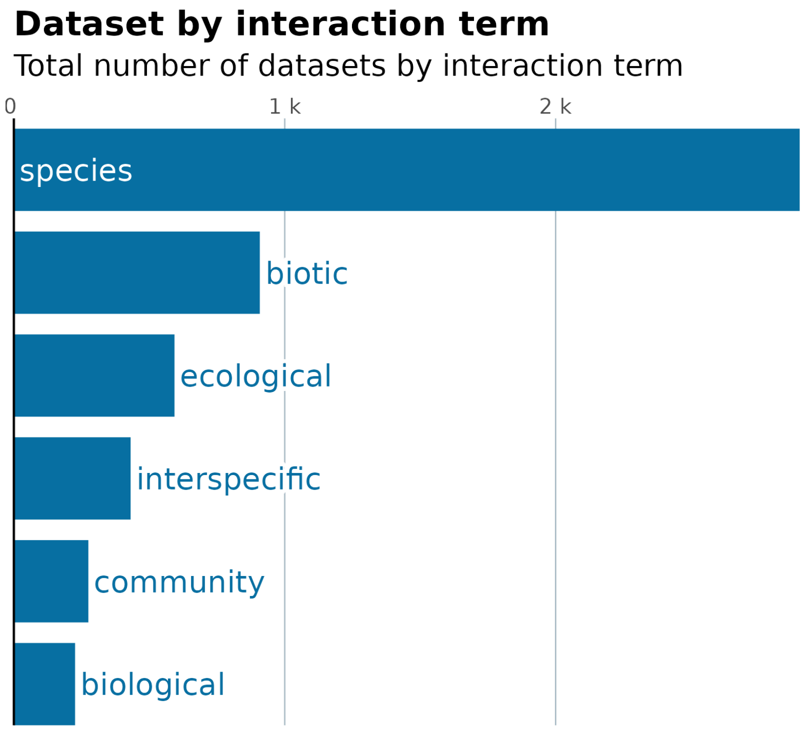


Figure 16: Total number of biotic interaction datasets returned by each “interaction term”

found to be either empty or contained textual information regarding the reuse and distribution of the datasets without specifying a license.

The most common license used is CC-BY-4.0[5], followed by CC0-1.0[6](public domain) (Figure 18). The CC0 is used for all datasets in Dryad as defined in its Publication Policy[7], but it is also common in Figshare and Zenodo. The less restricted publication policy of Figshare and Zenodo allows other open licenses to be applied like CC-BY-NC, MIT, GPL-3.0, CC-BY-SA-4.0 and others (grouped into “Other-Open” category in Figure 18). Only one dataset on restricted copyright was found in Figshare repository.

Regarding the file formats, despite some proprietary formats being used to share data, the most frequent file formats are open formats, except for Microsoft Excel XLS (Figure 19). The top 10 most frequent files corresponds to 99.9% of all file formats in the datasets. Comma Separated Values (CSV), plain text files and Tab-Separated Values (tsv) are among the most frequent formats. These formats have many advantages over other formats, mainly because they do not require dedicate software for reading and processing, CSV and tsv have specific open standards describing them (RFC4180 and IANA MIME type) and they are efficient for data exchange. OpenXML file formats like Microsoft Excel

[5] https://creativecommons.org/licenses/by/4.0/

[6] https://creativecommons.org/publicdomain/zero/1.0/

[7] ⟨https://datadryad.org/stash/terms#publication⟩

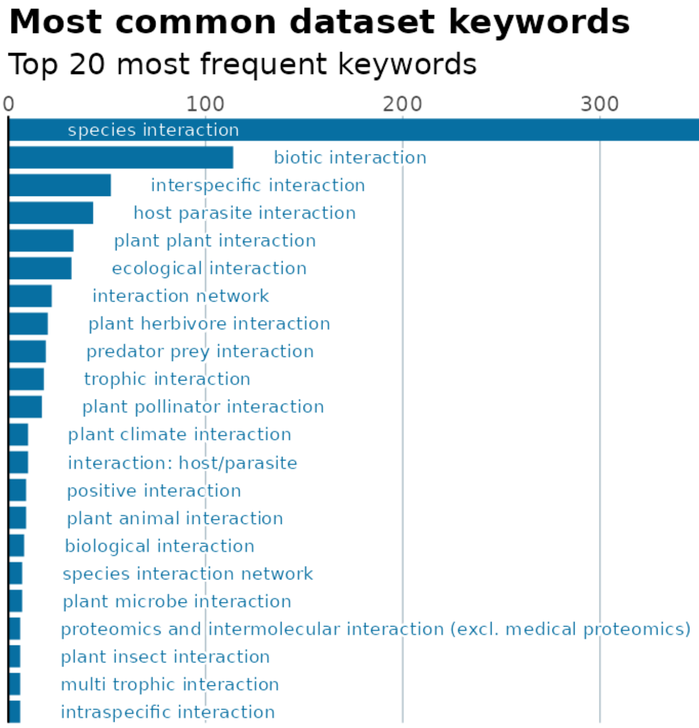


Figure 17: The 20 most frequent keywords in biotic interaction datasets metadata

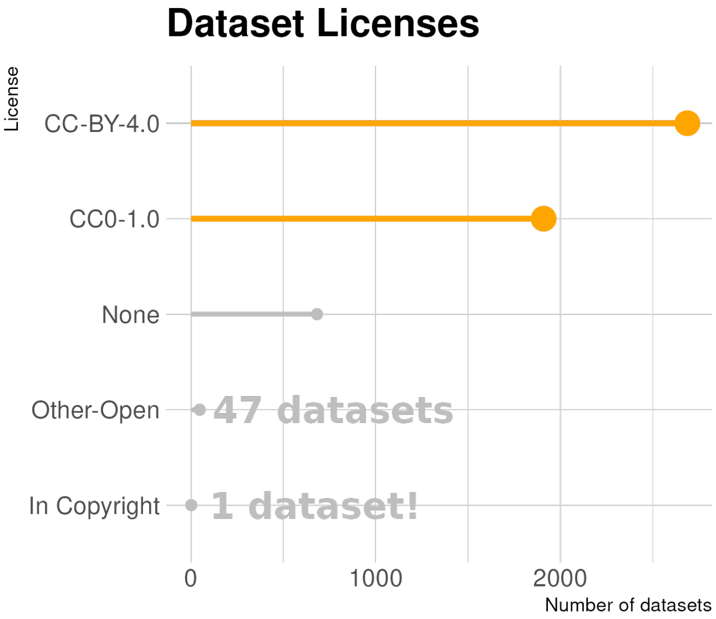


Figure 18: Relationship between dataset license and number of files from all biotic interaction datasets

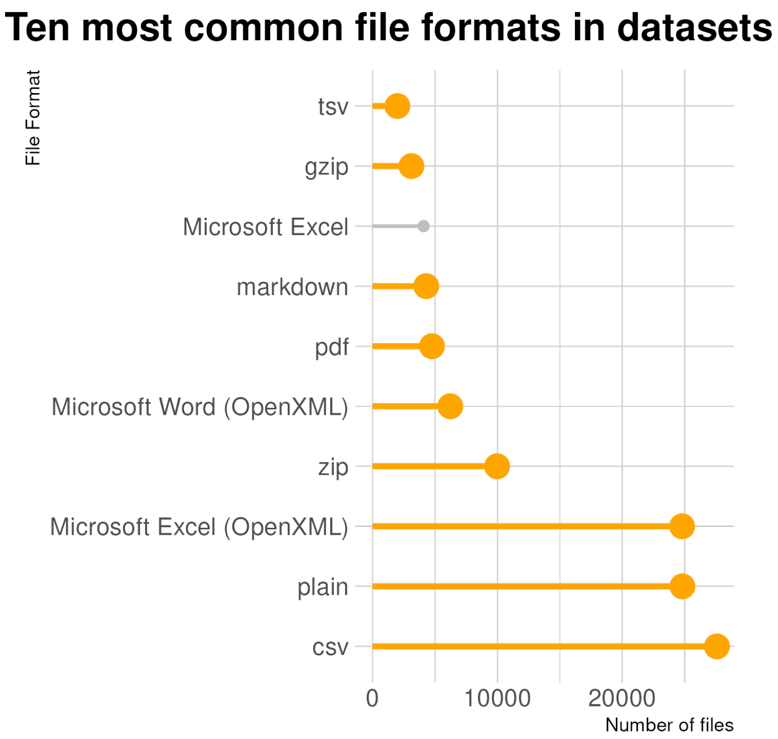


Figure 19: The 10 most common file formats found in the biotic interaction datasets

In gray the proprietary formats and in orange the open-formats

(XLSX) and Microsoft Word (DOCX) are also common used to share data. In the case of DOCX and Portable Document Format (PDF) file formats, it is mostly due to the usage of Figshare as a repository for publishing supplementary data by scientific journals.

Despite the increasing availability of biotic interactions data, this has not contributed to data reuse. The diversity of formats and lack of proper license attributions pose challenges to data indexing and aggregation efforts. This emphasizes the importance of developing data standards and best practices guides specifically for biotic interactions data in order to facilitate data exchange and enable data reuse.

When considering the data available in specialized databases, there are several initiatives that deserve to be mentioned. The Interaction Web DataBase (IWDB) (GUIMARÃES; RAIMUNDO; CAGNOLO, 2012; VAZQUEZ; MELIAN, 2008), created in 2003 by National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California (USA), was the first cooperative effort of scientists to help disseminate data on biotic interactions. The IWDB is a registry of interaction network datasets and does not offer a searching feature. Currently, IWDB is hosted by the University of São Paulo (Brazil) and it offers open-access to 55 datasets distributed in seven categories: anemone-fish, host-parasite, plant-ant, plant-herbivore, plant-pollinator, plant-seed disperser and predator prey. However, it is not receiving new updates since 2020 and it appears to be outdated.

Similar initiatives, like the Web of Life (FORTUNA; ORTEGA; BASCOMPTE, 2014), LifeWebs Project (FAYLE *et al.*, 2016) and GlobalWeb (THOMPSON *et al.*, 2012), provide access to many interaction network datasets, but, similarly to IWDB, they do not adopt any (meta)data standards for sharing data. Without a common data model, the datasets provided by these projects are incompatible with each other. To overcome this limitation and to centralize the access to biotic interactions data, the Global Biotic Interactions (POELEN; SIMONS; MUNGALL, 2014) was created in 2014 and mangal (POISOT *et al.*, 2016) was created in 2015. Mangal is an ecological network database which provides access to standardized data by enforcing the usage of its own template for sharing data through its Application Programming Interface (API) or R client (`rmangal`). The template provided by mangal allows the inclusion of optional data about the environment and species traits, besides the mandatory files describing the networks. GloBI uses a similar approach by defining its own data format and vocabulary of terms, along with other mechanisms of data sharing[8]. Each of these projects provides access to biotic interactions data using its own data formats and schemas, which makes data aggregation across multiple sources a complex task, since data aggregation relies on laborious and repeated transformations of the original datasets into custom, nonstandardized formats, making data integration and discovery a costly and time-consuming process.

Biotic interactions records are also found in GBIF data. The DwC standard defines many terms which can be used to document biotic interactions, however, GBIF is not currently indexing such terms, and therefore it is not possible to directly search for biotic interactions on the GBIF portal or API (but see Section 4.3).

Despite the increasing data availability, data gaps and biases still exist. For instance, there is a larger amount of biotic interactions data from temperate and high-latitude regions compared to the tropics (HORTAL *et al.*, 2015; VIZENTIN-BUGONI *et al.*, 2018), hampering the assessment of global patterns such as latitudinal gradients (ARZABE *et al.*, 2018; SCHEMSKE *et al.*, 2009). In addition, most of the available biotic interaction datasets do not adopt any standard for data or metadata. Moreover, for those that do, the lack of appropriate data standards largely contributes to the dispersion and heterogeneity in the data. As a result, biotic interactions data are often insufficient or biased for many types of analyses, and simply publishing data out of context would fail to produce the

[8] https://www.globalbioticinteractions.org/contribute

correct interpretation of the data nor would it be aligned with researchers needs for sharing and reusing data.

The code used to collect the datasets is available in the GitHub repository[9] and in Salim (2023).

4.2 Indexing and aggregating biotic interactions data - The Global Biotic Interactions (GloBI)

The Global Biotic Interactions (GloBI) is a platform which indexes open access biotic interaction data from many sources. GloBI can effectively process data from different formats, based on provided mappings between fields in the dataset to terms in GloBI's vocabulary [10]. GloBI also indexes biotic interactions data from iNaturalist[11] by utilizing its own mapping[12] between iNaturalist's observation fields[13] and Relation Ontology (MUNGALL *et al.*, 2023). It periodically searches for datasets in GitHub and Zenodo, and individual contributions can also be made by opening an issue on GloBI's GitHub repository, which may required manual intervention to prepare the dataset to be indexed by GloBI.

To facilitate data access and integration with other systems, GloBI provides an API[14]. Additionally, GloBI offers SPARQL and Cypher endpoints for querying the data, as well as an R package called `rglobi` for programmatic access and analysis. In addition, all GloBI's data can be downloaded in different formats, such as, Darwin Core Archive (DwC-A), delimiter-separated values (TSV and CSV), Neo4j dump and sqlite. Internally, the biotic interactions are stored in a graph database (Neo4j) using terms in the GloBI's vocabulary as node properties and Relation Ontology for the interaction types. Currently (May 9, 2023), GloBI is the main resource for sharing and accessing biotic interactions data, providing access to standardized data over 36 million interaction records[15] across 426 datasets.

Despite the success of GloBI in indexing and providing open access to biotic interactions data, like any other indexing system, it is only indexing a subset of data fields

[9] ⟨https://github.com/zedomel/thesis_2023⟩

[10] ⟨https://github.com/globalbioticinteractions/template-dataset#data-format-and-dictionary⟩

[11] ⟨https://www.inaturalist.org⟩

[12] ⟨https://github.com/globalbioticinteractions/inaturalist/blob/main/interaction_types.csv⟩

[13] ⟨https://www.inaturalist.org/observation_fields⟩

[14] ⟨https://github.com/globalbioticinteractions/globalbioticinteractions/wiki/API⟩

[15] `curl https://api.globalbioticinteractions.org/reports/sources | jq '.data[][4]' | awk 'BEGIN sum=0 for (i = 1; i <= NF; i++) sum+=$i END print sum'`

(i.e., the terms in GloBI's vocabulary). Even though the vocabulary can be expanded and new terms added to it, the vocabulary is not a formal vocabulary of terms and it is not maintained by a specific community. GloBI's vocabulary incorporates terms from DwC and ontologies like Uberon (HAENDEL *et al.*, 2009; MUNGALL *et al.*, 2012), Phenotype And Trait Ontology (CONSORTIUM, 2002) and Environment Ontology (BUTTIGIEG *et al.*, 2013), a community interested in creating a common vocabulary can sure benefit from GloBI's vocabulary to create richer representations of biotic interactions data.

Like GBIF, GloBI is capable of aggregating data based on its vocabulary of terms. The difference, however, is that GBIF uses a ratified community data standard (i.e., DwC) while GloBI uses its own data format and vocabulary. Despite providing a great advance in biotic interactions data sharing, reuse and interoperability, the entire community benefits even more from the adoption of common data schemas and vocabularies (SIELEMANN; HAFNER; PUCKER, 2020).

4.3 Biotic interactions data at the Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (GBIF) (GBIF Secretariat, 2022) has indexed beyond 2 billion occurrence records from 85,387 datasets. Many of the these datasets often include "hidden" biotic interaction data. The term "hidden" is derived from the fact that GBIF does not currently index the commonly used DwC terms for documenting biotic interactions. In addition, the different usages of DwC standard to document biotic interactions represents a challenge for data indexing and aggregation. Due to the absence of an GBIF official guide, like for DNA-derived data (ANDERSSON *et al.*, 2021), or community common practices on how to document biotic interactions, these data are completed hidden in GBIF and can not be found through regular search in GBIF's portal or API. Since GBIF is not indexing these terms (e.g., `dwc:associatedTaxa`, `dwc:ResourceRelationship`) it is not possible to know in advance which datasets contain biotic interactions data. To identify such data, the original datasets must be retrieved directly from the data provider and their contents scanned for biotic interactions data. The presence of biotic interactions data in DwC-A indexed by GBIF does not guarantee their availability in the future if GBIF decides to start indexing these terms. To simplify data aggregation, it is highly preferable for the data providers to update their data models according to established guidelines, rather than requiring GBIF to implement alternatives for each unique case. Simplifying data aggregation is a key objective of standardization.

Fortunately, GloBI already has a set of tools based on regular expressions to parse and extract biotic interaction records from DwC-As. A similar workflow employed in GloBI can be used to exploit the indexing of biotic interactions data at GBIF.

4.4 Data representation and formats

Biotic interactions data have been shared in many formats, often using conflicting representations of the interaction concepts. This section presents a non-exhaustive discussion on the utilization of DwC standard for documenting biotic interaction data. In addition, the section explores the implications of interaction networks data representation.

4.4.1 Biotic interactions and the Darwin Core standard

The Darwin Core standard defines many terms which can be used to document biotic interactions, varying by the level of details that can be provided and complexity for adoption.

Particularly, DwC provides a set of terms, known as *association terms*, that can be used to document generic associations between occurrence instances (formally `dwc:Occurrence`) and other resources (e.g., images, references, taxa, occurrences). From that set, two *association terms* are relevant for biotic interactions: the `dwc:associatedTaxa` and the `dwc:associatedOccurrences` terms.

The `dwc:associatedTaxa` term is defined as "*a list (concatenated and separated) of identifiers or names of taxa and the associations of this Occurrence to each of them.*"[16], and the `dwc:associatedOccurrence` is defined as "*a list (concatenated and separated) of identifiers of other Occurrence records and their associations to this Occurrence*"[17]. Thus, both terms, `dwc:associatedTaxa` and `dwc:associatedOccurrences` can be used to document biotic interactions. The term `dwc:associatedTaxa` is used to represent an interaction between an instance of `dwc:Occurrence` class and multiple instances of `dwc:Taxon` class or taxon names. On the other hand, the term `dwc:associatedOccurrences` is used to represent an interaction between an instance of `dwc:Occurrence` class and one or multiple `dwc:Occurrence` instances. However, both terms do not make any assumption about how association types should be documented or if a controlled vocabulary should be used. It creates some problems for machines, and even humans, since there is no sepa-

[16] ⟨http://rs.tdwg.org/dwc/terms/associatedTaxa⟩

[17] ⟨http://rs.tdwg.org/dwc/terms/associatedOccurrences⟩

ration between the association type and the associated taxon/occurrence, culminating in a myriad of different forms to fill data into these fields (MUKHIN; VLADYKINA, 2020; ROBERTS, 2023).

Another form to document biotic interactions using the DwC standard is through the `dwc:ResourceRelationship` class. The "*ResourceRelationship class is an alternative means of representing associations, and with more detail*"[18], and it is defined as "*a relationship of one rdfs:Resource to another*"[19], where a *rdfs:Resource* can be any "*identifiable records or instances of classes and may include, but need not be limited to dwc:Occurrence, dwc:Organism, dwc:MaterialSample, dwc:Event, dwc:Location, dwc:GeologicalContext, dwc:Identification, or dwc:Taxon*". Thus, the `dwc:ResourceRelationship` class can be used to document relationships (e.g., interactions) between two or more `dwc:Occurrence`'s or `dwc:Taxon`'s and create an interaction network representation using DwC. Currently, the `dwc:ResourceRelationship` class defines eight terms:

a) `dwc:resourceRelationshipID`: an identifier for the relationship, e.g., a global unique identifier for an interaction;

b) `dwc:resourceID`: an identifier for the subject of the relationship, e.g., the identifier of a `dwc:Occurrence` or `dwc:Taxon` participating in the interaction;

c) `dwc:relatedResourceID`: the identifier for the object of the relationship, e.g., the identifier of the other `dwc:Occurrence` or `dwc:Taxon` participating in the interaction;

d) `dwc:relationshipOfResource`: the literal form for the relationship type, e.g., *pollinates*;

e) `dwc:relationshipOfResourceID`: the identifier (non-literal) for the relationship type, e.g., ⟨*http://purl.obolibrary.org/obo/RO_0002455*⟩

f) `dwc:relationshipAccordingTo`: the source (e.g., person, organization, publication) which established the relationship;

g) `dwc:relationshipEstablishedDate`: the date when the relationship was established;

h) `dwc:relationshipRemarks`: additional remarks about the relationship.

[18] comments on ⟨http://rs.tdwg.org/dwc/terms/associatedTaxa⟩

[19] ⟨http://rs.tdwg.org/dwc/terms/ResourceRelationship⟩

Despite its complexity, the `dwc:ResourceRelationship` class has some advantages compared to the *association terms*: i) it provides distinct terms for the subject, the object and the relationship type, ii) multiple relationships can be documented without relying on recommendations which usually are not adopted ("recommended best practice is to separate the values in a list with space vertical bar space (|)[20]"), iii) the relationship has a (globally) unique identifier making possible to reference a specific relationship, and iv) additional details can be provided for each relationship using the terms `dwc:relationshipAccordingTo`, `dwc:relationshipEstablishedDate`, and `dwc:relationshipRemarks` (e.g. HERNÁNDEZ *et al.*, 2021).

Although these terms represent the "formal" representation of associations in DwC, biotic interactions have been documented using other terms, such as `dwc:dynamicProperties`, `dwc:habitat` and `dwc:occurrenceRemarks`. The `dwc:dynamicProperties` is a record-level term which can be used to provide "a list of additional measurements, facts, characteristics, or assertions about the record"[21] using a structured content. The recommended format is JavaScript Object Notation (JSON), but the term has been used with different formats (e.g. DEMBOSKI, 2023). Biotic interactions are included in the `dwc:dynamicProperties` as *key:value* pairs, where usually, the *key* is the interaction type (e.g., predates, pollinates) and the *value* is a list of taxon names. The terms `dwc:habitat` and `dwc:occurrenceRemarks` are descriptive terms, meaning that they accept text in natural language, what makes the interpretation of their values challenge for machines (e.g Queensland Department of Environment and Science, 2023).

However, simply adopting the official terms for documenting associations and relationships does not guarantee that data will be correctly interpreted afterward. Thus, for data to be reusable, interoperable and allow access to the context-dependent characteristics of the interactions, it is desirable for datasets to adhere to the following guidelines:

a) interactions should be individually identified and referenced;

b) datasets should use a common controlled vocabulary for the interaction types (e.g., OBO Relation Ontology (MUNGALL *et al.*, 2023));

c) datasets should use either the *association terms* or the `dwc:ResourceRelationship` to document biotic interactions, with preference to latter method;

d) context-dependent data can be linked to the interactions and to the organisms

[20] ⟨http://rs.tdwg.org/dwc/terms/associatedTaxa⟩

[21] ⟨http://rs.tdwg.org/dwc/terms/dynamicProperties⟩

participating on the interactions (e.g., interaction outcomes and effects, organisms traits);

e) bibliographic references should be used as evidence for the interactions.

These five items above ensure that biotic interactions data follow the recommendations for data standardization and provide minimal information regarding the biotic interactions.

4.4.2 Interaction networks

Interactions networks represent a distinct form of recording biotic interactions. Usually, interaction networks are a *species level* representation of the interactions, with non-binary networks encoding some metric to quantify/qualify the effects or strength of pairwise interactions. In interaction networks, nodes represent species or higher taxonomic groups and edges represent the interactions between a pair of nodes (DORMANN; FRÜND; SCHAEFER, 2017). Historically, food webs have been the most commonly studied type of interaction network, but over the last two decades, bipartite (two-group) networks have dominated the literature (DORMANN; FRÜND; SCHAEFER, 2017). Sampling interactions are time and cost consuming and require much effort to build up a complete network and the collected data tends to be biased and noisy (AGUIAR *et al.*, 2019). Thus, enabling data reuse and aggregation of interaction networks is essential to overcome biases and to compose large datasets which can help the understanding of how interactions drive ecological and evolutionary dynamics and the maintenance of ecosystems (ALBRECHT *et al.*, 2018; DELMAS *et al.*, 2019; LANDI *et al.*, 2018). Importantly, reusing and aggregating data from different interaction networks is high impacted by topological heterogeneity. Biological and environmental drivers, sampling strategies and network construction methods have great influence on the topology of interaction networks (BRIMACOMBE *et al.*, 2023) and therefore, without adequate metadata to describe these classes of heterogeneity, data reuse and aggregation is often limited.

In fact, to build an interaction network, pairwise interactions are sampled in the field, often including more details than just the “tetranomial”, and only after, the network is built and some metrics are calculated (e.g., frequencies, strength, visitations). An interaction network is a summary of the individual pairwise interactions sampled in the field, represented as an adjacency matrix or list of nodes and edges. While *organism level* interactions are passive to create an interaction network (*species level*), with some information loss, the opposite is not possible. There is much less information in an interaction network than in a dataset of individual interactions. Despite efforts for the standardization of interaction networks (POISOT *et al.*, 2016), it is generally preferable to share primary and standardized data on organism-level interactions. Biotic interactions can be individually documented using DwC, individual records can, therefore, be used to build interaction networks considering different aspects of these records (e.g., temporal, spatial, abiotic and biotic factors) and used to create other representations besides networks. It is important to note that while interaction networks are valuable, the way data are traditionally shared (e.g., adjacency matrix) may not be the most appropriate. This is because the construction of the network relies on sampled individual interactions. As discussed in Section 2.1, *species-level* interactions, in addition, assume species homogeneity and the biological unit of interest is mixed with the observational unit (LAZIC; CLARKE-WILLIAMS; MUNAFÒ, 2018).

Interaction networks will be continually created and shared, but if, for any reason, it is not possible to share the primary organism-level interactions data, at least, some standardization should be adopted by the community (POISOT *et al.*, 2016).

5 MINING BIOTIC INTERACTIONS IN GBIF DATA

In section 4.3 is described the existence of, here denominated, “hidden” biotic interactions data in the GBIF data. This section presents a methodology supported by the GloBI indexing system to extract biotic interactions data from Darwin Core Archive (DwC-A) retrieved from GBIF.

5.1 Methods

As described in Section 4.4.1, the diversity of practices adopted by the biodiversity community to document biotic interactions data impose challenges for data indexing and aggregation. Since GBIF primarily currently focuses on occurrence data, it does not index biotic interactions data. For that reason, it is not possible to search GBIF for interaction data within occurrence records, either through the web portal or using the API. However, it is possibly to use the GBIF index to retrieve all datasets directly from the data provider, and later perform a local indexing of the biotic interactions. The GBIF API offers an endpoint to access datasets metadata, which can be used to retrieve datasets from the original sources.

This section describes an alternative method to retrieve versioned copies of GBIF datasets using *preston* (ELLIOTT; POELEN; FORTES, 2020). The methods for processing and aggregating were also described in detail and summarized in Figure 20.

Preston is an open-source software that versions biodiversity datasets, enabling reproducible research by uniquely identifying a versioned copy of all or parts of GBIF-indexed datasets. It uses the PROV (LEBO *et al.*, 2013) and PAV (CICCARESE *et al.*, 2013) ontologies to provide provenance information generated in different systems and under different contexts for digital biodiversity datasets. A web-accessible copy of the biodiversity data graph generated with *preston* for keeping track of GBIF, Integrated Digitized Biocollections (iDigBio) and Biological Collection Access Service (BioCASe) registries is

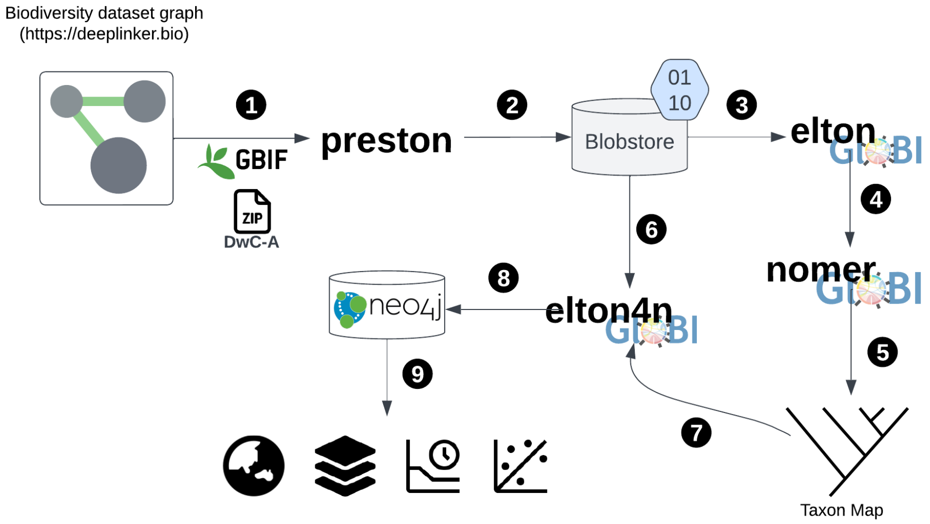


Figure 20: Diagram depicting the methodology for extracting biotic interaction data from GBIF

The diagram shows the methodology used to find and extract biotic interaction data from Darwin Core Archive (DwC-A) retrieved from Global Biodiversity Information Facility (GBIF) registry. 1) Download DwC-A from GBIF using *preston*; 2) *preston* creates local copies of the datasets in a blobstore; 3)a apply *elton* over the DwC-A to find and extract biotic interactions data; 4) use *nomer* to match names found in interaction records to names in provided taxonomic catalogues; 5) create a "Taxon Map" with all valid taxon names; 6) use *elton4n* to import interaction records to the Neo4j database; 7) use the "Taxon Map" and *elton4n* to create `Taxon` nodes in the Neo4j database; 8) use *elton4n* to import biotic interaction records with valid taxon names; 9) use Neo4j database in subsequent analysis. Source: created by the author.

available at ⟨http:s//linker.bio⟩. *Preston* works by crawling and downloading copies of biodiversity registries and their datasets, while it creates a new version of the biodiversity dataset graph. The complete record of the crawling activity is stored into a blob store and the *hexastore* (WEISS; KARRAS; BERNSTEIN, 2008). A simplified indexing scheme for RDF is used to describe the record as a version of the biodiversity dataset graph. The provenance of the biodiversity dataset graph is retained by the relations between each version of the biodiversity dataset graph and the previous crawling activity.

Thus, the first step in accessing the "hidden data" in GBIF was to create versioned copies of all DwC-A retrieved from the latest version of the web-accessible biodiversity dataset graph.

Subsequently, each DwC-A was locally processed in order to extract biotic interaction records using *elton*[1]. *Elton* is an open-source command-line tool specially designed to parse and organize biotic interactions data from various dataset formats. It implements multiple data parsers that facilitate the identification and extraction of biotic interaction

[1] github.com/globalbioticinteractions/elton

records from datasets, including a parser for DwC-A. The DwC-A parser employs a set of regular expressions to systematically search for and extract biotic interactions data from various DwC terms. The extracted data are mapped to terms in the GloBI's vocabulary, allowing aggregation of biotic interactions data from diverse sources. By default, *elton* aligns the extracted interaction types with the terms in the OBO Relations Ontology (POISOT *et al.*, 2016) (e.g., `visitsFlowersOf`, `eats`, `interactsWith`). However, it is possible to extend the interaction types alignment using custom mappings between interaction types identified in the data and the corresponding terms in the OBO Relations Ontology (RO). *Elton* is the software used by GloBI to extract and standardize biotic interactions data from datasets currently in its registry.

Since interaction types documented in the DwC-A retrieved from GBIF do not typically align with the terms defined in RO, determining the specific values used to document interaction types requires preliminary investigation. Throughout its execution, *elton* generates detailed information regarding each processed record, including warnings about unsupported interaction types found in data (i.e., interaction types not found in RO). In order to compile a list of unsupported interaction types and create a customized mapping for these interaction types to terms in the RO, *elton* was executed twice. The first execution was intended to obtain a list of unsupported interactions types (e.g., “accociated taxa”, “hospedador de”, “infectedBy”) and subsequently, manually create a customized mapping for the values found to terms in the RO. The second execution was performed using the customized interaction types mapping obtained from the previous execution. This mapping enabled *elton* to correctly recognize the interaction types present in data and aligned them with the terms in the RO. The custom interaction types mapping used here are available in Appendix A and also at ⟨unsupportedinteractiontypes⟩.

Elton does not validate taxonomic names in the biotic interaction records extracted from datasets. Thus, the validation and normalization of taxonomic names was performed separately using *nomer*[2]. *Nomer* is an open-source command-line tool which maps identifiers and names in data to identifiers and names in the provided taxonomic catalogs. It stores and indexes local versioned copies of different taxonomic catalogs, which allows *nomer* to efficiently find matches between input names and valid taxon names. *Nomer* was used to match the names found in the biotic interaction records against multiple taxonomic catalogs, such as GBIF Backbone Taxonomy (GBIF Secretariat, 2022), Integrated Taxonomic Information System (ITIS) (National Museum of Natural History, Smithsonian Institution, 2023), Catalogue of Life (CoL) (BÁNKI *et al.*, 2023), Index Fun-

[2] github.com/globalbioticinteractions/nomer

gorum (KIRK, 2000), National Center of Biotechnology Information (NCBI) Taxonomy (FEDERHEN, 2012), World Flora Online (WFO) (WFO, 2023) and Open Tree of Life (OTT) (REES; CRANSTON, 2017). The result was a *Taxon Map* mapping names found in data to their respective names in the catalogs. Names that did not yield any matches within the taxonomic catalogs were removed along with their corresponding interaction records.

In order to facilitate posterior analysis, the biotic interaction records were stored in the Neo4j[3] database. The process of importing the interaction records into the database, after taxonomic validation, was performed using *elton4n*. *Elton4n* is a tool used in GloBI to create a graph representation of the biotic interaction datasets in its registry. It uses the same logic employed by *elton* to find and extract biotic interactions from DwC-As, and using the *Taxon Map* it creates the mappings between original names found in data to taxon names in the catalogs. During importing, five types of nodes are created:

a) *Specimen*: contains original data of a record. It represents a physical specimen or an observation;

b) *Taxon*: represents a taxon name from a taxonomic catalog, and it is linked to the *Specimen* nodes through a `CLASSIFIED_AS` relationship type;

c) *Location*: stores the geographic information linked to a *Specimen* node through a `COLLECTED_AT` relationship type;

d) *Reference*: stores the bibliographic references for *Specimen* nodes;

e) *Dataset*: stores dataset metadata.

Figure 21 presents an example of nodes and relationships in the Neo4j database after importing the interaction records.

An advantage of this compact data model is its ability to represent not only interactions at the *organism* level but also at the *species* level. Species-level interactions are represented in the graph by edges between *Taxon* nodes. In contrast, individual-level interactions are represented by edges between *Specimen* nodes. In both cases, edges may have properties and additional data (e.g. number of visits, interaction strength, biomass consumption), stored or calculated, can be provided as a result of querying the database.

[3] ⟨https://neo4j.com/product/neo4j-graph-database/⟩

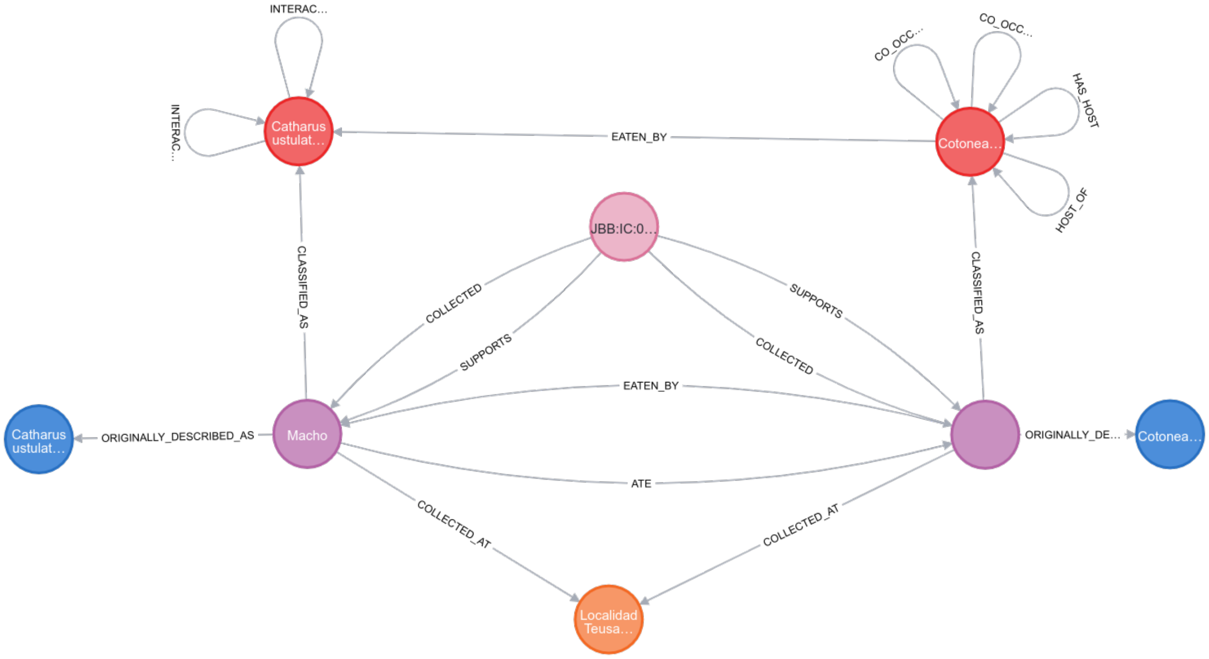


Figure 21: Example of a Neo4j graph model for a biotic interaction

In red are the two taxon nodes with taxonomic information about the two specimens (in purple). The specimen nodes are linked to the location node (in orange), the reference of the dataset (pink) and the original names provided in the record found by *elton* (blue).

After the successful import of the data, an exploratory analysis was conducted to understand various aspects of data coverage and potential bias within datasets. The objective of this analysis was not to generate novel biological knowledge, but rather to offer an overview of the potential of the unexplored "hidden" data in supporting studies in biology.

The exploratory analysis focused on investigating the taxonomic composition and coverage of the biotic interactions records, as well as the geographic and temporal distribution of the data. In the taxonomic coverage analysis, the number of species in the data was compared to the species count in the GBIF Backbone Taxonomy, National Center of Biotechnology Information (NCBI) Taxonomy and GloBI. It aimed to identify biases and gaps in the taxonomic information in the "hidden" data.

For the geographic coverage analysis, records lacking geographic coordinates were initially removed from the analysis. Subsequently, the geographic coordinates for these records were obtained by processing and geocoding their textual locations descriptions. Despite geocoding may introduce some errors and potentially reduce the data quality, it is important to note that the objective of the analysis was not to conduct an investigation of global patterns of biotic interactions. Instead, the aim was to gain insights into the general geographic distribution of the records. The temporal analysis considered only records with valid date/time values for `eventDate` property.

5.2 Results and Discussion

A total of 65,565 DwC-As, containing 691,181,259 records, were retrieved from the biodiversity dataset graph (BIOGOUDA, 2022) using *preston*. Subsequently, *elton* identified 10,699,053 biotic interaction records (1.5% of scanned records) in 1,079 datasets (1.64% of the datasets). Even with a custom mapping of interaction types, unsupported interaction types were reported for 4,271,621 records. Further investigation revealed that the unsupported interaction types did not express any form of association between taxa in the records. The regular expressions used by *elton* for extracting interaction data from DwC terms are generic and may produce matches even when there is no interaction being documented. This is especially true for terms such as `dwc:habitat` and `dwc:occurrenceRemarks`, which contain free text data, where the regular expressions used by *elton* found matches even when no interaction was documented. For example, the regular expression which implements the DwC recommendations for populating the `dwc:associatedTaxa` term (in the form "`interaction type: taxon name | another interaction type: another taxon name | ...`)", when applied to the `dwc:habit` term with a value of "`shrubScrub; slope aspect: 71.9; slope gradient: 14.35; soil type order: Histosols`" found matches for `slope aspect: 71.9`, `slope gradient: 14.35` and `soil type order: Histosols` which definitively are not biotic interactions.

It is important to note that although the fraction of datasets containing biotic interactions data represents only 1.64% of all datasets in GBIF, it accounts for almost one-third of the total number of records in GloBI. This substantial collection is one of the largest resources of biotic interactions data, but currently inaccessible due to limitations in GBIF indexing system.

In addition, the usage of the DwC terms and schemas to document biotic interactions was not uniform across the datasets analyzed. From all the terms processed by *elton*, the `dwc:associatedTaxa` term returned the highest number of records (90.53% of the total dataset, see Figure 22). Despite its limitations, the `dwc:associatedTaxa` term is the simplest form to document biotic interactions using DwC. However, the process of extracting biotic interactions data from this term is not always trivial due to the lack of standardization in the documented values. This becomes evident when considering the number of unsupported interaction types found by *elton* and the wide range of formatted

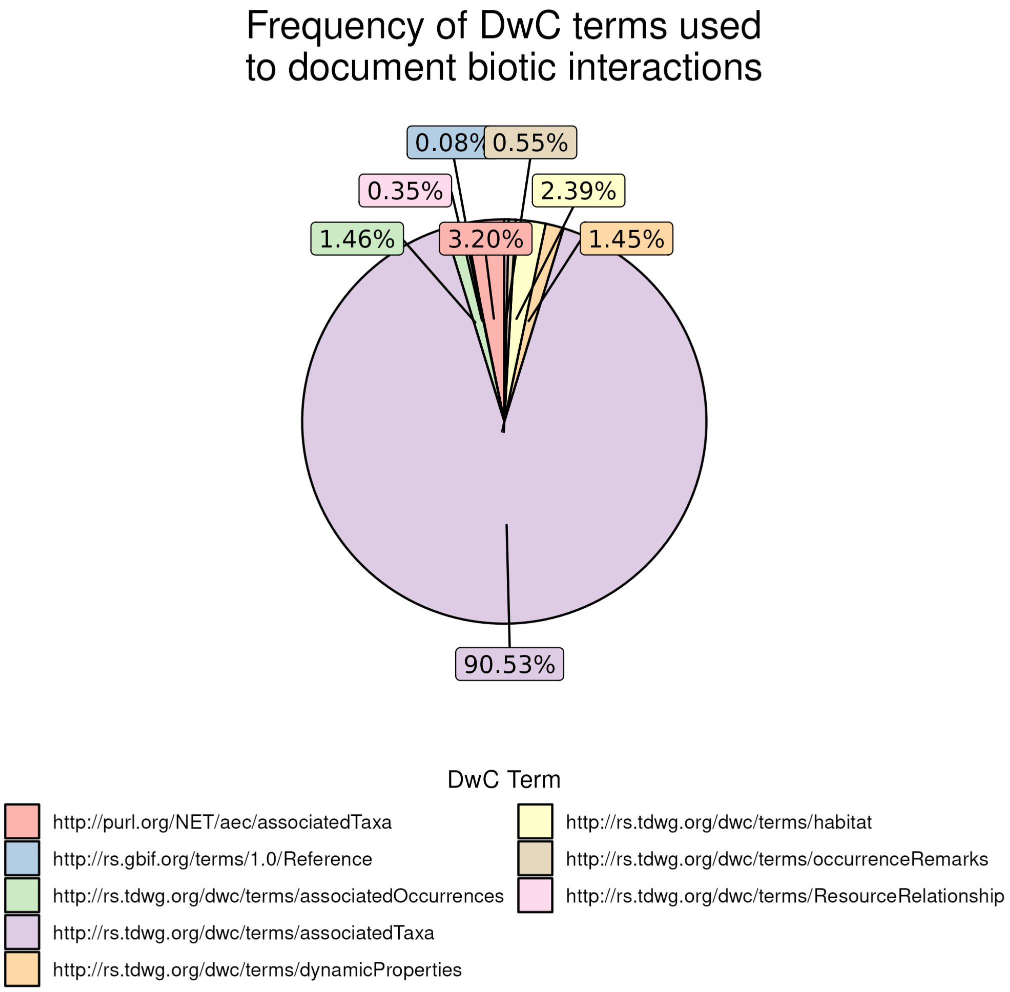


Figure 22: Frequency of Darwin Core terms used to document biotic interactions as percentage of the total number of records

and non-formatted values provided within the term.

The taxonomic catalogs used with *nomer* provided names for 8,215,365 interaction records (76.8% of total records found by *elton*). In the records with valid taxon names the generic interaction types, such as, *interacts with* and *co-occurs with* were the most frequently found, corresponding to 84.6% of all interaction records (Table 4). It is, partially, due to the mappings provided in the customized interaction types. In addition, the *interacts with* type was used as a fallback when the interaction type was missing from a term's value. It may produce "false" interaction records if the term which contains the value is generic or not appropriated for documenting associations in DwC (e.g `dwc:habitat`, `dwc:occurrenceRemarks`, `dwc:dynamicProperties`). However, when applied to the *association terms* (e.g. `dwc:associatedTaxa`) it is likely to produce a valid biotic interaction record, since associations can be interpreted, at least, as *co-occurrences*, but this may not be always the case. For the purposes of this analysis, these generic interaction records are interesting since they provide an estimation of the potential of using GBIF data for studying biotic interactions. Although, for biological studies involving these data, automated or manual curation of the data is essential. Since the objective here is not to provide insights about the biological implications of the global patterns found in the data, all interaction records found by *elton* were used in the subsequent analysis.

Interaction Type	**# Records**	**%**
interactsWith	7,751,658	72.45
coOccursWith	1,302,033	12.17
hasHost	854,006	7.98
epiphyteOf	601,609	5.62
parasiteOf	59,902	0.56
kills	52,220	0.49
ectoparasiteOf	36,706	0.34
visitsFlowersOf	17,995	0.17
symbiontOf	16,008	0.15
createsHabitatFor	3,072	0.03
mutualistOf	2,146	0.02
preysOn	879	0.01
eats	445	<0.01
visits	254	<0.01
parasitoidOf	31	<0.01
pathogenOf	30	<0.01
providesNutrientsFor	30	<0.01
hyperparasiteOf	16	<0.01
pollinates	9	<0.01

Table 4: Total number and proportion of records for each interaction type

Interestingly, the most common names not found in any taxonomic catalog were “vehicle” (2.0% of the records without valid taxon names), “løvtræ” (deciduous tree in Danish, 1.4%) and “unidentified” (1.1%). The “vehicle” was originated from “associations” of type “killed by vehicle“. It shows the importance of having specific forms to document biotic interactions separately from general associations. Otherwise, there is no simply way to differentiate valid names, but not found in a taxonomic catalog, from invalid names (“vehicle”) and misspelled names (e.g., “Trichillia martiana”, the correct is “Trichilia martiana”).

5.2.1 Taxonomic coverage and bias

The biotic interaction records found in GBIF data are mostly for plants (70.8% of the taxon names are classified in the Plantae, Viridiplantae, Chloroplastida and Archaeplastida kingdoms), followed by animals (17.3% in Animalia, Metazoa and Protozoa kingdoms), fungi (6.6% in Fungi and Chromista kingdoms) and viruses (3.9% in Viruses kingdom). Bacteria and Archaea correspond to less than 2% of the records, and only 0.2% of records have their taxonomic classification unknown or undefined (*incertae sedis*) (Table 5).

Kingdom	**# Taxon names**	**%**
Plant/Viridiplantae/Chloroplastida/Archaeplastida	11,629,115	70.8
Animalia/Metazoa/Protozoa	2,846,794	17.3
Fungi/Chromista	1,085,827	6.6
Viruses	634,230	3.9
Bacteria	202,339	1.2
incertae sedis	30572	0.2
Archaea	1,709	<0.1
Protista	144	<0.1

Table 5: Total number of taxon names and proportions found in the interaction records for each kingdom

The most common interactions are plant-plant (56.9% of interaction records), plant-animal (14.8%) and plant-fungi (11.2%) interactions, but a substantial number of records was found for animal-virus (6.7%) and animal-animal interactions (5.4%). The other interactions account for less than 6% of the records. A complete list of interactions summarized by kingdom is in Appendix B.

Within plants, the most abundant families were Asteraceae (10.0%), Poaceae (6.7%) and Fabaceae (4.8%), but the remaining of the records involving plant families like Fagaceae (3.3%), Rosaceae (2.4%), Pinaceae (2.4%), Cyperaceae (2.1%), Cactaceae (1.7%) and others (see (SALIM, 2023) for a complete list of plant families). The most common interaction types for plants were *interacts with* and *co-occurs with* (95.1% of interaction records involving plants). The high number of generic interaction types is due to the equivalent number of original interaction types documenting associations involving plants (e.g., “associated with”, “in association with”, “found near of”) mapped to *interact with* or *co-occurs with* terms of RO. In addition to the generic interaction types, the *has host* interaction type corresponds to 4.4% of the total interactions involving plants, followed by *visits flowers of* (0.3%). Other interaction types account for less than 0.2% of the records (a complete list of interaction types for plants is in Appendix C).

For animals, the most abundant orders are Hemiptera (10.8%), Hymenoptera (8.3%), Artiodactyla (4.2%), Passeriformes (2.5%) and Anseriformes (2.1%). Similarly to the plants, the most common interaction types are *interacts with* (67.4%) and *co-occurs with* (17.1%), followed by *has host* (12.2%), *ectoparasite of* (1.5%) and *visits flowers of* (0.7%). The remain interaction types together account for 1.1% of the total interactions involving animals (a complete list of interaction types for animals is presented in Appendix D).

The taxonomic coverage of the analyzed data was performed against the NCBI Taxonomy and GBIF Backbone Taxonomy, the taxonomies which returned the highest number of matches using *nomer* (Table 6). Only records identified at species rank were considered to calculate the taxonomic coverage of biotic interactions records over the two taxonomies, and the results are summarized by kingdom.

Taxonomy	#Taxon names
NCBI	385,932
GBIF	345,873
OTT	194,462
CoL	183,819
ITIS	118,019
WFO	67,028

Table 6: Total number of matches (taxon names found) by *nomer* for each taxonomic catalogue

Compared to GBIF Backbone Taxonomy, the NCBI Taxonomy returned almost twice as many matches, despite NCBI having a smaller taxonomy in relation to the number of species names. While the taxon names found in data correspond to 11% of the names in the NCBI Taxonomy, for GBIF Taxonomy it was under 6%. The comparison by kingdom revealed that GBIF taxonomy did not return matches for viruses, what is explained by the nomenclature used to provide names for viruses (Figures 23 and 24). Viruses taxonomic names often include specific strains or variants obtained and cultured from a particular source (e.g., *Macrobrachium nudivirus CN-SL2011*). Although the GBIF Taxonomy is updated regularly, it does not provide lower classification for viruses (below family). Unlike GBIF Taxonomy, the NCBI Taxonomy uses the International Committee on Taxonomy of Viruses (ICTV) for viruses' names and currently (2022 Release) it includes 11,273 viruses species names.

In conjunction with taxonomic information, geographic and temporal factors are key information in ecological studies. The completeness of a biotic interaction record can

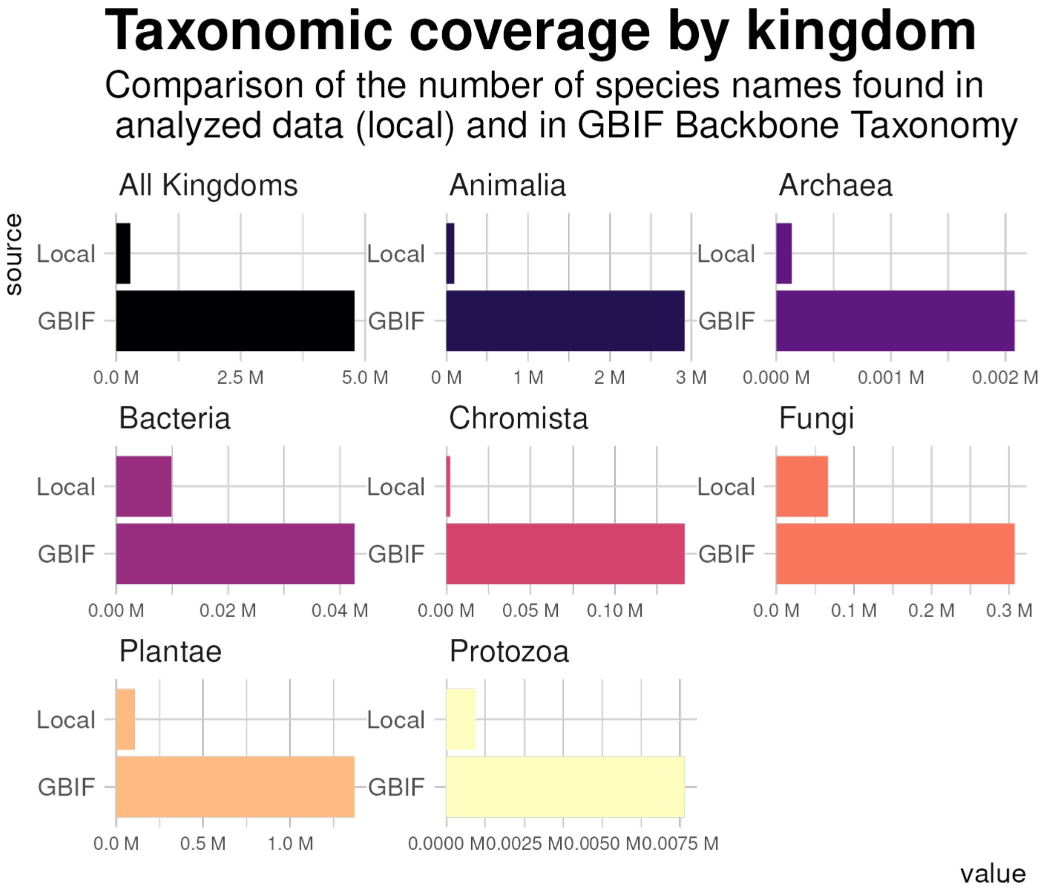


Figure 23: Overall and by kingdom taxonomic coverage in the GBIF Taxonomic Backbone

The comparison considered only the records with full taxonomic identification at the species rank. Source: created by the author.

be defined in terms of these three components. Thus, records with data about these three components are beyond the "tetranomials" view of biotic interactions. By including geographic and temporal data, they contribute to a more realistic context-dependent view of biotic interactions. In the studied data, the geographic and temporal distribution were unequal among kingdoms (Figure 25). Geographic and temporal data are relatively common in interaction records involving plants, animals and even for the polyphyletic Protozoa kingdom, but less common for other kingdoms, especially Bacteria and Archeae. Viruses were excluded for completeness calculations, since GBIF did not return matches bellow family.

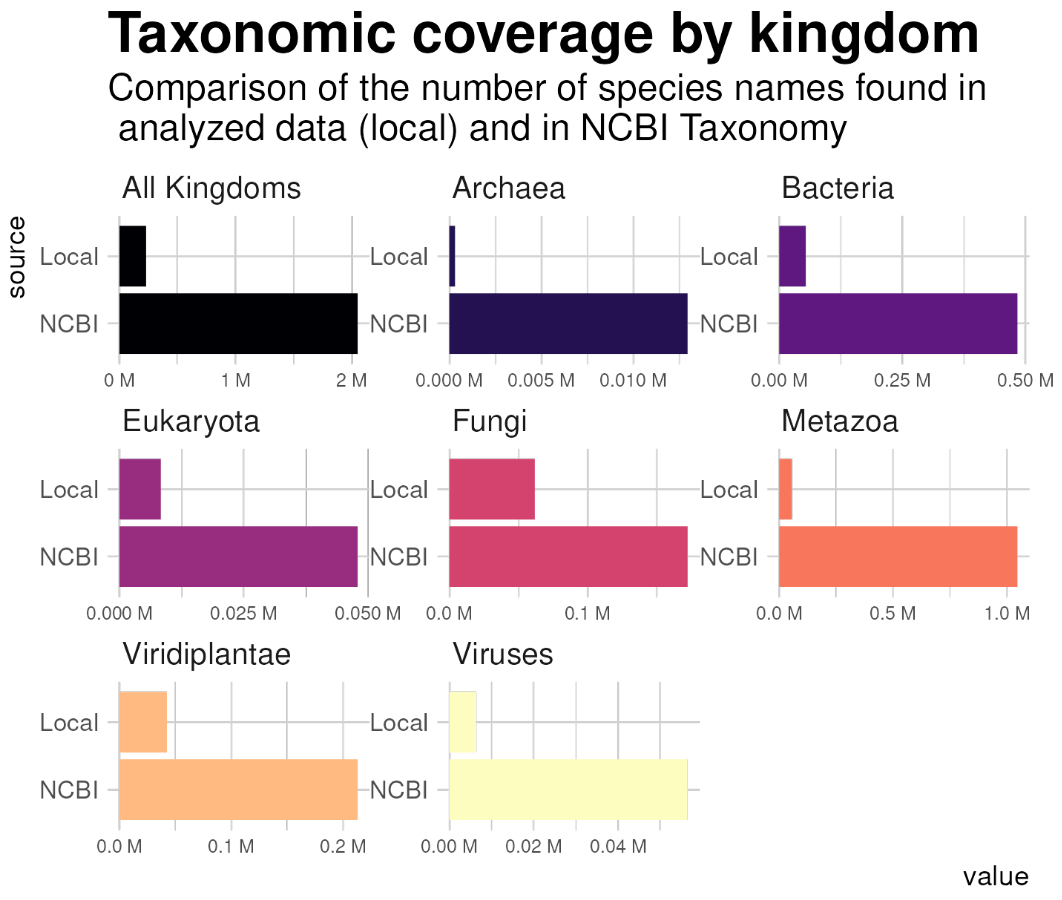


Figure 24: Overall and by kingdom taxonomic coverage in the NCBI Taxonomic

The comparison considered only the records with full taxonomic identification at the species rank.

5.2.2 Geographic and temporal distribution of the interactions

Access to taxonomic data is essential in biological research, and numerous studies can greatly benefit from systematic access to data. Furthermore, geographic and temporal data are crucial for understanding how life has been organized and evolved on Earth. The biotic interactions hidden in GBIF data can, potentially, provide complementary and supplementary information to many studies in Biology. Thus, the study of the various aspects of the data currently available is crucial for deriving their meaning and context.

The geographic coverage of the biotic interactions data was studied considering only the records with complete taxonomic information at the species level (i.e., valid taxon names for both of the interacting specimens). Furthermore, only the records with valid

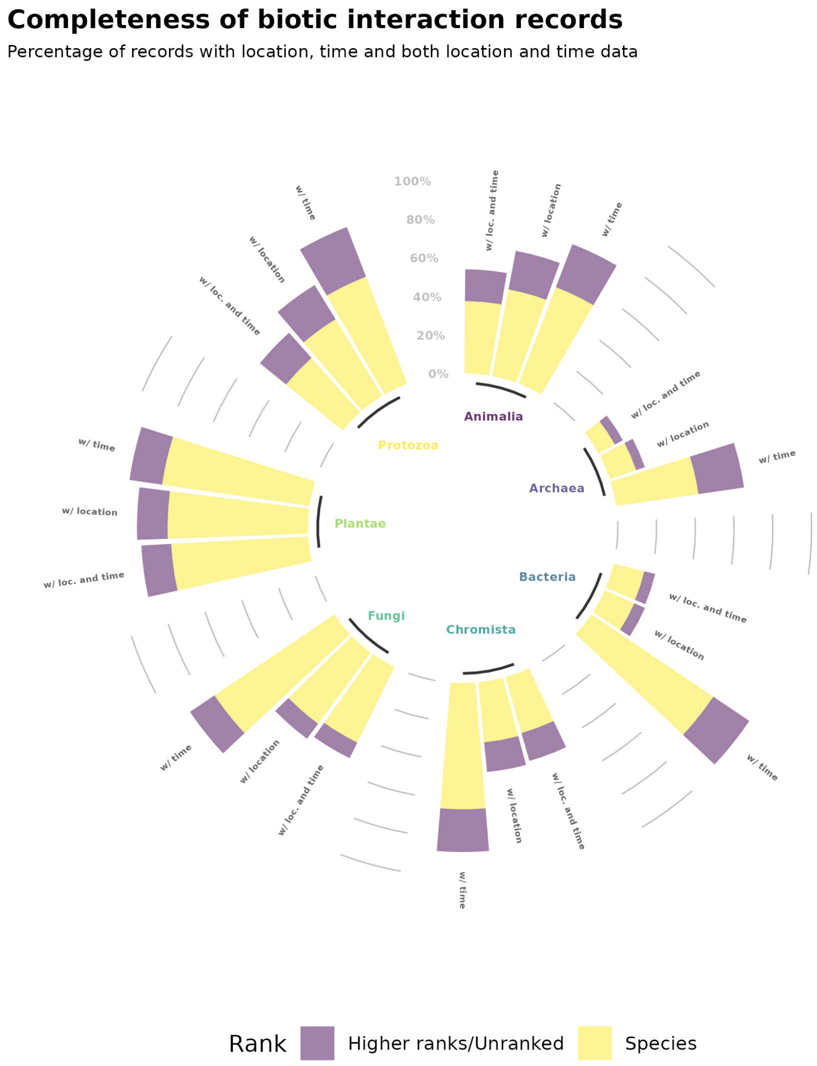


Figure 25: Completeness of biotic interaction records regarding the number of non-null values of geographic coordinates and temporal data (`eventDate`) compared to the total number of records of each kingdom. Source: created by the author.

geographic coordinates (i.e., decimal latitude and longitude) are considered. Since few of the records provide information about the coordinate system and horizontal datum used, the World Geodetic System 1984 (WGS84) was assumed to represent the locations of all biotic interactions records in data. After removing records with coordinates at the "Null Island" (KURGAN, 2013), i.e., at zero degrees latitude and zero degrees longitude (0°N 0°E), the analysis was carried out on 4,734,169 records (79.6% of total records).

The distribution of the biotic interactions records on the Earth's surface is not different from observed for the occurrence data which they were derived from (BECK *et al.*, 2014; ROCHA-ORTEGA; RODRIGUEZ; CÓRDOBA-AGUILAR, 2021; QIAN; ZHANG; JIANG, 2022). The interaction records are mostly concentrated in the north hemisphere (96.4%), especially in North America and Europe (Figures 26 and 27). Additionally, 79.6%

of the interaction records are concentrated in the Nearctic and 13.0% in the Palearctic realms.

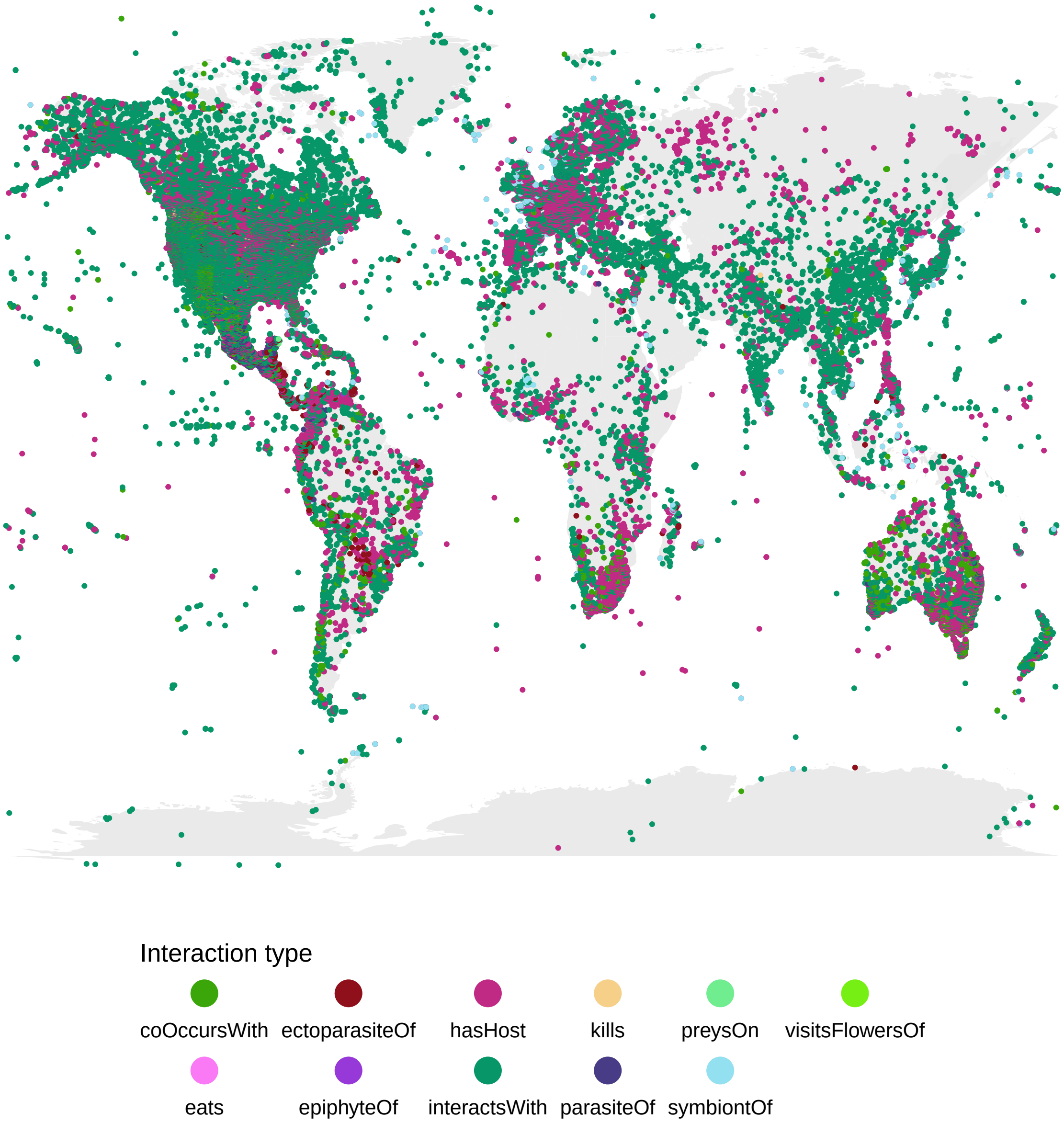


Figure 26: Geographic distribution of interaction records across the globe

The distribution of the interaction records shows high intersection with protected areas of the world. According to World Database on Protected Areas (WDPA) the terrestrial protected areas cover 66,489,319 km^2 of the Earth's surface. The interaction records are found in 31,990,373 km^2 of protected areas (48%), mostly in protected areas of the Nearctic and Palearctic realms.

In addition, the interactions were verified against the International Union for Conservation of Nature (IUCN) Red List (IUCN, 2022) to retrieve the threat category of the species involved in the interactions. For each interaction record, the IUCN Red List

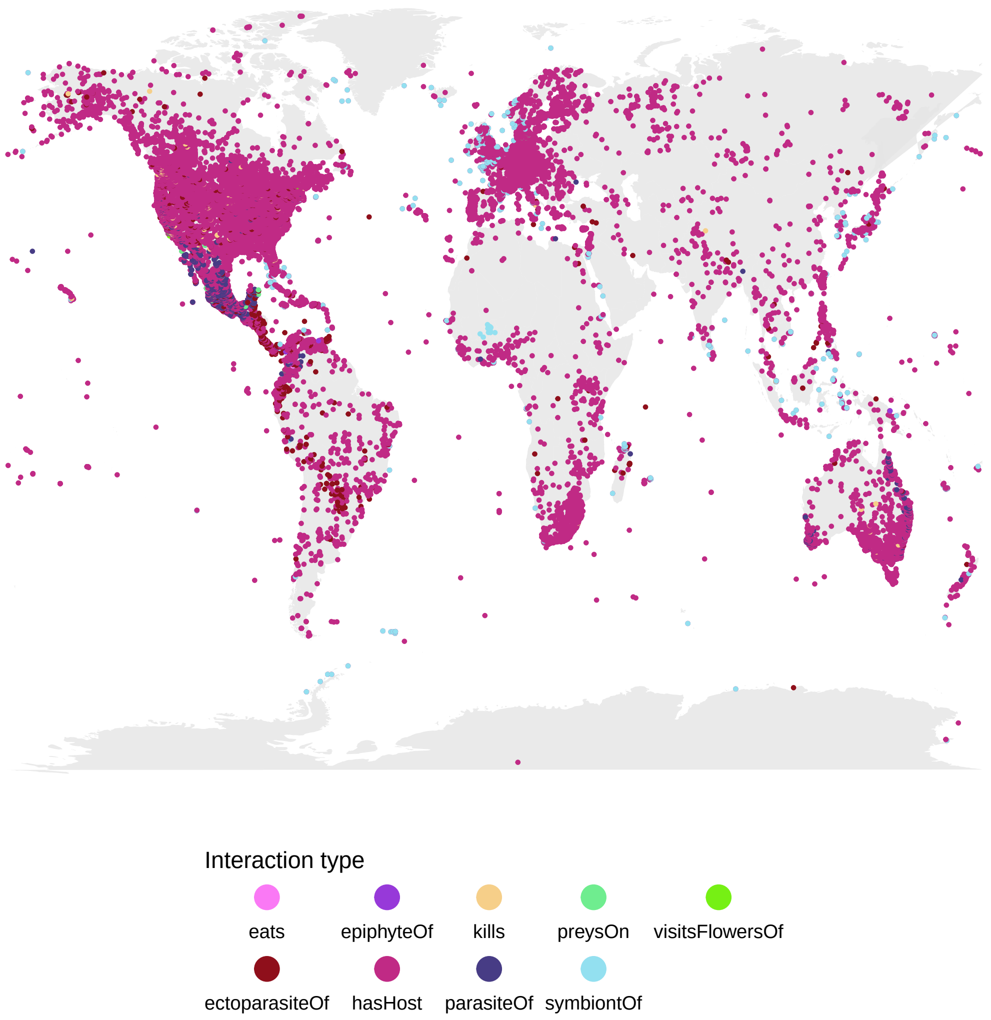


Figure 27: Geographic distribution of interaction records across the globe after removing generic interaction types (i.e., *interacts with* and *co-occurs with*).

threat category was assigned to both the taxa participating in the interaction. Thus, an "interaction threat" was assigned for each interaction record, taking the highest risk between the two species in the interaction record. The IUCN Red List retrieved from GBIF in DwC-A format (IUCN, 2022) contains 254,583 records, where 150,490 are valid taxon names and 104,093 are synonyms. The number of interactions records with category higher than vulnerable (VU) was 8,162, with prevalence of categories vulnerable and endangered (EN) (Table 7).

The classes of Animalia kingdom with the highest number of threatened interactions were Aves (30.1%), Insecta (20.5%), Anthozoa (14.7%) and Mammalia (10.3%). The Amphibia class, which in the IUCN Red List is estimated to have 41% of species threatened with extinction, corresponds to 4.5% of all threatened interaction in the Animalia kingdom. For plants, the classes Magnoliopsida (dicotyledons, 78.6%), Liliopsida (monocotyledons, 8.9%) and Pinopsida (conifers, 8.2%) have the highest number of threatened interactions. In the Magnoliopsida class, Myrtales (20.1%), Caryophyllales (12.2%) and Proteales (10.8%) are the orders with the highest number of threatened interactions. The IUCN threatened categories ($\geq$ VU) for each kingdom is shown in Table 8, Figure 28 shows the classes with the highest number of threatened species, and the complete list of threatened interaction by species level is available in Salim (2023).

Considering biotic interactions from all IUCN categories, 26.7% are inside protected areas, covering an area of 1,677,180.7 km^2. Interactions with high risk of "extinction" (categories equal or higher than vulnerable) correspond to only 0.01% (2.52 million km^2) of the total protected areas.

IUCN Category	**#Species**
EX	5
EW	16
CR	1,236
EN	3,039
VU	3,866
DD	2,643
NT	4,024
LC	280,353
Total	295,182

Table 7: Number of species in the biotic interaction records in the IUCN Red List separated by threat categories

Kingdom	**IUCN Category**	**#Species**
Animalia	EX	3
Animalia	CR	17
Animalia	EN	39
Animalia	VU	97
Fungi	CR	7
Fungi	EN	23
Fungi	VU	29
Plantae	EW	2
Plantae	CR	61
Plantae	EN	175
Plantae	VU	211

Table 8: Number of threatened species in the biotic interaction records for each kingdom, with classified with an IUCN category above or equal to vulnerable (VU)

Despite the low taxonomic coverage of the biotic interactions records when compared with all known forms of life (Section 5.2.1), biotic interactions data for threatened species

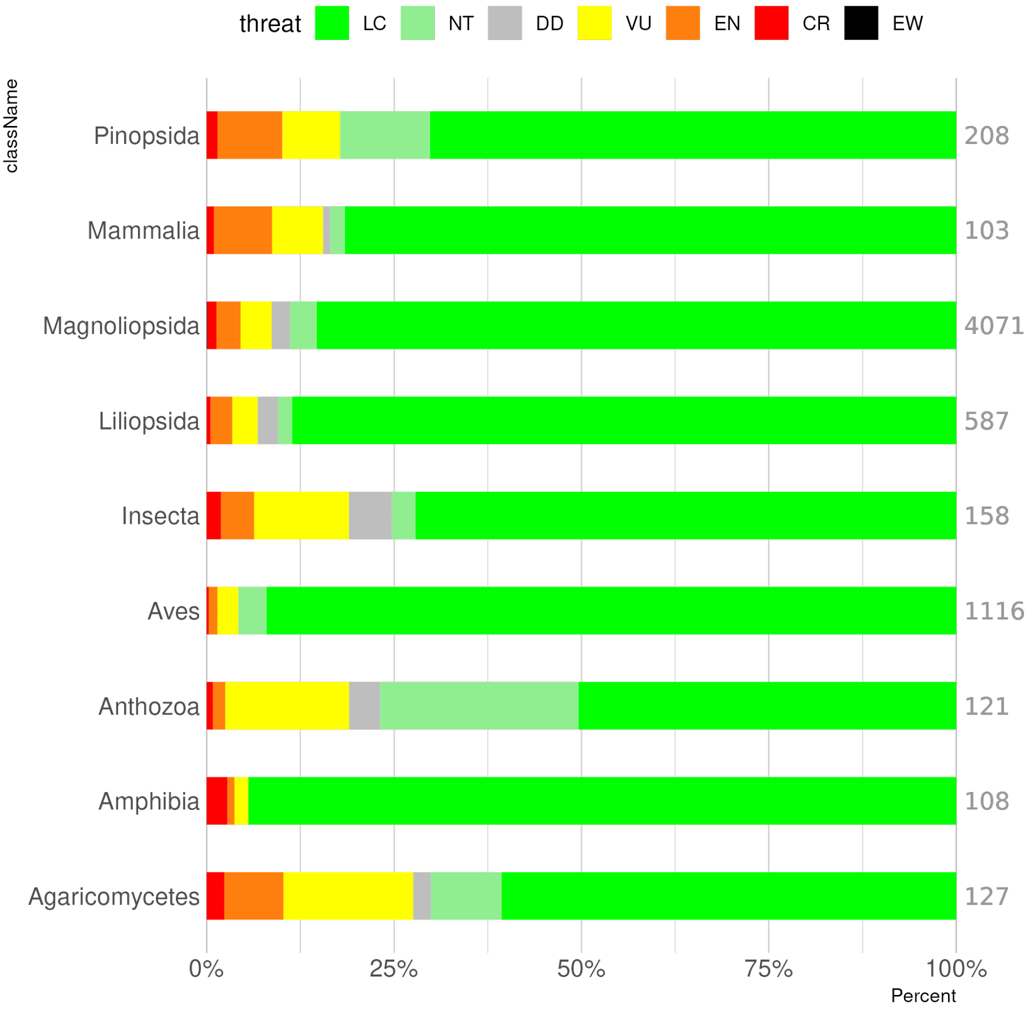


Figure 28: Classes with the highest number of threatened species in the biotic interactions records according to IUCN Red List

on the IUCN Red List are under-represented.

6 BIOTIC INTERACTION DATA STANDARDIZATION

"is undeniable that vast funds are expended on data creation and acquisition. It is false economy, and poor scientific practice, not to ensure that the data are present and useful to all users in the future"

-- JONES *et al.*, 2006

This chapter describes a generic data schema for documenting and sharing biotic interactions data using the DwC standard. The GBIF Unified Model is also presented and compared to the data schema defined here.

6.1 Advancements in standardization of biotic interactions data

The standardization of biotic interactions data began even before the ratification of DwC as a TDWG standard, as shown in Figure 29. The initial proposal for standardization was published on the DwC Wiki in 2007 (Biodiversity Information Standards (TDWG), 2007b). The extension, known as `InteractionExtension`, introduced a new term called `RelationshipType`, which allows for documenting the interaction type in conjunction with the record-level and taxon terms of DwC (Table 9). At that time, the *association terms* and the `dwc:ResourceRelationship` class were not yet part of the standard, and DwC was just a vocabulary of terms without defined classes of terms. One problem with the proposed extension was that, despite the initial proposal being limited to the record-level and taxon terms, it had the potential to be extended to other DwC terms and allowed for an arbitrary number of nested "copies" of DwC records within the same record. Besides the `InteractionExtension`, two other extensions were proposed specifically for plant-pollinator interactions: the `PollinationExtension` (Biodiversity Information Standards (TDWG), 2007c) and the `EnvironmentMeasurementsExtension` (Biodiversity Information Standards (TDWG), 2007a). However, the discussion about the extensions did not progress further, and the proposal was abandoned. This could be due to simultaneous efforts of another group working on what eventually became known as the “TDWG Ontology”. The “TDWG Ontology” introduced the concept of a `TaxonOccurrenceInteraction` in RDF format, specifically designed to document interactions between two `dwc:Occurrence` (TDWG Ontology, 2015).

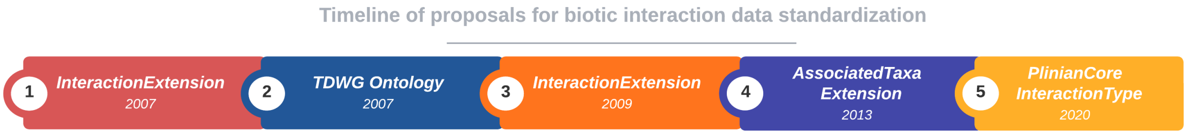

Timeline of proposals for biotic interaction data standardization
1 InteractionExtension 2007
2 TDWG Ontology 2007
3 InteractionExtension 2009
4 AssociatedTaxa Extension 2013
5 PlinianCore InteractionType 2020

Figure 29: Timeline of proposals for biotic interaction data standardization.

Element	Description
Interaction Elements	
RelationshipType	A descriptive term indicating the type of relationship between an organism represented with DarwinCore and the related organism represented with this extension. Examples: VisitedFlowerOf - FlowerVisitedBy, PreyedUpon - PreyedUponBy, DispersedSeedOf - SeedWasDispersedBy, HostOf - ParasiteOf, ExtractedResinFrom - ResinExtractedBy, NestedIn - UsedAsNestBy, PathogenOf - InfectedBy
Record-level Elements	GlobalUniqueIdentifier, BasisOfRecord, InstitutionCode, CollectionCode, CatalogNumber
Taxonomic Elements	ScientificName, HigherTaxon, Kingdom, Phylum, Class, Order, Family, Genus, SpecificEpithet, InfraspecificRank, InfraSpecificEpithet, AuthorYearOfScientificName, NomenclaturalCode
Identification Elements	IdentificationQualifer
Biological Elements	Sex, LifeStage

Table 9: `InteractionExtension` proposed in TDWG Wiki in 2007

The ‘TDWG Ontology” introduced the `TaxonOccurrenceInteraction` class, which includes properties for documenting an interaction between two `TaxonOccurrence` instances. These properties, such as `fromOccurrence` and `toOccurrence`, are used to reference the two occurrences involved in the interaction. Additionally, the property `interactionCategory` is used to describe the type of the interaction. On the other hand, the `TaxonOccurrence` class has the property `hasInteraction` to reference a `TaxonOccurrenceInteraction`. However, in 2013, the TDWG Vocabulary Management Task Group (VoMaG) recommended the deprecation of several terms in the “TDWG Ontology”, including the `TaxonOccurrenceInteraction` (TDWG Vocabulary Management Task Group, 2013). The VoMaG report stated that the terms in the “TDWG Ontology”, including its component vocabularies, lacked extensive semantic restrictions and well-defined relationships typically found in formal ontologies. The same report emphasized that the widespread adoption of Life Science Identifiers (LSID) never occurred, resulting in many parts of the ontology, including the `TaxonOccurrenceInteraction`, being unused. Ultimately, the report recommended clarifying that the "TDWG Ontology" is no longer under active development and should not be used.

Later, in 2009, another proposal for the `InteractionExtension` was presented at the TDWG Conference (SARAIVA *et al.*, 2009). In this proposal, an interaction was represented as an independent record, separate from the occurrences records, by incorporating the global unique identifiers of the two occurrences participating in the interaction (Table 10). A conceptual model was never proposed for this extension. Instead, a publication was made with an XML representation, which exhibited a confusion between the database model and the exchange data model. The extension incorporated three global unique identifiers: one for the interaction record `InteractionGlobalUniqueIdentifier`, and two for the occurrences records (`GlobalUniqueIdentifer1` and `GlobalUniqueIdentifier2`) (CARTOLANO, 2009). The issue with that representation is that the interaction records in the XML were represented similarly to how a *joining table* is represented in a relational database. In other words, it resembled an auxiliary table used to link two other tables in a many-to-many relationship by storing the identifiers of records from both joined tables. The DwC records are designed to be self-contained, meaning that all the data related to a record should be present within that record itself (e.g., flat tables, Simple Darwin Core). Although the introduction of the DwC-A has addressed the issue of representing one-to-many relationships in DwC, the representation of many-to-many relationships did not fit well with the *star-schema* model. The `InteractionExtension` was never formally adopted as an official extension of DwC, but it found practical application for standardizing numerous records within the InterAmerican Biodiversity Information Network–Pollinators Thematic Network (IABIN-PTN) and the Pollination Information Management System of the Food and Agriculture Organization of the United Nations (CARTOLANO, 2009).

Another attempt at standardizing biotic interactions was the "associatedTaxa extension" developed by Encyclopedia of Life (EOL). The "associatedTaxa extension" primarily focused on capturing the taxonomic characteristics of the interactions, rather than their ecological and functional aspects[1]. This DwC extension was created in 2013 and includes 16 new terms in the namespace `aec`. These terms are essentially duplications of some DwC terms, with the prefix `associated` appended to them (Table 11). Although, it is

[1] ⟨http://purl.org/NET/aec/associatedTaxa⟩

Term	Data type	Mandatory
InteractionGlobalUniqueIdentifier	Text	Yes
DateLastModified	DateTime	Yes
GlobalUniqueIdentifer1	Text	Yes
InteractionType	Text	Yes
GlobalUniqueIdentifer2	Text	Yes
RelatedInformation	Text	Yes
LocalityElements	XML elements	No
CollectingElements	XML elements	

Table 10: The `InteractionExtension` as proposed in Saraiva *et al.* (2009)

apparently in use in the Tri-Trophic Thematic Collection Network[2], it has not been formally proposed as a DwC extension nor has it been brought for community discussion in TDWG.

`dwc`	`aec`
associatedTaxa	-
-	associatedOccurenceID
family	associatedFamily
genus	associatedGenus
specificEpithet	associatedSpecificEpithet
scientificName	associatedScientificName
author	associatedAuthor
commonName*	associatedCommonName
-	associatedRelationshipTerm
-	associatedRelatinshipURI
-	associatedNotes
determinedBy	associatedDeterminedBy
condition	associatedCondition
-	associatedLocationOnHost
-	associatedEmergenceVerbatimDate
-	associatedCollectionLocation
-	isCultivar
accessURI	associatedAccessURI
creator	associatedImageCreator
rights	associatedImageRights

Table 11: The "associatedTaxa extension" introduced by Encyclopedia of Life (EOL)

In the Plinian Core (Plinian Core Task Group, 2021), which is a standard for sharing information mainly at the species level (yet to be ratified as a TDWG standard), the `InteractionClass` was defined for documenting biotic interactions. The `InteractionClass`

[2] ⟨https://www.discoverlife.org/tttcn/⟩

is built upon the `dwc:ResourceRelationship` class, and its primary purpose is to document "mutual or reciprocal actions or influences" at species-level. However, it is unclear whether the `InteractionClass` has been deprecated in the latest version of Plinian Core, as indicated in the deprecation list found in the Plinian Core repository[3]. Although, Plinian Core serves a different purpose than DwC, which primarily focuses on documenting occurrence data, the approach adopted in Plinian Core, utilizing the `dwc:ResourceRelationship` class, is currently the recommended method for documenting associations using DwC[4].

After investigating the advantages and disadvantages of each of the methods described above, a generic data schema was elaborated and is detailed in the following section. It not only incorporates all the advantages of the previous proposals but also addresses the disadvantages associated with them. The data model includes independent representation of interaction records separate from occurrence records, the ability to include additional data about the interactions such as effects and outcomes, documentation of organism-level and species-level interactions, and importantly, it is purely based on DwC without introducing any new terms into the data model.

6.2 A generic data schema for biotic interactions

The main concern regarding the approaches outlined in Section 6.1 is that the interactions themselves are not the primary focus of the data documentation. Instead, they are derived from associations established between occurrences or taxon names.

The *co-action* definition proposed by Haskell (1949), further refined by Lidicker (1979) for biotic interactions, and the more recent, the biotic interactions model of *interaction events* introduced by Gómez, Iriondo & Torres (2023), provide the basis for the development of a data schema having the interactions as the core information. The `dwc:Event` class of the DwC standard is defined as "an action that occurs at some location during some time"[5]. This generic definition of an event encompasses the concepts of both *co-actions* and *interaction events*. By adopting such a generic definition, the DwC standard enables for flexibility in capturing various types of events, including biotic interactions.

In the proposed data schema, instances of `dwc:Event` class serve as representations

[3] ⟨https://github.com/tdwg/PlinianCore/blob/master/Deprecated/terms/txt/plinCelementList_18-07-2012.txt⟩

[4] see Notes of the `associatedTaxa` term in DwC: ⟨http://rs.tdwg.org/dwc/terms/associatedTaxa⟩

[5] ⟨http://rs.tdwg.org/dwc/terms/Event⟩

of interaction events. These instances capture essential information about the interaction events, such as temporal data and sampling details. Additionally, geographic information can be included in the event using terms from the `dwc:Location` class [6]. The interacting organisms or taxa are represented by their respective instances of the `dwc:Occurrence` and `dwc:Taxon` classes. These classes serve as the basis of documenting data about individual organisms or species involved in the interactions. By linking these instances to an instance of `dwc:Event` the data schema enables the representation of a pairwise interaction. As highlighted in Section 2.1, it is common that biotic interactions are sampled with a particular interest in organisms' traits and the effects and outcomes of the interactions in which they participate. In DwC it is possible to include these data using the `dwc:MeasurementOrFact` class, but more complex representations can be achieved using extensions like the Extended Measurement or Fact (eMoF)(POOTER *et al.*, 2017), developed by Ocean Biodiversity Information System (OBIS), or the Ecological Trait-data Standard Vocabulary (SCHNEIDER *et al.*, 2019). The eMoF is particularly useful when using the DwC-A format because it addresses the limitations of the *star-schema* in representing one-to-many relationships. Instances of `dwc:MeasurementOrFact` class, or its extensions, can be associated with instances of the `dwc:Event` class to represent interaction effects and outcomes. Similarly, these measurements or facts cat also be linked to instances of the `dwc:Occurrence` class, representing the traits of specific organisms or effects of the interaction on an individual organism or group of organisms.

The proposed data schema goes beyond the simplicity of using the `dwc:associatedTaxa` term or the `dwc:ResourceRelationship` in conjunction with `dwc:Occurrence`. It offers a more comprehensive and standardized approach to documenting and sharing biotic interaction data. This schema recognizes the importance of context-dependent information, which drives biotic interaction in nature (HOEKSEMA *et al.*, 2010; BUTTERFIELD; CALLAWAY, 2013; CHAMBERLAIN; BRONSTEIN; RUDGERS, 2014; MARON; BAER; ANGERT, 2014; HOEKSEMA; BRUNA, 2015; FREDERICKSON, 2017). It is important to emphasize that the proposed schema is not intended for direct use by end users (e.g., researchers, data authors). Instead, its implementation is intended for data providers for facilitating data entry, such as Integrated Publishing Toolkit and GloBI. Systems can provide multiple abstraction layers to simplify user input and facilitate data entry. An example of this is presented in Chapter 8. By adopting a common schema, data can be easily exchanged and transformed into different internal representations.

However, in order to ensure the reusability of biotic interactions data in different

[6] ⟨http://purl.org/dc/terms/Location⟩

contexts, it is crucial to populate a minimal set of terms with relevant data. Ideally, the more information provided, the better. But, depending on the amount of data collected or the specific objectives of the studies, only a portion of the terms can be filled. Therefore, to ensure a minimum level of essential information, interaction records should include the following data, referred by Jordano (2021) as "tetranomials":

a) `dwc:Event`: the unique identifier of the event `dwc:eventID`, which is used to link the `dwc:Occurrence`'s to the same interaction event;

b) `dwc:Occurrence`: at least the taxonomic data must be present (e.g., `dwc:scientificName`);

c) `dwc:ResourceRelationship`: the type and direction of the interactions (e.g., `resourceID`, `relatedResourceID`, `relationshipOfResource`).

The data schema depicted in Figure 30 is agnostic regarding the serialization format used to represent data (e.g. DwC-A, RDF, XML). Depending on the serialization format, different approaches are considered. The DwC standard was originally designed to handle tabular data, and later, expanded to support Linked Data (BASKAUF; WEBB, 2016). However, certain terms within the standard may not be applicable in an RDF context. The following sections present how the data schema can be serialized in different formats: DwC-A, XML and RDF.

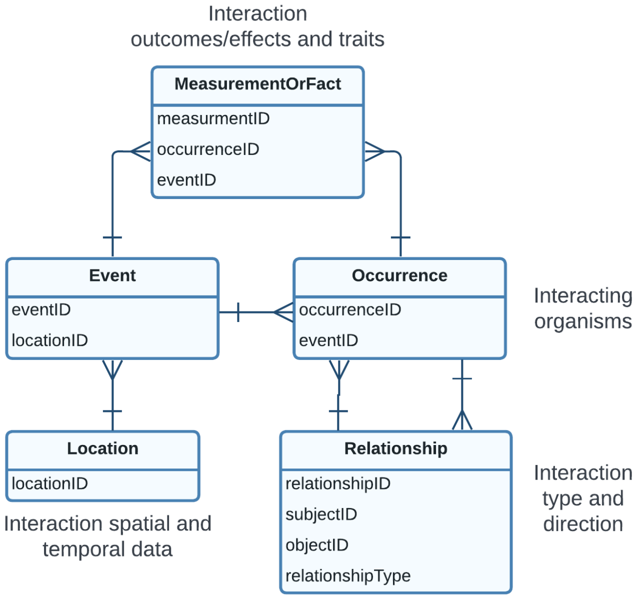


Figure 30: Conceptual model of the proposed data schema showing the main entities and their relationships in the context of biotic interactions data

6.2.1 Biotic interactions data as Darwin Core Archive

The DwC-A "*is a biodiversity informatics data standard that makes use of the Darwin Core terms to produce a single, self-contained dataset for sharing species-level (taxonomic), species-occurrence data, and sampling-event data*" (GBIF, 2021). It is a platform-independent machine-readable format and it is organized in a *star-schema*. The star-schema consists of a central/core file linked to multiple extension files through record identifiers in the core file (Figure 31). In addition to the data files, a DwC-A includes two XML files: the `meta.xml` file and the `eml.xml` file. The `meta.xml` file describes the relationship between the data files within the DwC-A and defines the mappings between data fields and terms in the DwC standard. It provides information on the structure and organization of the dataset. The `eml.xml` file contains metadata about the dataset using the Ecological Metadata Language (EML) (JONES *et al.*, 2019). This file includes information such as the dataset's title, creator, publication details, license and other descriptive attributes of the dataset.

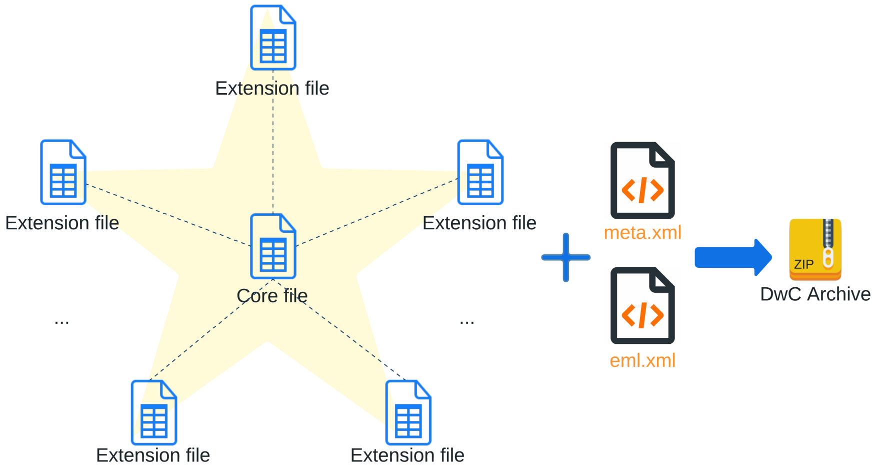


Figure 31: A diagram illustrating the star-schema used in Darwin Core Archive (DwC-A)

However, the star-schema has a limitation in that all the extension files within a DwC-A must be linked to the core table, thus extensions themselves cannot be linked to other extensions directly ("subextensions"). One possible solution is to associate the "subextensions" to the core table in the same manner as normal extensions (i.e., using the core identifiers). However, an additional field must be included to link the "subextension" to the parent extension. Although, this introduces some redundancy in the relationships between the tables, it does not violate the star-schema definition. Actually, this is the solution employed by some DwC extensions (POOTER *et al.*, 2017; WIECZOREK *et* *al.*, 2014). While this approach is not mandatory for documenting biotic interactions in DwC-A, it becomes necessary when occurrences have associated measurements or facts which need to be linked to both the core file (`dwc:Event`) and the occurrence extension (`dwc:Occurrence`).

The DwC-A model for biotic interactions data, depicted in Figure 32, is built upon the widely used "sampling event data model"[7]. In the proposed model, similar to the "sampling event data model", the core file in the DwC-A contains rows that correspond to instances of the `dwc:Event` class (e.g., 12). The core file is linked to the `dwc:Occurrence` extension, which documents the occurrences of the interacting organisms or group of organisms (e.g., Table 13). The extension `dwc:ResourceRelationship` is used for documenting the interaction type and direction. Although it may seem redundant to have both instances of `dwc:ResourceRelationship` and `dwc:Occurrence` linked to the core `dwc:Event`, it is necessary due to the limitations of the star-schema (i.e., direct linking between `dwc:Occurrence` and `dwc:ResourceRelationship` is not possible in the star-schema). The terms `dwc:resourceID` and `dwc:relatedResourceID` from the `dwc:ResourceRelationship` class are used to designate the subject of the interaction (also called source in graph theory) and the object of the interaction (also called target in graph theory), respectively. In conjunction with the term `dwc:relationshipOfResource`, and its non-literal form `dwc:relationshipOfResourceID`, this terms reflect the type and direction of a interaction (e.g., Table 14).

eventID	**eventDate**	**locality**	**decimalLatitude**	**decimalLongitude**
evt:0001	2008-02-01	Parque Municipal de Mucugê, Chapada Diamantina, Bahia, Brasil	–12.98833333	-41.34083333)
evt:0002	2008-02-15	Jaqueira, Pernambuco	-8.71138889	-35.84166667

Table 12: Example of an event table (`dwc:Event` instances) of a DwC-A containing biotic interaction data

In addition to documenting the "tetranomials" (who interacts with whom), the data schema allows for the inclusion of organisms' traits and interaction outcomes and effects. This is achieved by using the `dwc:MeasurementOrFact` class and its extensions, such as eMoF. The `dwc:MeasurementOrFact` class is used to represent one-to-many relationships between the `dwc:Event` class and various characteristics of the interactions. However,

[7] ⟨https://www.gbif.org/sampling-event-data⟩

eventID	occurrenceID	scientificName	individualCount
evt:0001	occ:0001	Walteria cinerescens	1
evt:0001	occ:0002	Augastes lumachella	1
evt:0002	occ:0003	Aechmea fulgens	1
evt:0002	occ:0004	Thalurania watertonii	2

Table 13: Example of an occurrence table (`dwc:Occurrence` instances) of a DwC-A contain biotic interaction data

eventID	resourceRelationshipID	resourceID	relationshipOfResource	relatedResourceID
evt:0001	int:0001	occ:0001	pollinatedBy	occ:0002
evt:0002	int:0002	occ:0003	flowerVisitedBy	occ:0004

Table 14: Example of a resource relationship table (`dwc:ResourceRelationship` instances) of a DwC-A containing biotic interaction data

when using DwC-A, the `dwc:MeasurementOrFact` class cannot be used to document measurements or facts regarding the characteristics of the interacting organisms (e.g., traits). This limitation arises from the constraints imposed by the star-schema, as discussed earlier. In such cases, the eMoF extensions should be used. The eMoF was specially designed to handle environmental data in conjunction with occurrence data. It extends the `dwc:MeasurementOrFact` class by incorporating the `dwc:occurrenceID` term along with three additional terms: `obis:measurementTypeID`, `obis:measurementValueID`, and `obis:measurementUnitID`. The `dwc:occurrenceID` term is used to circumvent the limitations of the star-schema and associate measurement or fact records in the eMoF extension to occurrence records in the `dwc:Occurrence` extension. The additional terms are used to constrain and standardize the measurement types, values, and units. Unlike the unconstrained terms of the `dwc:MeasurementOrFact` class, these new terms require use of controlled vocabularies referenced by URIs.

6.2.2 Biotic interactions data as XML

The implementation based on DwC using XML is similar to the implementation using DwC-A. However, in XML, one-to-many relationships can be handled naturally compared to the limitations of the star-schema. Thus, there is no need to use the eMoF extension to document additional characteristics of the `dwc:Occurrence` instances. Although the eMoF can be useful for referencing measurement types, values, and units using URIs. Instead, the `dwc:MeasurementOrFact` class should be used, providing the `dwc:measurementID` and the respective `dwc:eventID` and `dwc:occurrenceID` which the measurement is linked to. Lines 71-76 in Code 1 show an example of measurement or

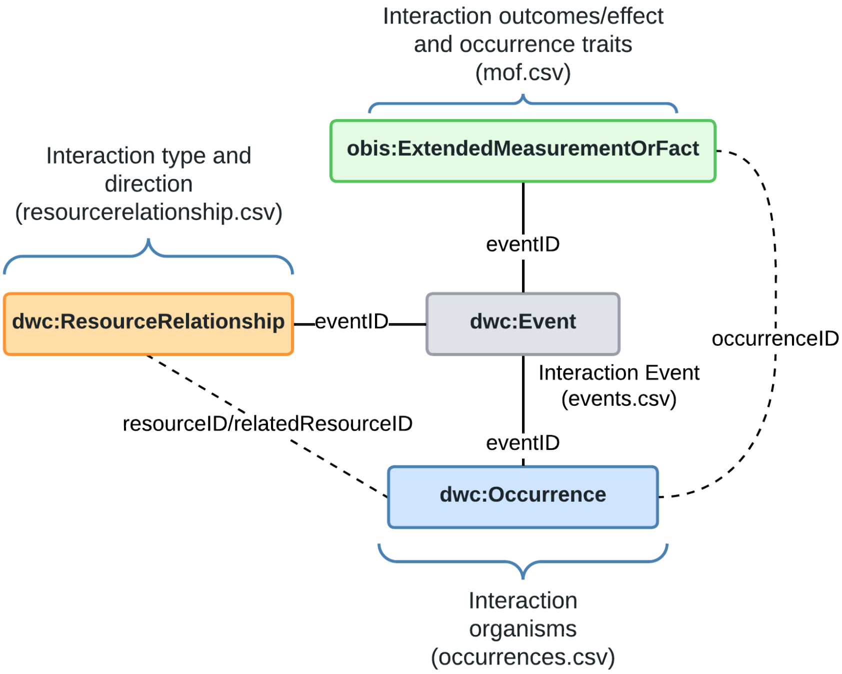


Figure 32: Diagram illustrating the contents of a `meta.xml` file of DwC-A for biotic interaction data

fact for a `dwc:Event`, and lines 78-101 in Code 1 show three different measurements for the `dwc:Occurrences`. It is important to note that the `dwc:occurrenceID`) is used to link `dwc:MeasurementOrFact` to the respective `dwc:Occurrence`. Because the DwC XML does not define any constraint regarding the duplication of "ID terms" within an XML element representing a DwC class (Darwin Core Maintenance Group, 2021), it is possible for multiple instances of the `dwc:MeasurementOrFact` class can to refer to the same `dwc:Occurrence` or `dwc:Event` instances.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <dwr:DarwinRecordSet xmlns:dwr="http://rs.tdwg.org/dwc/dwcrecord/"
3    xmlns:dcterms="http://purl.org/dc/terms/"
4    xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
5    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
6    xsi:schemaLocation="http://rs.tdwg.org/dwc/dwcrecord/
7    http://rs.tdwg.org/dwc/xsd/tdwg_dwc_classes.xsd">
8     <dwc:Occurrence>
9       <dwc:occurrenceID>INT:OCC:00101</dwc:occurrenceID>
10      <dwc:basisOfRecord>HumanObservation</dwc:basisOfRecord>
```

```
    <dwc:recordedBy>José A. Salim</dwc:recordedBy>
    <dwc:eventID>INT:EVENT:0001</dwc:eventID>
  </dwc:Occurrence>
  <dwc:Occurrence>
    <dwc:occurrenceID>INT:OCC:00102</dwc:occurrenceID>
    <dwc:basisOfRecord>HumanObservation</dwc:basisOfRecord>
    <dwc:recordedBy>José A. Salim</dwc:recordedBy>
    <dwc:eventID>INT:EVENT:0001</dwc:eventID>
  </dwc:Occurrence>
  <dwc:Event>
    <dwc:eventID>INT:EVENT:0001</dwc:eventID>
    <dwc:eventDate>2000-01-05</dwc:eventDate>
    <dwc:locationID>INT:LOC:0001</dwc:locationID>
  </dwc:Event>
  <dcterms:Location>
    <dwc:locationID>INT:LOC:0001</dwc:locationID>
    <dwc:country>Brazil</dwc:country>
    <dwc:countryCode>BR</dwc:countryCode>
    <dwc:decimalLatitude>-47.0680352</dwc:decimalLatitude>
    <dwc:decimalLongitude>-22.8261888</dwc:decimalLongitude>
  </dcterms:Location>
  <dwc:Identification>
    <dwc:identifiedBy>John Doe</dwc:identifiedBy>
    <dwc:dateIdentified>2000-01-05</dwc:dateIdentified>
    <dwc:occurrenceID>INT:OCC:00101</dwc:occurrenceID>
    <dwc:taxonID>https://www.gbif.org/species/5293403</dwc:taxonID>
  </dwc:Identification>
  <dwc:Taxon>
    <dwc:taxonID>https://www.gbif.org/species/5293403</dwc:taxonID>
    <dwc:scientificName>Euterpe edulis Mart.</dwc:scientificName>
    <dwc:taxonRank>species</dwc:taxonRank>
    <dwc:genus>Euterpe</dwc:genus>
    <dwc:specificEpithet>edulis</dwc:specificEpithet>
  </dwc:Taxon>
  <dwc:Identification>
    <dwc:identifiedBy>John Doe</dwc:identifiedBy>
```

```
      <dwc:dateIdentified>2000-01-05</dwc:dateIdentified>
      <dwc:occurrenceID>INT:OCC:00102</dwc:occurrenceID>
      <dwc:taxonID>https://www.gbif.org/species/1341976</dwc:taxonID>
    </dwc:Identification>
    <dwc:Taxon>
      <dwc:taxonID>https://www.gbif.org/species/1341976</dwc:taxonID>
      <dwc:scientificName>Apis mellifera (Linnaeus, 1758)</dwc:scientificName>
      <dwc:taxonRank>species</dwc:taxonRank>
      <dwc:genus>Apis</dwc:genus>
      <dwc:specificEpithet>mellifera</dwc:specificEpithet>
    </dwc:Taxon>

    <dwc:ResourceRelationship>
      <dwc:eventID>INT:EVENT:0001</dwc:eventID>
      <dwc:resourceID>INT:OCC:00101</dwc:resourceID>
      <dwc:relationshipOfResource>
        has flowers visited by
      </dwc:relationshipOfResource>
      <dwc:relationshipOfResourceID>
        http://purl.obolibrary.org/obo/RO_0002623
      </dwc:relationshipOfResourceID>
      <dwc:relatedResourceID>REBIPP:OCC:00102</dwc:relatedResourceID>
    </dwc:ResourceRelationship>

    <dwc:MeasurementOrFact>
      <dwc:measurementID>INT:MOF:0001</dwc:measurementID>
      <dwc:eventID>INT:EVENT:00001</dwc:eventID>
      <dwc:measurementType>resourceCollected</dwc:measurementType>
      <dwc:measurementValue>pollen</dwc:measurementValue>
    </obis:ExtendedMeasurementOrFact>

     <dwc:MeasurementOrFact>
      <dwc:measurementID>INT:MOF:0002</dwc:measurementID>
      <dwc:eventID>INT:EVENT:00001</dwc:eventID>
      <dwc:occurrenceID>INT:OCC:00101</dwc:occurrenceID>
      <dwc:measurementType>habit</dwc:measurementType>
    <dwc:measurementValue>whole plant arborescent</dwc:measurementValue>
  </dwc:MeasurementOrFact>

  <dwc:MeasurementOrFact>
    <dwc:measurementID>INT:MOF:0003</dwc:measurementID>
    <dwc:eventID>INT:EVENT:00001</dwc:eventID>
    <dwc:occurrenceID>INT:OCC:00101</dwc:occurrenceID>
    <dwc:measurementType>flowerLongevity</dwc:measurementType>
    <dwc:measurementValue>168</dwc:measurementValue>
    <dwc:measurementUnit>days</dwc:measurementUnit>
  </dwc:MeasurementOrFact>

  <dwc:MeasurementOrFact>
    <dwc:measurementID>INT:MOF:0004</dwc:measurementID>
    <dwc:eventID>INT:EVENT:00001</dwc:eventID>
    <dwc:occurrenceID>INT:OCC:00102</dwc:occurrenceID>
    <dwc:measurementType>caste</dwc:measurementType>
    <dwc:measurementValue>worker</dwc:measurementValue>
  </dwc:MeasurementOrFact>
</dwr:DarwinRecordSet>
```

Listing 1: Example of documenting biotic interaction data using DwC XML.

6.2.3 Biotic interactions data as Resource Description Framework

During the World Wide Web Consortium (W3C) meeting in 1997, Tim Berners-Lee formulated the two major goals of what is now known as *Semantic Web*: enabling people to work collaboratively by allowing them to share knowledge and incorporating tools to assist people analyzing and managing information in a meaningful way (BERNERS-LEE, 1997). The vision of the Semantic Web is to enhance the annotations of resources with *semantic markups*, making them easily interpreted by machines. To this end, the Resource Description Framework (RDF) (CYGANIAK; WOOD; LANTHALER, 2014) was developed by the W3C to support the creation, exchange and annotation of Web resources. RDF defines a data model which serves for annotation in the Semantic Web. The core structure of the data model is a set of triples (Figure 33), each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph (CYGANIAK; WOOD; LANTHALER, 2014).

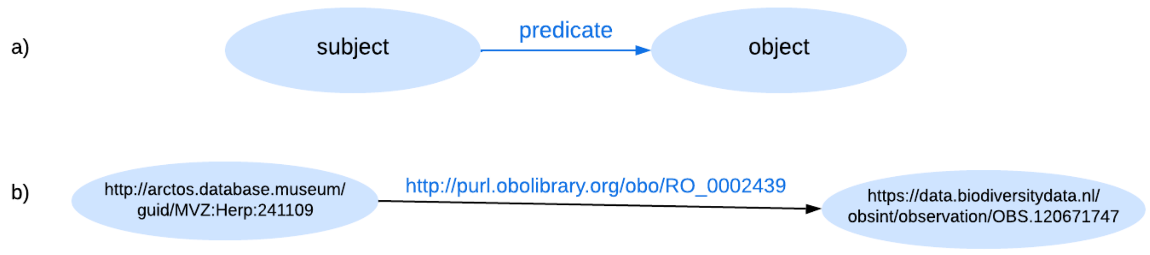

a)
subject
predicate
object
b)
http://arctos.database.museum/ guid/MVZ:Herp:241109
http://purl.obolibrary.org/obo/RO_0002439
https://data.biodiversitydata.nl/ obsint/observation/OBS.120671747

Figure 33: Example of an RDF triplet

In short, any Internationalized Resource Identifier (IRI) or literal (e.g., strings, numbers, dates) are called resources. Anything can be a resource, including physical things (e.g., specimens), documents (e.g., publications), abstract concepts (e.g., interaction), numbers and strings.

Though RDF has been adopted by some communities (MORITZ *et al.*, 2011; BASKAUF; WEBB, 2016; PAGE, 2016; PENEV *et al.*, 2019), its usage is still in the early stages by the biodiversity community. The DwC RDF Guide (Darwin Core and RDF/OWL Task Groups, 2015) was created in 2015, but its implementation is still timid compared to other implementations (e.g., DwC-A), mostly because some parts of the DwC standard lack formal definitions and recommendations on how to express all the potential uses of DwC in RDF (e.g. `dwc:ResourceRelationship`, agents annotation).

DwC defines a number of "ID terms" intended to designate identifiers (e.g. `dwc:occurrenceID`, `dwc:taxonID`, `locationID`). The "ID terms" serve two functions: specifying the class of the resource being described or referenced and, indicating that, the value of the term is an identifier. However, in Resource Description Framework, these functions are handled separately using the `rdf:type` declarations and URIs for expressing the identifier of the resource. For most DwC "ID terms", the `dcterms:identifier` in Dublin Core (DCMI Usage Board, 2020) can be used as replacement to indicate the identifier of an RDF resource. But the same can not be applied to `dwc:ResourceRelationship` class. The `dwc:ResourceRelationship` class defines additional "ID terms", besides the resource identifier (`dwc:resourceRelationshipID`): `dwc:resourceID` ("an identifier for the resource that is the subject of the relationship"), `dwc:relatedResourceID` ("an identifier for a related resource; the object, rather than the subject of the relationship"), `relationshipOfResourceID` ("an identifier for the relationship type (predicate) that connects the subject identified by resourceID to its object identified by relatedResourceID").

Because the usage of `dcterms:identifier` term implies "an unambiguous reference to the resource within a given context" (DCMI Usage Board, 2020), it is not clear whether `dwc:ResourceRelationship` would make sense in the context of RDF, and probably it does not (BASKAUF; WEBB, 2016; GitHub, 2023). The `dcterms:identifier` would make no distinction between each term it serves as a replacement for when more than one "ID term" is used within the same resource.

Baskauf & Webb (2016) proposed an RDF vocabulary, called Darwin Core Semantic Web (DwC-SW), to complement the DwC RDF terms[8], which are intended for use exclusively with non-literal objects (i.e., IRI). The DwC IRI enables exposing data in the form of RDF using DwC terms, including biotic interactions data.

In the context of RDF, the biotic interactions are still represented using the `dwc:Event` class. However, the relationships between instances of `dwc:Occurrence` and `dwc:Event` are documented using the `dsw:atEvent` predicate from the DwC-SW, as illustrated in Appendix E. The type and direction of an interaction are given naturally by the RDF triplet composed by two instances of the `dwc:Occurrence` class (one as subject and another as object) and a term from the RO as predicate (or any other vocabulary and ontology which defines biotic interaction types).

In addition, interaction outcomes, effects and organism traits (and any other characteristics of an interaction) are also represented using the `dwc:MeasurementOrFact`, but in RDF the terms from the DwC IRI should be used. Instead of using the term `dwc:measurementType`, the `dwciri:measurementType` term should be used for non-literal objects. The same is valid for other terms in DwC which have analog terms in `dwciri`: `dwc:measurementUnit`, `dwc:measurementDeterminedBy` and `dwc:measurementMethod`. The exception is the `dwc:measurementValue` which does not have a correspondent term in DwC IRI. Because it is used to represent measurements values or facts, it will mostly contain literal objects, but there are cases where terms from controlled vocabulary can be used with non-literal objects. In RDF the `dwc:measurementValue` should be used with the attribute `rdf:datatype`, which defines the data type of the object (value) for the term (e.g. `xsd:int` for integer, `xsd:string` for text). Thus, when using a term from a controlled vocabulary, the value of the attribute `rdf:datatype` must be set to `rdf:Resource`, indicating that the data type of `dwc:measurementValue` is an RDF resource (i.e., non-literal value).

The relationships between RDF resources (e.g., `dwc:Event`, `dwc:Occurrence`) are

[8] namespace ⟨http://rs.tdwg.org/dwc/iri/⟩

represented using the term `dcterms:relation` from Dublin Core (DCMI Usage Board, 2020), following the Darwin Core RDF Guide (Darwin Core and RDF/OWL Task Groups, 2015).

The Linked Data model is in its early stages of adoption by the biodiversity community (BASKAUF; WEBB, 2016). As the use of RDF grows, the RDF representation of relationships between resources can be revised to properly meet the requirements of the linked open data principles.

6.3 The GBIF Unified Model

The decision to prioritize flexibility and simplicity during the ratification of DwC as a standard (WIECZOREK *et al.*, 2012) resulted in the absence of explicit relationships among the classes, as commonly found in formal ontologies (BASKAUF; WEBB, 2016). The flexibility of the standard allows that any relationship between DwC classes can be created and implemented in different formats (e.g. RDF, XML, DwC-A). It poses many challenges for data aggregators, such as GBIF. In order to ingest as much data as possible, data aggregators face the challenge of handling different schemas and implementing specialized methods for each schema. Currently, GBIF fully supports three types of biodiversity data: checklist data, occurrence data and sampling event data. The checklist data are assumed to have `dwc:Taxon` core, the occurrence data are mapped to `dwc:Occurrence` core and the sampling event data to the `dwc:Event` core. Although, GBIF accepts datasets structured in different schemas (e.g ANDERSSON *et al.*, 2021), not all terms and data relationships within those schemas are fully supported, indexed and aggregated.

The "Unified Data Model" under development by GBIF seeks to overcome these barriers by the investigation of a wide range of real use cases (e.g., camera trap, biotic interactions, eDNA metabarcoding). The goal of the new model is to define a common schema to be adopted by the data provider to accommodate various types of biodiversity data. A draft schema has been presented to the biodiversity community, revealing that the new model is a extension of the existing sampling event data model. The generality of the `dwc:Event` class allows that instances of the class represent any kind of action (i.e., "an action that occurs at some location during some time"), and then, it is a key aspect in the new model.

Although the development of the "Unified Data Model" is still ongoing, the current data schema for biotic interactions proposed in the GBIF model shares similarities with the schema proposed in this study, which in turn, is a further development of a schema proposed earlier in the Brazilian Network on Plant-Pollinator Interactions (REBIPP) community[9] and in the TDWG Interest Group on "Biological Interactions"[10]. This is a sign of alignment and progress towards a unified approach to representing biotic interactions in biodiversity data.

The upcoming release of the new model is expected to have a significant impact on how biodiversity data are shared, enabling richer data documentation. However, the current state of the model does not clearly differentiate between the abstract model and the application profile. This indicates that the primary focus of the model's development is centered around its compatibility with the DwC-A format. As the usage of RDF grows in the biodiversity community, RDF graph models will need to be developed considering different use cases, similarly to what GBIF has been doing with the "Unified Data Model".

At the present time, there is no official data schema specifically dedicated to documenting biotic interactions data in the biodiversity community. However, the proposed schema holds great potential to serve as a solution or serve as a starting point for the development of a definitive solution for a broader community.

[9] https://github.com/rebipp/ppi/issues/60

[10] https://github.com/tdwg/interaction/issues/24

7 A VOCABULARY OF TERMS FOR PLANT-POLLINATOR INTERACTIONS

This chapter outlines the process of developing a vocabulary of terms for creating standardized plant-pollinator interactions datasets. The aims of the vocabulary are to provide a set of terms for documenting organism traits and interaction outcomes and effects.

First, it describes a historical introduction providing an overview of previous studies which have focused on the standardization of plant-pollinator interaction data. Subsequently, it describes the detailed process of vocabulary development, emphasizing a community-driven approach.

7.1 Introduction

The data schema presented in the Chapter 6 is sufficiently generic to be adopted for documenting plant-pollinator interactions data. However, to fully exploit the capabilities of DwC and the proposed data schema, the organisms' traits and interaction outcomes and effects data should be standardized using community-specific vocabularies. The vocabularies of terms are useful to provide standardized data for measurement types, values and units in conjunction with `dwc:MeasurementOrFact` class and its extensions.

For that reason, one of the initial activities of the Brazilian Network on Plant-Pollinator Interactions[1] (REBIPP in Portuguese) was to develop a solution for the standardization of plant-pollinator interactions data. REBIPP is an open collaborative network of experts in Pollination Biology, founded in 2016, with a focus on the study of plant-pollinator interactions in their various dimensions. The network aims to i) generate diagnosis of plant-pollinator interactions in Brazil (WOLOWSKI *et al.*, 2019), ii) integrate knowledge in pollination of natural and anthropogenic ecosystems, iii) identify knowledge gaps, iv) support public policy for ecosystem services and food production and,

[1] ⟨https://www.rebipp.org.br⟩

v) encourage collaborative research among members of the network.

The current vocabulary of terms has its roots in previous initiatives which elaborated vocabularies and data models for documenting plant-pollinator interactions. The vocabulary began to be assembled in 2006 during the creation of the InterAmerican Biodiversity Information Network–Pollinators Thematic Network (IABIN-PTN). The IABIN-PTN contributed to the digitization of many datasets on plant-pollinator interactions. In a joint effort with Food and Agriculture Organization of the United Nations (FAO), the first model proposed defined three extensions to the emerging DwC standard (not yet a TDWG ratified standard). The extensions consisted of i) an *InteractionExtension* (Biodiversity Information Standards (TDWG), 2007b) intended to document generic interactions between two `dwc:Occurrence` instances (described in Section 6.1), ii) a *Pollination Extension* (Biodiversity Information Standards (TDWG), 2007c) which included terms for documenting pollen and nectar removal and iii) an `Environmental Measurement Extension` (Biodiversity Information Standards (TDWG), 2007a) which defined terms for documenting the environmental conditions during the observation of an interaction. At that time, a discussion within the TDWG community regarding the creation of "TDWG Ontology" involving a broader scope of biotic interactions data did not resolve to a consensus (TDWG Ontology, 2015), and given the time constraints of the IABIN-PTN project, only the plant-pollinator interaction were considered. The proposal received criticism regarding the creation of new terms with identical definitions but outside the official DwC namespace. Additionally, some terms included in the extensions were domain-specific, which contradicted the general nature of the DwC standard, designed to be applicable across various biodiversity domains. Another concern was the inclusion of terms from different DwC classes within the *InteractionExtension*.

In a subsequent version, the simplification of the previous model converged to document each interaction as a record of type `InteractionExtension` including data about the observers, location, date/time and bibliographic references for two interacting instances of `dwc:Occurrence`s class. According to the discussion presented in Section 6.1, this model has two disadvantages. First, the usage of the `InteractionExtension` enforces the usage of the `dwc:Occurrence` as the central entity in the model, which imposes some challenges to documenting characteristics of the interaction, such as, outcomes and effects, especially in the context of DwC-A. A solution for this may require new terms to be added to the extension covering all possible concepts regarding interaction outcomes and effects, leading to a unnecessary growth of the extension, or the creation of a extension similar to `emof`, but restricted only to be used with the `InteractionExtension`.

However, as previously mentioned in Section 6.1, this model contributed to digitization and standardization of data collected within IABIN-PTN and the FAO Global Pollination Project, a collaborative effort involving United Nations Environment Programme (UNEP) and Global Environment Facility (GEF) (CARTOLANO, 2009). Together, these projects were responsible for digitization and standardization of thousands of plant-pollinator interactions datasets, enhancing data accessibility and reusability.

The `PollinationExtension` introduced four new terms:

a) `PollinationEvidence`: indicates if pollination occurred. Use "1" if the event happened, "0" if the event did not happen and "0.5" to indicate uncertainty. NULL should be used if there was no attempt to obtain this information;

b) `PollenRemoval`: indicates if pollen was removed. Use "1" if the event happened, "0" if the event did not happen and "0.5" to indicate uncertainty. NULL should be used if there was no attempt to obtain this information;

c) `NectarRemoval`: indicates if nectar was removed. Use "1" if the event happened, "0" if the event did not happen and "0.5" to indicate uncertainty. NULL should be used if there was no attempt to obtain this information;

d) `OilRemoval`: indicates if oil was removed. Use "1" if the event happened, "0" if the event did not happen and "0.5" to indicate uncertainty. NULL should be used if there was no attempt to obtain this information;

e) `FlowerPredation`: indicates if at least part of the flower was consumed. Use "1" if the event happened, "0" if the event did not happen and "0.5" to indicate uncertainty. NULL should be used if there was no attempt to obtain this information.

From an ontological perspective, events are generally considered to be discrete occurrences that either happen or do not happen. There is typically no middle ground or partial occurrence of an event. The basic evidence for the occurrence of an event is often the existence of an object or entities in a specific location and time (GUIZZARDI *et al.*, 2013; ALMEIDA; FALBO; GUIZZARDI, 2019). This evidence helps establish the occurrence of the event and provides a basis for further analysis and interpretation. Thus, the definitions of the terms in `PollinationExtension` can be improved to align with the binary nature of events, either occurring or not occurring. Currently, allowing a value of "0.5" to indicate a partially happening event introduces ambiguity and goes against the consideration of events as discrete occurrences. Additionally, the use of "NULL" to indicate that no attempt was made to obtain certain information is not intuitive. A better approach would be to leave the field empty or unused to indicate the absence of information.

However, it should be noted that these four terms alone may not be sufficient to fully characterize the various aspects of a plant-pollinator interactions. The limited set of predefined resources collected from flowers (i.e., pollen, nectar and oil) restricts the scope of the extension, potentially omitting other important characteristics involved in interactions. While the encoding schema (e.g., “1”, “0”, “0.5”) may provide a simple representation, it lacks explicit information without proper interpretation or decoding. In metadata design, terms which accept boolean values (i.e., “true” or “false”) are commonly referred to as “indicators”. Furthermore, it appears that there is a discrepancy between the datatype definition and the specified values for the terms in the extension. While the terms were defined to have a “probabilityType”, which typically allows for fractional values between 0 and 1, the defined values in the terms’ definition are limited to “1”, “0”, and “0.5”. Additionally, adding terms in that way could lead to excessive growth of the extension, resulting in several “indicator”-like metadata terms.

The `EnvironmentExtension` included four terms: `Temperature`, `RelativeiHumidity`, `Luminosity` and `WindSpeed`. Despite these data elements could potentially hold relevance for plant-pollinator interactions, they are more generic and applicable to various contexts beyond the scope of a specific vocabulary for plant-pollinator interactions. Considering the broader applicability of environmental data, it is advisable to separate the representation of environmental elements into a distinct vocabulary.

The limitations of the data descriptors in the `PollinationExtension` were recognized and addressed within the GEF project. As part of this initiative, an additional effort was made to expand the vocabulary used in the project by including information about organism traits and interaction outcomes and effects. Supported by FAO, a survey of potential variables for description of plant-pollinator interactions was conducted involving researchers from five continents (SARAIVA; GEMMILL-HERREN; RUGGIERO, 2010). The participants were requested to share the data fields or variables they utilized for documenting interaction data. Due to the wide range of research questions explored by the participants, an extensive list of data fields (more than 200) was compiled. This list covered various aspects related to plant-pollinator interactions, including plant and pollinator taxonomy and traits, experimental settings and protocols, environmental factors, interaction outcomes, and bibliographic references. However, many of the suggested fields seemed to be synonyms, but the lack of proper definition and conceptualization hindered the comparison of these fields and the development of the vocabulary did not evolve.

Although the GEF project did not ultimately produce a finalized vocabulary, the information collected throughout the project remains highly valuable. The involvement of experts from pollination ecology, botany, zoology, information and computing science contributed to the development of a list of potential terms that reflects the vision and the needs of a broad community, a prerequisite for data standardization (CARVALHEIRO; SARAIVA; GIANNINI, 2016).

In this context, REBIPP seized the opportunity to continue the work done by previous studies and expanded the collaboration with the international community in a joint development of a solution for sharing standardized plant-pollinator interaction data.

7.2 Community-driven development of a vocabulary of terms

As expected in any scientific field, researchers have different views and concerns about which data are relevant and must be preserved and shared (TREMBLAY *et al.*, 2017). Reaching a consensus on the terms that should compose the vocabulary is particularly challenging, especially within diverse communities. Based on the knowledge and findings from the plant-pollinator interactions vocabulary development, a guide for collaboratively developing new domain-specific vocabularies of terms was created (SALIM *et al.*, 2022). This guide aims to assist other communities in the process of creating their own specialized vocabularies.

The creation of a vocabulary of terms is a multifaceted task that encompasses both empirical and sociological components of collaborative work. While the technical aspects of constructing a vocabulary are well documented in the TDWG Vocabulary Maintenance Standard (Vocabulary Maintenance Specification Task Group, 2017b) and the TDWG Standard Documentation Standard (Vocabulary Maintenance Specification Task Group, 2017a), there is a lack of guidance on how a community should effectively organize itself and promote collaboration among its members during the development process. The proposed guidelines aim to provide assistance to communities in addressing their specific needs and requirements during the process of vocabulary development. The guidelines and the workflow, denominated *Community-driven Vocabulary Development Life Cycle*, are illustrated in the Figure 34 and fully described in Salim *et al.* (2022). The guidelines were elaborated based on challenges and successes of previous works and current work on the elaboration of the plant-pollinator interactions vocabulary (CARTOLANO, 2009; CARVALHEIRO; SARAIVA; GIANNINI, 2016; SARAIVA *et al.*, 2009; SARAIVA; GEMMILL-HERREN; RUGGIERO, 2010).

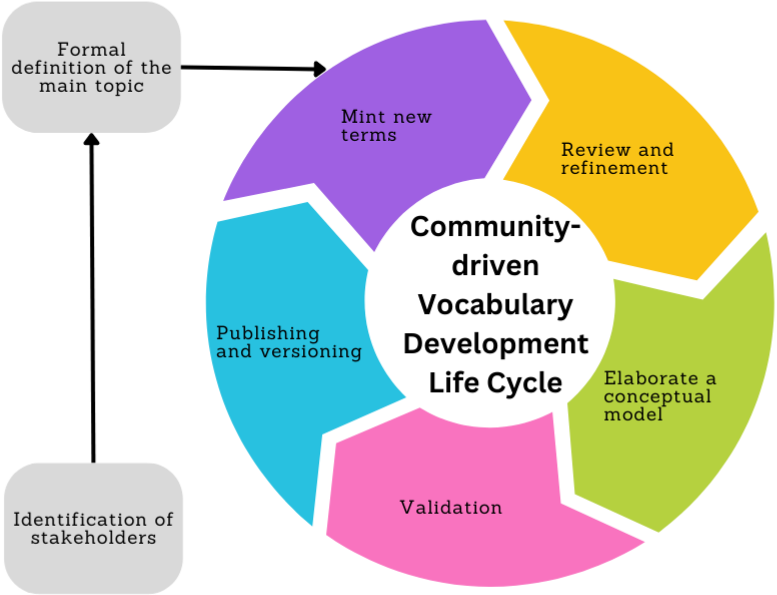

Formal definition of the main topic
Identification of stakeholders
Mint new terms
Review and refinement
Elaborate a conceptual model
Validation
Publishing and versioning
Community-driven Vocabulary Development Life Cycle

Figure 34: The *Community-driven Vocabulary Development Life Cycle* guidelines for development of new vocabularies of terms

Before beginning the development of a vocabulary, there are some initial concerns that a community interested in developing their domain-specific vocabulary should be aware of. First, the stakeholders should be carefully identified and invited to collaborate. Engagement from community members is not only vital for the development of the vocabulary, but also, more importantly, for its subsequent adoption. It is crucial that members of the community have a clear understand of the benefits of data standardization. The division of the community into different, more homogeneous groups, focusing on specific aspects of the vocabulary, facilitates the process of reaching a consensus. Next, it is essential to clearly define the main topic of the vocabulary to avoid any ambiguity regarding its scope and application. Conceptualizing the main topic involves the abstraction of processes and entities, which may sometimes lack a formal description or nomenclature within the community. Thus, while expert knowledge can facilitate the definition process, it should not be the sole source of knowledge. Literature, glossaries, nomenclature codes as well as concepts and terms borrowed from existing standards and vocabularies, should be leveraged to aid the process and included in a bibliographic reference. Once the main topic has been defined and the community has reached a consensus on the scope of the vocabulary, the development process can be initiated.

The cycle should begin with the creation of new terms. During this phase, the terms must be formally defined according to the concepts they represent. This may involve an assessment of terms from existing standards and vocabularies. The groups should collaborate to compile an inclusive set of terms that encompasses a general understanding of the main topic and scope. New terms should only be defined if they are not already defined by other existing data standards.

With a compiled list of formally defined terms, the terms should undergo multiple rounds of review and refinement. Here the members of the community play an important role as they should provide examples of use cases, elaborate controlled vocabularies (if applicable), and address conflicting definitions. The community may suggest merging or splitting terms if conflicts or ambiguities are identified. In such cases, conflicts can be resolved by defining terms that represent more general concepts. Adding new terms at this point is generally not recommended, unless it involves the splitting of existing terms. Introducing new terms at this stage could lead to unnecessary vocabulary growth and require additional rounds to achieve consensus. The review and refinement phase should be iterated until full consensus is reached. Once the review and refinement process is completed, the conceptual model should incorporate and define the rules for term usage. It is crucial to consider data models adopted by the broader community to maintain the interoperability and consistency among models and schemas. It is also advisable to develop at least one representation model for the vocabulary, such as RDF, DwC-A or XML. Once the data model is formalized, it is important to validate the solution by compiling a set of real data that covers different scientific questions. This validation process involves verifying if the vocabulary can effectively capture all or the most relevant information required for each use case, as previously considered during the conceptualization phase. Terms which present impediments to usage, such as ambiguity, missing or conflicting definitions, or definitions that are too narrow or too broad, should be identified and selected for further review and refinement. Finally, the terms that successfully pass the validation will form the next version of the vocabulary. The vocabulary should be published on suitable repositories or platforms that facilitate public access and review. For this purpose, GitHub[2] has proven to be a valuable platform for tracking and maintaining standards (CRYSTAL-ORNELAS *et al.*, 2021). In addition, alternative layouts can be considered

[2] www.github.com

for audiences who are not familiar or comfortable working on GitHub. Creating a simple and user-friendly webpage that provides a description of the vocabulary, its purpose, and allows users to easily browse for terms and definitions is an excellent way to make the vocabulary widely accessible. A good example is the DwC Quick Reference Guide[3].

The *Community-driven Vocabulary Life Cycle* described above was elaborated based on the four years of collaborative work involving members and collaborators of REBIPP. It incorporates the empirical findings that emerged from the development of the Plant-Pollinator Interactions vocabulary (presented in Section 7.3), which also serves as example for the application of the proposed guidelines.

7.3 The Plant-Pollinator Interactions vocabulary of terms

The development of the vocabulary followed the guidelines described in Section 7.2. The identification phase considered all members and collaborators of REBIPP as stakeholders. REBIPP is composed of researchers from diverse areas such as zoology, botany, ecology, agriculture, who conduct studies in Pollination Biology producing data in different contexts and with different objectives. Thus, it is crucial to establish a formal definition for the concept of "biotic interaction" to ensure that all stakeholders have a common reference point throughout the various phases of the vocabulary development life cycle.

The first topic of discussion was related to the level at which interactions and recorded and documented. Members of the community reached a consensus that the main focus is the standardization of primary data on plant-pollinator interactions. It is generally easier to summarize species-level interactions based on organism-level interactions, but, the reverse process often results in information loss. After extensive discussions, the community members agreed in favor of prioritizing organism-level interactions over species-level interactions.

The other major topics discussed regarded the differentiation between evidence and knowledge, as well as defining the concepts of "behavioral interactions" and "ecological interactions" within the context of plant-pollinator interactions. Evidences refers to the available information, data, observations, or facts that support or provide justification for a claim or hypothesis (LEHRER, 1965; ROUSH, 2005). Thus, the observation of

[3] ⟨https://dwc.tdwg.org/terms/⟩

a hummingbird visiting a flower can be considered as evidence for an interaction. The observation provides empirical data that supports the interaction. On the other hand, knowledge refers to the understanding or awareness that is acquired through the interpretation and synthesis of evidence (LEHRER, 1965; ROUSH, 2005). Qualifying that interaction as mutualistic is knowledge that requires evidence of "mutually beneficial, interspecific interactions, regardless of their specificity, intimacy or evolutionary history" (BRONSTEIN, 2001). Therefore, the interactions, when sampled and documented, serve as evidence that support a hypothesis, leading to a new state of mind (i.e., knowledge). Having reached an agreement on the distinction between evidence and knowledge in the context of plant-pollinator interaction, the community proceeded to the definitions of behavioral and ecological interactions. "Behavioral interactions" refer to all types of interactions as result of actions, responses and activities performed by one organism towards another organism (IMMELMANN, 2012). On the other hand, "ecological interactions" refer to the types of interactions that enable and support ecological functioning. Thus, the community decided to categorize plant-pollinator interactions as "behavioral interactions" due to their intimate relationship with evidence-based interactions.

The consensus was crucial because different perspectives led to divergent and conflicting concepts. The collaborative formulation of Definition 7.1 ensured a shared understanding within the community.

Definition 7.1. ***Biotic interaction:*** *a context-dependent action that a particular organism or group of organisms (considered to be taxonomically homogeneous) performs on another particular organism or group of organisms (taxonomically homogeneous) living together in a community at a particular location during some time.*

After identifying the divergences, the stakeholders were divided into three working groups (WG): the Plant WG, Animal WG and Interaction WG. While the Plant WG primarily consisted of botanists, the Animal WG and Interaction WG included specialists from other areas, such as, zoology, ecology and agriculture. This division was necessary to address the bias towards Botany among REBIPP members and to ensure that the groups had relatively equal sizes. The groups were provided with a list of potential terms from previous initiatives and were asked to review it before starting the development of the vocabulary. Reviewing the existing list of terms was useful in preventing the duplication of terms that had already been considered, and also helped the members become familiar with the terminologies and representations used in the field of Information Science.

After the members reviewed the initial list of potential terms, a survey was conducted to identify key data elements that required attention during the development of the vocabulary. Among those who answered the survey (n=29), 90% considered the initial list of terms to be sufficient for describing plant-pollinator interactions, and all respondents agreed that the terms encompassed the data usually collected during research. When asked about the most suitable data elements for their research needs, the respondents indicated that collected resources (100%), interaction type (96.6%), animal behavior (89.7%), and place of contact between the animal and the plant (86.2%) are some of the most relevant data elements in their research, among others (Figure 35). Specifically regarding animals, the respondents indicated data elements such as biological elements (82.8%), behavior (79.3%) and activity season (79.3%) are among the most important data elements (Figure 36). For plants, they indicated that flower longevity (86.2%), bloom intensity (86.2%) and flower type (82.8%) are considered some of the most important data elements. Figure 37 presents a list of the top 20 most cited data elements for plants, as indicated by respondents. While taxonomic elements were among the most frequently cited data elements, for both animals and plants, some respondents may have considered this information to be implicit or fundamental and therefore did not explicit mention it. Additionally, in certain studies, researchers may choose to use functional groups or guilds instead of specific taxonomic information for their analysis.

Indeed, based on the respondents' feedback, it can be summarized that sampling details play a crucial role in the studying of plant-pollinator interactions (and certainly for biotic interactions too). Additionally, considering the life history and traits of the organisms involved in the interactions is essential for understanding the dynamics and implications of these interactions. Fortunately, the DwC standard already has terms for documenting sampling details in the `dwc:Event` class, and the `dwc:MeasurementOrFact` and its extensions can be used to document organisms' traits and interaction effects and outcomes, as already described in Section 6.2.

To facilitate understanding, formalize the definitions of terms, and aid in the organization of the vocabulary, a common template (Table 15) was used for defining the terms.

Periodically, workshops were organized by REBIPP with funding support from the project "Safeguarding Pollination Services in a Changing World: theory into practice (SURPASS2)"[4]. These workshops served to foster collaboration and facilitate the exchange of knowledge among members of each group. During the first workshop, which took place between April 10 and 11, 2017, at the University of São Paulo (USP), the participating researchers focused on defining the working groups (WG) and discussing

[4] ⟨https://bv.fapesp.br/pt/auxilios/104850/⟩

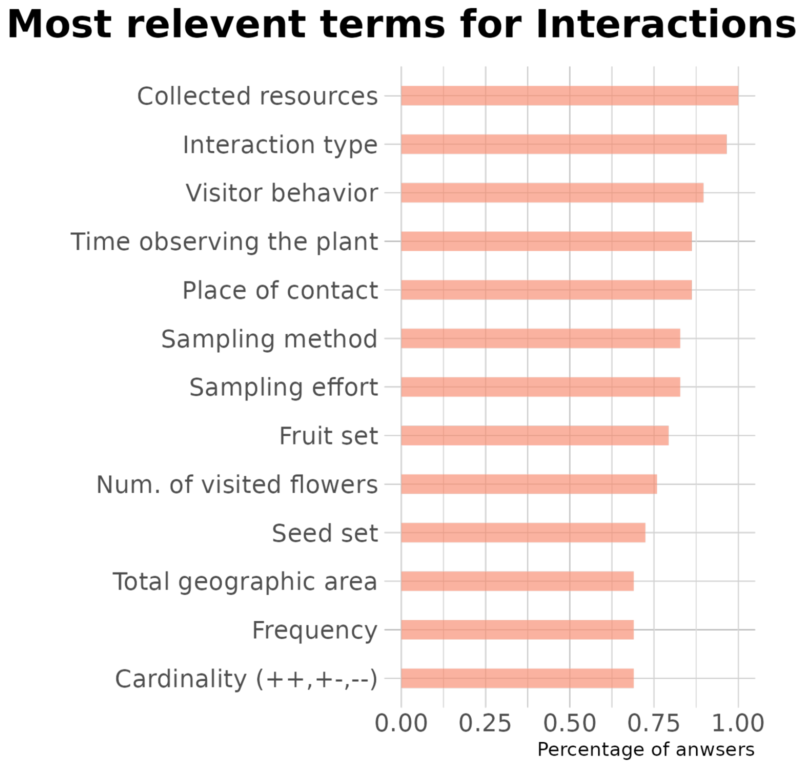


Figure 35: Most relevant interaction terms listed by survey respondents

the results of the survey that had been conducted prior to the workshop. The WGs also engaged into discussions about representing biological information using the DwC standard, including its limitations and the need for additional terms to extend the standard. The subsequent workshops, held in August 2017, May 2018 and January 2019, were important for accelerating discussions and fostering consensus among critic topics. These in-person meetings provided a valuable opportunity for participants to gain insights into the progress and decisions made by the WGs, in-depth conversations, knowledge sharing, and the exchange of different perspectives. The meetings aimed to promote extensive discussions and ultimately achieve a consensus on the terms that would compose the vocabulary. By bringing together experts and researchers in the field of Pollination Biology and Biodiversity Informatics, the workshops resulted in advancements in understanding and resolving critical topics related to the representation of plant-pollinator interactions data using the DwC standard.

Each group worked on the creation of new terms, which they deemed to be important and missing from the initial list of terms. The result was a long list of 278 terms, some of which highly specific to certain protocols or rarely collected and measured. The principle

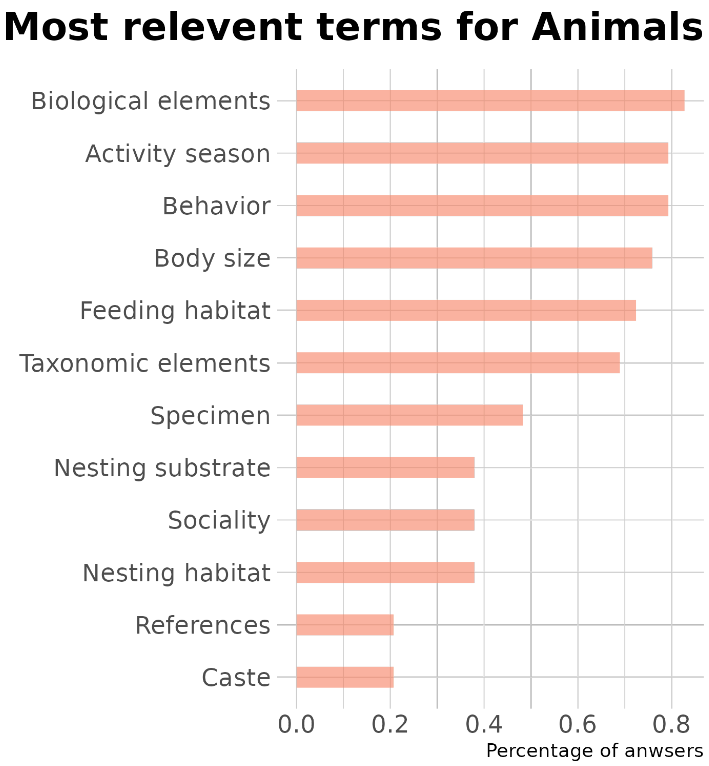


Figure 36: Most relevant animal terms listed by survey respondents

of simplicity in metadata development is based on the premise that a vocabulary with an excessive number of terms can be challenging to use, and complexity can create barriers to its adoption (CHAN; ZENG, 2006; DUVAL *et al.*, 2002; POMERANTZ, 2015). Furthermore, the high level of abstraction of certain concepts, such as mutualism and parasitism, requires subjective causal inference that is beyond what primary data alone can represent. Consequently, during the review and refinement phase, the groups were encouraged to reduce the number of terms by merging similar terms and eliminating those that are rarely recorded or inferred from primary data. During the review and refinement phase, the GitHub Issues Tracking system[5] was used for discussions, ensuring transparency and open access throughout the vocabulary development. After the review and refinement, the size of the vocabulary was significantly reduced to include only the necessary terms for the representation of plant-pollinator interactions and the elaboration of the conceptual data model.

From the beginning, it was understood that the vocabulary of terms for Plant-Pollinator Interactions data should be used in conjunction with the DwC standard, more

[5] ⟨https://github.com/rebipp/ppi⟩

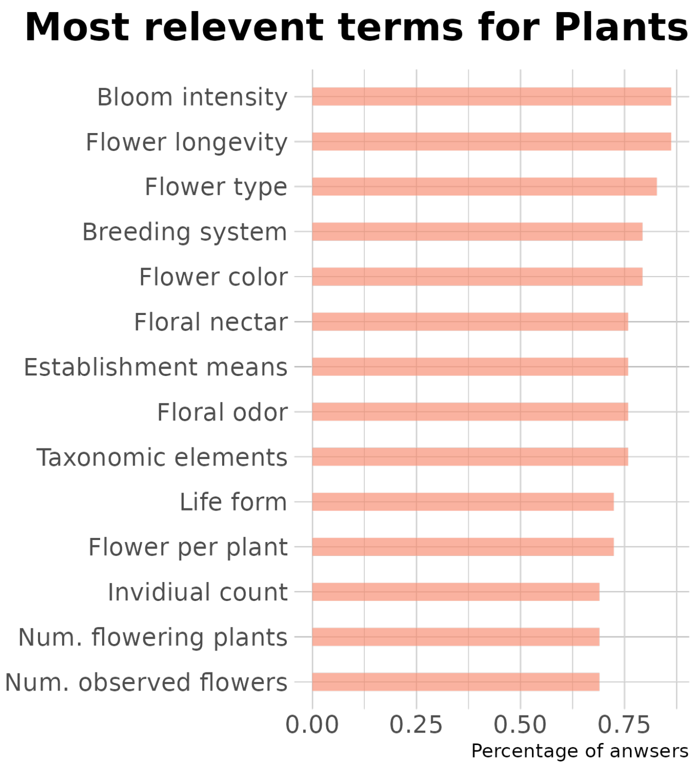


Figure 37: Most relevant plant terms listed by survey respondents

specifically with the `dwc:MeasurementOrFact` class and its extensions. Therefore, the data schema for biotic interactions was adopted. During this phase, the groups were asked to classify the terms as properties of the interactions (`dwc:Event`) or of the organisms documented in the occurrences (`dwc:Occurrence`). Terms that represent an interaction outcome resulting from the co-occurrence and encounter of two organisms were classified as measurements or facts of the `dwc:Event` representing the interaction. On the other hand, the remaining terms that represent any trait or effect on one of the interacting organisms were classified as measurements or facts of the `dwc:Occurrence` representing such organisms.

A draft version of the vocabulary was completed after multiple rounds of the development process. Once the vocabulary reached a certain level of maturity, where no new terms were being added, the groups were dissolved and all members of the community came together to collaboratively review and refine of all terms. It was particularly important to revisit and refresh discussions that were relatively consolidated within groups, but had not been fully explored by other groups.

After nearly three years of collaborative and voluntary effort, the first official version of

Term Label: Flower Opening Type	
Identifier	http://rs.rebipp.org.br/ppi/terms/flowerOpeningType
Class	Flower
Definition	The type of flower describing whether the flower's corolla opens or not, exposing its reproductive parts
Comments	Recommended best practice is to use a controlled vocabulary
Details	Proctor, M. P. et al. 1996. The natural history of pollination. HarperCollins. Inouye DW, Favre DW, Lanum JA, Levine DM, Meyers JB, Roberts MS, Tsao FC, Wang Y-Y. 1980. The effects of nonsugar nectar constituents on estimates of nectar energy content. Ecology 61: 992–996
Protocol	Observation of the floral development from the bud stage to senescence (Dafni et al. 2005)
Controlled vocabulary	cleistogamous; chasmogamous; both
Examples	cleistogamous; chasmogamous; both

Table 15: Template used to define the terms in the plant–pollinator interactions vocabulary

Term label: a human readable name; *identifier*: a unique IRI; *class*: the category in which the term is defined; *definition*: the term definition in human readable form; *comments*: additional comments to the term definition and its use; *details*: list of references to the concept represented by the term; *protocol*: recommended protocols to measure or to record the value for the term (if applicable); *controlled vocabulary*: list of recommended values, such as terms from existing thesauri and ontologies (if applicable).

the Plant-Pollinator Interactions (PPI) vocabulary was released. This version includes 48 terms specifically defined for documenting plant-pollinator interactions. The vocabulary is easily accessible in human-readable format through the open access and stable repository in GitHub [6] and on the REBIPP website [7]. Furthermore, a Controlled Vocabularies (CV) which includes new definitions and imported terms from other existing CV is available[8]. The CV can be used in conjunction with the terms defined in the PPI vocabulary. To ensure compatibility and interoperability, machine-readable formats based on content negotiation for HyperText Transfer Protocol (HTTP) are also available.

[6] ⟨https://github.com/rebipp/ppi⟩

[7] ⟨https://ppi.rebipp.org.br/terms/⟩

[8] ⟨https://ppi.rebipp.org.br/cv/⟩

8 IMPLEMENTING THE PLANT-POLLINATOR INTERACTIONS VOCABULARY

This chapter describes the implementation of the Plant-Pollinator Interactions vocabulary within the Brazilian Network on Plant-Pollinator Interactions. It highlights the solutions that were developed to facilitate the digitization and standardization of plant-pollinator interactions data.

8.1 A template spreadsheet for data digitization

The data model presented in Section 6.2 is not intended for use by data authors. Instead, it is designed to be used by software developers and biodiversity informaticians to generate and exchange standardized data in machine-readable formats. While it is important for data authors to be aware of available standards and vocabularies and know how to use them when documenting data, it is also necessary to provide some level of abstraction to simplify the process of data preparation and annotation. In the GBIF network, this is done by the Integrated Publishing Toolkit (IPT). The IPT provides data authors with a user-friendly interface where they can prepare and annotate their datasets without relying on the complexities of producing machine-readable formats such as DwC-A.

However, the IPT is primarily designed for handling flat data organized in two-dimensional structure. When it comes to working with relational data, the IPT may not provide the same level of convenience. It involves the attribution of unique identifiers for each record in the dataset and using these identifiers to create links between records in different tables, typically through the use of *foreign keys*. Thus, even with small datasets, the process of data preparation can become complex and time-consuming. The Tables 16 and 17 show an example of relational data that has been annotated with terms from the DwC standard. In Table 16, the unique identifiers (`dwc:occurrenceID`) need to be created by the data authors. These identifiers are then used in Table 17 (`dwc:occurrenceID`) to establish the linkage between the measurements and the corresponding occurrences. Preparing relational data, whether in the IPT or any other system, can be challenging for individuals who are only familiar with working with flat data such as spreadsheets.

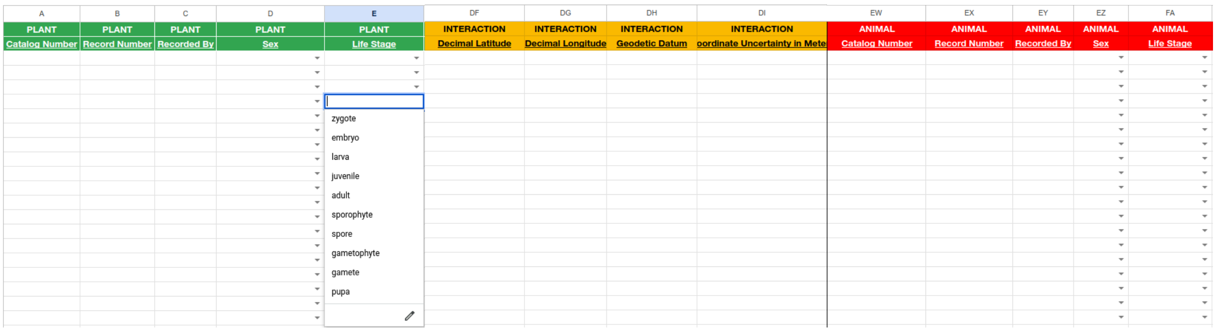
occurrenceID	**scientificName**	**recordedBy**
occ:1001	*Curatella americana*	JAS
occ:1002	*Acacia mangium*	JAS

Table 16: Example of a table with occurrence data (`dwc:Occurrence`)

occurrenceID	**measurementID**	**measurementType**	**measurementValue**
occ:1001	mof:0001	floralSymmetry	actinomorphic
occ:1001	mof:0002	plantHabit	tree
occ:1002	mof:0003	flowerShape	chamber-shaped

Table 17: Example of a table with measurements or facts data (`dwc:MeasurementOrFact` or its extensions).

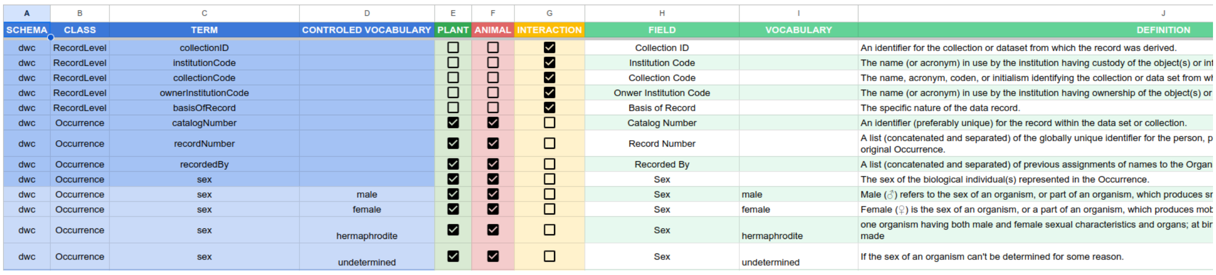
In order to simplify the data digitization and standardization process, the approach adopted in REBIPP was the utilization of a template spreadsheet (Figure 38). This spreadsheet allows for the mapping of columns to corresponding terms from DwC and PPI vocabulary. It has columns dedicated to documenting the characteristics of plants, animals and interactions, using a color scheme to differentiate each group. For terms that have defined a controlled vocabulary, the spreadsheet includes a list of predefined values for data authors to choose from. The simplest form of using DwC is the creation of a flat-file, where columns are mapped to the corresponding terms in the DwC standard (the "Simple Darwin Core"). In the "Simple Darwin Core" format, it is not allowed to have duplicate terms, meaning that the spreadsheet cannot have two columns mapped to the same DwC term. However, in the template spreadsheet, the DwC terms are duplicated for both plants and animals, as they each have their own set of characteristics to be documented. Thus, the template spreadsheet is not a "Simple Darwin Core", but a flat-file with a known format which can be transformed into a machine-readable representation (e.g. DwC-A, RDF). This solution has been implemented in the plant-pollinator interactions information system, which will be described in the next section. The template spreadsheet also includes other worksheets for metadata documentation, using terms imported from the Ecological Metadata Language (EML), for consulting the glossary of terms (Figure 39 and the controlled vocabularies (Figure 40).



A	B	C	D	E	DF	DG	DH	DI	EW	EX	EY	EZ	FA
PLANT	PLANT	PLANT	PLANT	PLANT	INTERACTION	INTERACTION	INTERACTION	INTERACTION	ANIMAL	ANIMAL	ANIMAL	ANIMAL	ANIMAL
Catalog Number	Record Number	Recorded By	Sex	Life Stage	Decimal Latitude	Decimal Longitude	Geodetic Datum	oordinate Uncertainty in Mete	Catalog Number	Record Number	Recorded By	Sex	Life Stage

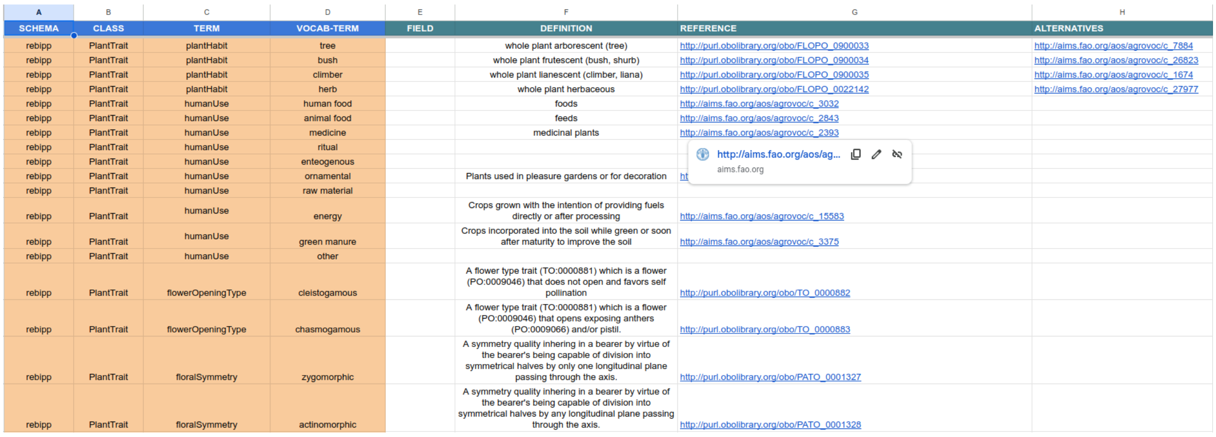
Figure 38: The template spreadsheet used in the Brazilian Network on Plant-Pollinator Interactions (REBIPP) community to facilitate data sharing

This solution proved to be efficient in the context of REBIPP as many datasets could be standardized or are in process of standardization by collaborators of the network. Those who wish to contribute data can simply make a copy of the template spreadsheet and fill it with their data, without needing to worry about relationships and foreign keys.



SCHEMA	CLASS	TERM	CONTROLED VOCABULARY	PLANT	ANIMAL	INTERACTION	FIELD	VOCABULARY	DEFINITION
dwc	RecordLevel	collectionID		☐	☐	☑	Collection ID		An identifier for the collection or dataset from which the record was derived.
dwc	RecordLevel	institutionCode		☐	☐	☑	Institution Code		The name (or acronym) in use by the institution having custody of the object(s) or inf
dwc	RecordLevel	collectionCode		☐	☐	☑	Collection Code		The name, acronym, coden, or initialism identifying the collection or data set from wh
dwc	RecordLevel	ownerInstitutionCode		☐	☐	☑	Onwer Institution Code		The name (or acronym) in use by the institution having ownership of the object(s) or
dwc	RecordLevel	basisOfRecord		☐	☐	☑	Basis of Record		The specific nature of the data record.
dwc	Occurrence	catalogNumber		☑	☑	☐	Catalog Number		An identifier (preferably unique) for the record within the data set or collection.
dwc	Occurrence	recordNumber		☑	☑	☐	Record Number		A list (concatenated and separated) of the globally unique identifier for the person, p original Occurrence.
dwc	Occurrence	recordedBy		☑	☑	☐	Recorded By		A list (concatenated and separated) of previous assignments of names to the Organ
dwc	Occurrence	sex		☑	☑	☐	Sex		The sex of the biological individual(s) represented in the Occurrence.
dwc	Occurrence	sex	male	☑	☑	☐	Sex	male	Male (♂) refers to the sex of an organism, or part of an organism, which produces sr
dwc	Occurrence	sex	female	☑	☑	☐	Sex	female	Female (♀) is the sex of an organism, or a part of an organism, which produces mob
dwc	Occurrence	sex	hermaphrodite	☑	☑	☐	Sex	hermaphrodite	one organism having both male and female sexual characteristics and organs; at bir made
dwc	Occurrence	sex	undetermined	☑	☑	☐	Sex	undetermined	If the sex of an organism can't be determined for some reason.

Figure 39: Glossary of terms in the template spreadsheet used in the Brazilian Network on Plant-Pollinator Interactions (REBIPP) for biotic interactions data sharing.



SCHEMA	CLASS	TERM	VOCAB-TERM	FIELD	DEFINITION	REFERENCE	ALTERNATIVES
rebipp	PlantTrait	plantHabit	tree		whole plant arborescent (tree)	http://purl.obolibrary.org/obo/FLOPO_0900033	http://aims.fao.org/aos/agrovoc/c_7884
rebipp	PlantTrait	plantHabit	bush		whole plant frutescent (bush, shurb)	http://purl.obolibrary.org/obo/FLOPO_0900034	http://aims.fao.org/aos/agrovoc/c_26823
rebipp	PlantTrait	plantHabit	climber		whole plant lianescent (climber, liana)	http://purl.obolibrary.org/obo/FLOPO_0900035	http://aims.fao.org/aos/agrovoc/c_1674
rebipp	PlantTrait	plantHabit	herb		whole plant herbaceous	http://purl.obolibrary.org/obo/FLOPO_0022142	http://aims.fao.org/aos/agrovoc/c_27977
rebipp	PlantTrait	humanUse	human food		foods	http://aims.fao.org/aos/agrovoc/c_3032	
rebipp	PlantTrait	humanUse	animal food		feeds	http://aims.fao.org/aos/agrovoc/c_2843	
rebipp	PlantTrait	humanUse	medicine		medicinal plants	http://aims.fao.org/aos/agrovoc/c_2393	
rebipp	PlantTrait	humanUse	ritual				
rebipp	PlantTrait	humanUse	enteogenous				
rebipp	PlantTrait	humanUse	ornamental		Plants used in pleasure gardens or for decoration	ht	
rebipp	PlantTrait	humanUse	raw material				
rebipp	PlantTrait	humanUse	energy		Crops grown with the intention of providing fuels directly or after processing	http://aims.fao.org/aos/agrovoc/c_15583	
rebipp	PlantTrait	humanUse	green manure		Crops incorporated into the soil while green or soon after maturity to improve the soil	http://aims.fao.org/aos/agrovoc/c_3375	
rebipp	PlantTrait	humanUse	other				
rebipp	PlantTrait	flowerOpeningType	cleistogamous		A flower type trait (TO:0000881) which is a flower (PO:0009046) that does not open and favors self pollination	http://purl.obolibrary.org/obo/TO_0000882	
rebipp	PlantTrait	flowerOpeningType	chasmogamous		A flower type trait (TO:0000881) which is a flower (PO:0009046) that opens exposing anthers (PO:0009066) and/or pistil.	http://purl.obolibrary.org/obo/TO_0000883	
rebipp	PlantTrait	floralSymmetry	zygomorphic		A symmetry quality inhering in a bearer by virtue of the bearer's being capable of division into symmetrical halves by only one longitudinal plane passing through the axis.	http://purl.obolibrary.org/obo/PATO_0001327	
rebipp	PlantTrait	floralSymmetry	actinomorphic		A symmetry quality inhering in a bearer by virtue of the bearer's being capable of division into symmetrical halves by any longitudinal plane passing through the axis.	http://purl.obolibrary.org/obo/PATO_0001328	

Figure 40: Controlled vocabularies terms in the template spreadsheet used in the Brazilian Network on Plant-Pollinator Interactions (REBIPP) for biotic interactions data sharing.

8.2 The REBIPP plant-pollinator interactions database

From the beginning of REBIPP, it was understood that a database of plant-pollinator interactions is crucial to achieving the network's objectives. Since then, the development of an Information System (IS) to provide access to plant-pollinator interactions data was initiated. One of the requirements of the IS was that it must provide access to standardized and aggregated data. The IS is still in development, but a production version is available and accepting contributions[1].

For the reasons explained in Section 8.1, the IPT was not a feasible solution for data authors when they have to work with relational data. For that reason, a solution using the template spreadsheet was developed. The IS is capable of importing data from the template spreadsheets and properly populating the database with standardized plant-pollinator interactions data. The data are stored in a instance of MongoDB community server, a JSON-like document-oriented NoSQL database, which is a subclass of the key-value store. Key-value stores are a collection of objects, or records, with different fields (i.e., keys) within them, each containing data (i.e., values) (WIKIPEDIA, 2023). It means that each row in the template spreadsheet is a document, composed by key-value pairs, in the database. Similar to DwC, the keys cannot repeat within the same document, and thus, the duplicated keys (those from DwC) are prefixed with `plant` and `animal` depending on if it is a plant or animal occurrence. The terms related to the interactions are prefixed with `interaction` to maintain consistence with the format of other keys.

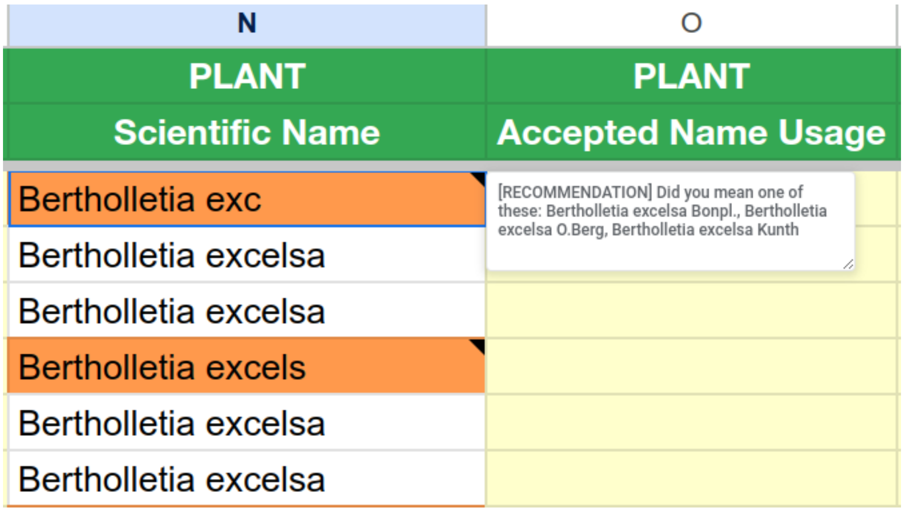
The PPI vocabulary is used to document interaction outcomes and species traits stored in the database, but the data model described in Section 6.2 is used only when exporting data from the database. Although, it is possible to access the original data or exporting results from a query as CSV files in the same format as the template spreadsheet. Since the IS can produce DwC-A from stored data, it can easily be integrated with other systems to exchange standardized data, including GBIF. Despite GBIF is currently unable to interpret the interactions in the DwC-A, it can still index the occurrences and events in the datasets, providing some level of access to the data.

Before importing data into the database, the spreadsheet passes through different validation rules. First, the IS checks if the spreadsheet is in the correct format (i.e., template spreadsheet), then, it checks the taxon names against GBIF Species API[2]. If there is any error in the spreadsheet or any invalid or incorrect spelled taxon name is

[1] How to Contribute: ⟨http://db.rebipp.org.br/how-to-contribute⟩

[2] ⟨https://www.gbif.org/developer/species⟩

found, the problematic spreadsheet cells are annotated with error messages (Figure 41). The data quality assessment and management is based on the framework proposed by Veiga *et al.* (2017), which incorporates data fitness for use principles. The complete workflow of data input and integration is shown in Figure 42.



N	O
PLANT Scientific Name	PLANT Accepted Name Usage
Bertholletia exc	
Bertholletia excelsa	
Bertholletia excelsa	
Bertholletia excels	
Bertholletia excelsa	
Bertholletia excelsa	

[RECOMMENDATION] Did you mean one of these: Bertholletia excelsa Bonpl., Bertholletia excelsa O.Berg, Bertholletia excelsa Kunth

Figure 41: Error reporting example in the template spreadsheet used in Brazilian Network on Plant-Pollinator Interactions (REBIPP) for biotic interaction data sharing

Furthermore, the current version of the IS provides a user-friendly Web interface that enables users to query the database based on plant and/or animal scientific names and/or interaction types. The query results can be visualized as tabular data, interactive maps of interactions and Sankey diagrams. Additionally, users have the option to download the query results in formats such as CSV, JSON or DwC-A for further analysis. In the upcoming version, the IS will include advanced filtering options based on the content of other DwC terms and terms from the PPI vocabulary.

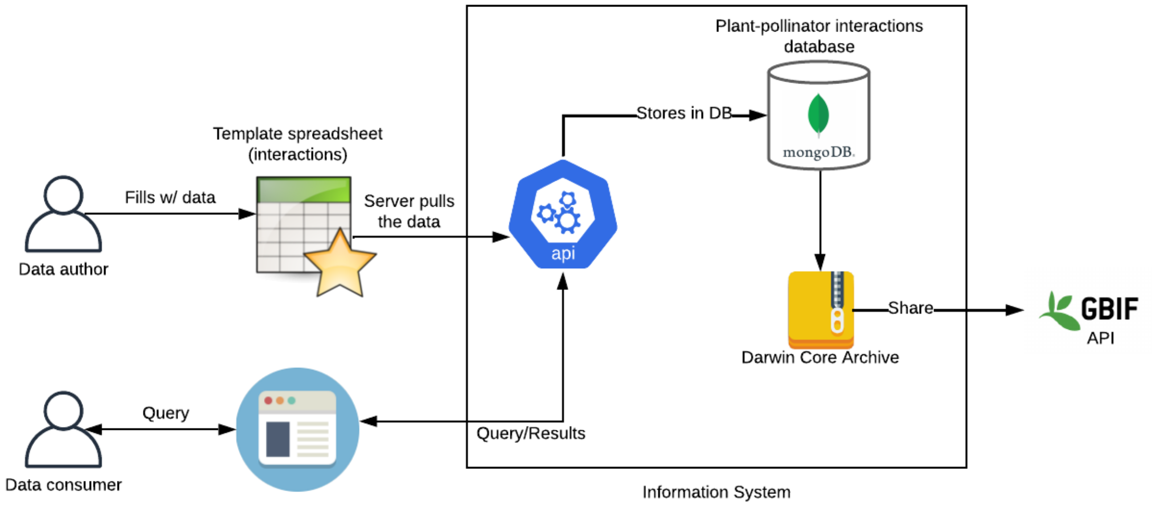


Figure 42: Diagram of the Brazilian Network on Plant-Pollinator Interactions (REBIPP) information system of plant-pollinator interactions

9 CONCLUSION

This chapter provides an overview of the main contributions and ideas for future work.

9.1 Overview of main contributions

The review of concepts and terminology used by the research community for designating biotic interactions revealed a mixture of inconsistent and unresolved aspects. Representations of biotic interactions, such as the “interaction grid”, while still valuable for various studies, have contributed to confusion between the observational unit and the biological unit of interest (LAZIC; CLARKE-WILLIAMS; MUNAFÒ, 2018) when defining biotic interactions. Decades of sampling and recording biotic interactions have contributed to the knowledge embedded in these representations. However, it is important to note that these representations do not apply to the classification of observed interactions, as they go beyond what primary data can represent. Thus, these representations have contributed little to the development of a definition of biotic interactions that aligns with what primary data can document. On the other hand, more recent approaches that consider biotic interactions as dynamic processes provide a contextualized and more realistic representation of these interactions. These approaches do not seek to classify biotic interactions into distinct categories based on their effects and outcomes in populations and communities. Instead, they represent biotic interactions as processes with immediate and long-term effects and outcomes. This representation of biotic interactions is closer to what is sampled in the field and recorded by primary data. It acknowledges the complexity and variability of interactions in natural systems, capturing the dynamic nature of these ecological processes without incorporating knowledge in the definition of biotic interactions. The understanding of interactions as discrete events, with arbitrary durations, rather than static outcomes of net effects, allows for the documentation of primary data without confusing evidence and knowledge, observational units, and the biological units of interest. This approach recognizes the importance of capturing the specific interactions as they occur, providing a clearer and more accurate representation of the underlying processes. However, in order to support knowledge construction for the elucidation or classification of biotic interactions, it is necessary to extend the scope of primary data beyond simple “tetranomials”.

While capturing the basic components of an interaction is important, incorporating additional data, such as environmental factors, organisms’ traits, and interaction outcomes and effects, is needed to provide a more holistic view of the interactions. When aggregated and summarized, these expanded data provide the foundation for inferring the impacts of biotic interactions on populations, communities, and ecosystems. Given the plasticity and dynamics of these interactions, it is crucial that primary data on biotic interactions incorporates these additional elements and, ideally, adheres to standardized formats. This enhanced data collection and standardization will facilitate more robust comparisons, analyses, and interpretations across different studies and ecological contexts. Standardization simplifies the process of aggregating and integrating data from different sources. This simplification, in turn, facilitates data reuse and interoperability, allowing for the seamless exchange and integration of data across studies and research projects. Standardization also enables the continuous update and refinement of knowledge on the processes involved in biotic interactions. Ultimately, the goal of standardization is to enhance collaboration, improve data quality, and promote the advancement of scientific understanding in the field of biotic interactions. Standardization can also be very helpful in the early stages of the research process, as it aids in identifying potential variables of interest to be collected and provides guidance on collecting and digitization methods.

While it may be possible to define and differentiate *interaction terms* based on their intrinsic meanings, such as biotic, ecological, or interspecific, it is important to note that the terminology commonly used in the literature may not align with these distinctions. In scientific discourse, the *interaction terms* and the concepts they represent are used interchangeably or with overlapping meanings, leading to potential confusion or ambiguity.

It is therefore essential to consider the existing terminology and usage within the field when developing a standardized vocabulary or classification system for biotic interactions. By aligning with the established terminology and terminology commonly used in the literature, it becomes easier to communicate and collaborate effectively within the scientific community and ensure consistency in the interpretation and analysis of biotic interaction data. This is supported by the frequent use of multiple *interaction terms* within the same publication, indicating a lack of consensus and consistency in the terminology used for biotic interactions. Ensuring clarity and minimizing ambiguities between concepts is crucial from a data perspective. By doing so, data can capture information more efficiently and effectively, enabling the development of more robust data models.

Additionally, the review of biotic interaction datasets has revealed that although the number of available datasets has been increasing, there are still limitations in terms of interoperability and data reuse. The investigation of datasets from four data repositories showed a diversity of formats, licenses, and metadata details across the datasets. The examination of the keywords present in the datasets revealed a wide range of terms used to refer to interactions. Due to the lack of standardization and specific metadata annotation of these datasets, they cannot be easily indexed and aggregated without manual transformations. The importance of sharing data aligned with open data principles is hindered by the heterogeneity of data formats and structures, which limits data reuse and interoperability. Most datasets have been made available primarily to meet the requirements of scientific journals for data publication, with limited attention given to their potential for reuse and interoperability. In this scenario, initiatives such as GloBI play a crucial role in enhancing and promoting access to open data that would otherwise be inaccessible or underutilized due to the lack of appropriate metadata. It also highlights the necessity of widely adopted community data models and standards for sharing biotic interactions.

GBIF has proven to be a valuable source of biotic interactions data with the potential to support studies at different geographical and taxonomic scales. The biases found in biotic interactions data are not different from what is commonly observed in occurrence data and biodiversity data in general (BECK *et al.*, 2014; BOAKES *et al.*, 2010; GEURTS; REYNOLDS; STARZOMSKI, 2023; ROCHA-ORTEGA; RODRIGUEZ; CÓRDOBA-AGUILAR, 2021; RUETE, 2015; TROIA; McManamay, 2016; TROUDET *et al.*, 2017). However, the lack of common data schema and standardized vocabularies for documenting and annotating biotic interactions data presents a challenge for the indexing of these data by GBIF, resulting in a significant amount of "hidden" data within the GBIF registry.

These findings strongly contributed to the development of the data schema presented in Chapter 6 and the vocabulary of terms in Chapter 7. By incorporating the most up-to-date terminology and concepts related to biotic interactions, the proposed data schema can capture a more realistic and contextualized representation of these interactions. The data schema is designed to be flexible enough to capture both immediate and long-term effects of interactions by allowing for the creation of a cascade of `dwc:Event` instances. It also accommodates context-dependent data such as organisms' traits, enhancing the representation of biotic interactions. Multiple `dwc:Event` instances can be linked using the terms `dwc:eventID` and `dwc:parentEventID`, enabling the representation of a chain of events that capture both immediate and long-term effects and outcomes of interactions. The exclusive use of the DwC standard in the data schema simplifies its implementation by data providers, enabling them to efficiently share standardized biotic interactions data. Data providers can easily structure and format their data according to the established guidelines, ensuring consistency and compatibility with other datasets. In conjunction with specialized vocabularies, such as the Plant-Pollinator Interactions vocabulary, the data schema can be improved to include standardized traits data and interactions effects and outcomes.

The *Community-driven Vocabulary Development Life Cycle* (SALIM *et al.*, 2022) was found to be a beneficial and efficient approach for managing and collaborating with the research community during the development of the PPI vocabulary of terms. It was particularly valuable in the context of a heterogeneous community with diverse expertise, including biologists and information scientists, as well as different perspectives, such as phytocentric and zoocentric views of the interactions. The development of standards and vocabularies is a complex process that requires attention to both practical and theoretical aspects of the subject. Equally important is the effective management of community members, ensuring their active participation, contribution, and satisfaction in expressing their needs and concerns. The outcome was the creation of a pioneering vocabulary specifically tailored for the standardization of plant-pollinator interactions data. This vocabulary represents a significant step towards facilitating data exchange and interoperability in the field of Pollination Biology. The successful development of the PPI vocabulary has been driven by the engagement and input of the community, ensuring that it incorporates the necessary data elements to effectively document plant-pollinator interactions in a standardized manner. The independence of the PPI vocabulary from the DwC standard, allows for its integration with other biodiversity data standards (e.g., Access to Biological Collection Data, Plinian Core), enabling its use in different contexts.

9.2 Future work

This study has focused on the development of a vocabulary of terms specifically for plant-pollinator interactions. However, an important direction for future research is the expansion of vocabularies to cover other crucial domains for ecosystem functioning, such as plant-herbivore (BURKEPILE; PARKER, 2017), plant-seed-dispersal (JORDANO *et al.*, 2010) and plant-fungi (DIGHTON, 2018) interactions. Each domain encompasses a unique set of concepts that require careful formalization to enable the definition of terms for capturing the necessary data elements. The development of vocabularies specific to each domain is essential for representing contextualized and more realistic interactions. By creating domain-specific vocabularies, data can be effectively documented to express the complexities of different ecological processes, leading to a more comprehensive understanding of ecosystem dynamics and functioning.

Another area of potential future exploration is the formalization of the knowledge related to biotic interactions into an ontology. While the development of a vocabulary of terms is a crucial step in standardizing and capturing data elements, an ontology goes beyond this by providing a formal and structured representation of the relationships, properties, and concepts within a domain. While the OBO Relations Ontology (RO) provides valuable terms for representing interaction types, there are still other important concepts and relationships within biotic interactions that have not been addressed. An ontology for biotic interactions would provide a semantic framework that allows for more sophisticated knowledge representation, reasoning, and integration across diverse datasets and research domains.

The evolution of the PPI vocabulary is an ongoing process that requires continuous participation and engagement from the community interested in sharing standardized data on plant-pollinator interactions. As new research findings emerge and new data requirements arise, the vocabulary will need to be updated and expanded to accommodate these changes. This can involve adding new terms, refining existing definitions, and incorporating feedback from the community. While the responsibility of vocabulary maintenance lies on REBIPP, it is important to recognize that this task may require further research to investigate and review the underlying terminology and definitions.

REFERENCES[1]

ABRAMS, P. A. On Classifying Interactions between Populations. **Oecologia**, v. 73, n. 2, p. 272–281, 1987. ISSN 0029-8549.

AGGARWAL, C. C. Mining text data. In: AGGARWAL, C. C. (Ed.). **Data Mining: The Textbook**. Cham: Springer International Publishing, 2015. p. 429–455. ISBN 978-3-319-14142-8. Disponível em: ⟨doi:10.1007/978-3-319-14142-8_13⟩.

AGUIAR, M. A. M.; NEWMAN, E. A.; PIRES, M. M.; YEAKEL, J. D.; BOETTIGER, C.; BURKLE, L. A.; GRAVEL, D.; GUIMARÃES JR, P. R.; O'DONNELL, J. L.; POISOT, T.; FORTIN, M.-J.; HEMBRY, D. H. Revealing biases in the sampling of ecological interaction networks. **PeerJ**, v. 7, p. e7566, 2019. ISSN 2167-8359. Disponível em: ⟨doi:doi:10.7717/peerj.7566⟩.

ALBRECHT, J.; CLASSEN, A.; VOLLSTÄDT, M. G. R.; MAYR, A.; MOLLEL, N. P.; COSTA, D. S.; DULLE, H. I.; FISCHER, M.; HEMP, A.; HOWELL, K. M.; KLEYER, M.; NAUSS, T.; PETERS, M. K.; TSCHAPKA, M.; STEFFAN-DEWENTER, I.; BÖHNING-GAESE, K.; SCHLEUNING, M. Plant and animal functional diversity drive mutualistic network assembly across an elevational gradient. **Nature Communications**, v. 9, n. 1, p. 3177, 2018. ISSN 2041-1723. Disponível em: ⟨doi:10.1038/s41467-018-05610-w⟩.

ALIGULIYEV, R. M. Performance evaluation of density-based clustering methods. **Information Sciences**, v. 179, n. 20, p. 3583–3602, 2009. ISSN 0020-0255. Disponível em: ⟨doi:10.1016/j.ins.2009.06.012⟩.

ALLEN-PERKINS, A.; MAGRACH, A.; DAINESE, M.; GARIBALDI, L. A.; KLEIJN, D.; RADER, R.; REILLY, J. R.; WINFREE, R.; LUNDIN, O.; MCGRADY, C. M.; BRITTAIN, C.; BIDDINGER, D. J.; ARTZ, D. R.; ELLE, E.; HOFFMAN, G.; ELLIS, J. D.; DANIELS, J.; GIBBS, J.; CAMPBELL, J. W.; BROKAW, J.; WILSON, J. K.; MASON, K.; WARD, K. L.; GUNDERSEN, K. B.; BOBIWASH, K.; GUT, L.; ROWE, L. M.; BOYLE, N. K.; WILLIAMS, N. M.; JOSHI, N. K.; ROTHWELL, N.; GILLESPIE, R. L.; ISAACS, R.; FLEISCHER, S. J.; PETERSON, S. S.; RAO, S.; PITTS-SINGER, T. L.; FIJEN, T.; BOREUX, V.; RUNDLÖF, M.; VIANA, B. F.; KLEIN, A.-M.; SMITH, H. G.; BOMMARCO, R.; CARVALHEIRO, L. G.; RICKETTS, T. H.; GHAZOUL, J.; KRISHNAN, S.; BENJAMIN, F. E.; LOUREIRO, J.; CASTRO, S.; RAINE, N. E.; DE GROOT, G. A.; HORGAN, F. G.; HIPÓLITO, J.; SMAGGHE, G.; MEEUS, I.; EERAERTS, M.; POTTS, S. G.; KREMEN, C.; GARCÍA, D.; MIÑARRO, M.; CROWDER, D. W.; PISANTY, G.; MANDELIK, Y.; VEREECKEN, N. J.; LECLERCQ, N.; WEEKERS, T.; LINDSTROM, S. A. M.; STANLEY, D. A.; ZARAGOZA-TRELLO, C.; NICHOLSON, C. C.; SCHEPER, J.; RAD, C.; MARKS, E. A. N.; MOTA, L.; DANFORTH, B.; PARK, M.; BEZERRA, A. D. M.; FREITAS,

[1] ACCORDING TO THE BRAZILIAN ASSOCIATION OF TECHNICAL STANDARDS (ABNT NBR 6023).

B. M.; MALLINGER, R. E.; OLIVEIRA DA SILVA, F.; WILLCOX, B.; RAMOS, D. L.; DA SILVA E SILVA, F.; LÁZARO, A.; ALOMAR, D.; GONZÁLEZ-ESTÉVEZ, M. A.; TAKI, H.; CARIVEAU, D. P.; GARRATT, M. P. D.; JODAR, D. N. N.; STEWART, R. I. A.; ARIZA, D.; PISMAN, M.; LICHTENBERG, E. M.; SCHÜEPP, C.; HERZOG, F.; ENTLING, M. H.; DUPONT, Y. L.; MICHENER, C. D.; DAILY, G. C.; EHRLICH, P. R.; BURNS, K. L. W.; VILÀ, M.; ROBSON, A.; HOWLETT, B.; BLECHSCHMIDT, L.; JAUKER, F.; SCHWARZBACH, F.; NESPER, M.; DIEKÖTTER, T.; WOLTERS, V.; CASTRO, H.; GASPAR, H.; NAULT, B. A.; BADENHAUSSER, I.; PETERSEN, J. D.; TSCHARNTKE, T.; BRETAGNOLLE, V.; CHAN, D. S. W.; CHACOFF, N.; ANDERSSON, G. K. S.; JHA, S.; COLVILLE, J. F.; VELDTMAN, R.; COUTINHO, J.; BIANCHI, F. J. J. A.; SUTTER, L.; ALBRECHT, M.; JEANNERET, P.; ZOU, Y.; AVERILL, A. L.; SAEZ, A.; SCILIGO, A. R.; VERGARA, C. H.; BLOOM, E. H.; OELLER, E.; BADANO, E. I.; LOEB, G. M.; GRAB, H.; EKROOS, J.; GAGIC, V.; CUNNINGHAM, S. A.; ASTRÖM, J.; CAVIGLIASSO, P.; TRILLO, A.; CLASSEN, A.; MAUCHLINE, A. L.; MONTERO-CASTAÑO, A.; WILBY, A.; WOODCOCK, B. A.; SIDHU, C. S.; STEFFAN-DEWENTER, I.; VOGIATZAKIS, I. N.; HERRERA, J. M.; OTIENO, M.; GIKUNGU, M. W.; CUSSER, S. J.; NAUSS, T.; NILSSON, L.; KNAPP, J.; ORTEGA-MARCOS, J. J.; GONZÁLEZ, J. A.; OSBORNE, J. L.; BLANCHE, R.; SHAW, R. F.; HEVIA, V.; STOUT, J.; ARTHUR, A. D.; BLOCHTEIN, B.; SZENTGYORGYI, H.; LI, J.; MAYFIELD, M. M.; WOYCIECHOWSKI, M.; NUNES-SILVA, P.; OLIVEIRA, R. Halinski de; HENRY, S.; SIMMONS, B. I.; DALSGAARD, B.; HANSEN, K.; SRITONGCHUAY, T.; O'REILLY, A. D.; GARCÍA, F. J. C.; PARRA, G. N.; AES PIGOZO, C. M.; BARTOMEUS, I. CropPol: A dynamic, open and global database on crop pollination. **Ecology**, v. 103, n. 3, p. e3614, 2022. ISSN 1939-9170. Disponível em: ⟨doi:10.1002/ecy.3614⟩.

ALLGEIER, J. E.; ADAM, T. C.; BURKEPILE, D. E. The importance of individual and species-level traits for trophic niches among herbivorous coral reef fishes. **Proceedings. Biological Sciences**, v. 284, n. 1856, p. 20170307, 2017. ISSN 1471-2954. Disponível em: ⟨doi:10.1098/rspb.2017.0307⟩.

ALMEIDA, J. P. A.; FALBO, R. A.; GUIZZARDI, G. Events as entities in ontology-driven conceptual modeling. In: LAENDER, A. H. F.; PERNICI, B.; LIM, E.-P.; DE OLIVEIRA, J. P. M. (Ed.). **Conceptual Modeling**. Salvador, Brazil: Springer International Publishing, 2019. ER 2019, p. 469–483. ISBN 978-3-030-33223-5.

ANDERSSON, A.; BISSET, A.; FINSTAD, A. G.; FOSSøY, F.; GROSJEAN, M.; HOPE, M.; JEPPESEN, T. S.; KÕLJALG, U.; LUNDIN, D.; NILSSON, R.; PRAGER, M.; SVENNINGSEN, C.; SCHIGEL, D. **Publishing DNA-derived Data through Biodiversity Data Platforms. v.1.0**. 2021. Disponível em: ⟨doi:10.35035/doc-vf1a-nr22⟩.

ARNAUD, E.; COOPER, L.; MENDA, N.; NELSON, R.; MATTEIS, L.; SKOFIC, M.; BASTOW, R.; JAISWAL, P.; MUELLER, L.; MCLAREN, G. Towards a Reference Plant Trait Ontology For Modeling Knowledge of Plant Traits and Phenotypes. In: . [s.n.], 2012. Disponível em: ⟨doi:10.13140/2.1.2550.3525⟩.

ARTHUR, W.; MITCHELL, P. A Revised Scheme for the Classification of Population Interactions. **Oikos**, v. 56, n. 1, p. 141–143, 1989. ISSN 0030-1299. Disponível em: ⟨doi:10.2307/3566099⟩.

ARZABE, A. A.; AGUIRRE, L. F.; BALDELOMAR, M. P.; MOLINA-MONTENEGRO, M. A. Assessing the geographic dichotomy hypothesis with cacti in South America. **Plant Biology**, v. 20, n. 2, p. 399–402, 2018. ISSN 1438-8677. Disponível em: ⟨doi:10.1111/plb.12669⟩.

BÁNKI, O.; ROSKOV, Y.; DÖRING, M.; OWER, G.; VANDEPITTE, L.; HOBERN, D.; REMSEN, D.; SCHALK, P.; DEWALT, R. E.; KEPING, M.; MILLER, J.; ORRELL, T.; AALBU, R.; ABBOTT, J.; ADLARD, R.; ADRIAENSSENS, E. M.; AEDO, C.; AESCHT, E.; AKKARI, N.; ALEXANDER, S.; ALFENAS-ZERBINI, P.; ALONSO-ZARAZAGA, M. A.; ALTENBURGER, K.; ALVAREZ, B.; ALVAREZ, F.; ANDERSON, G.; ANDRELLA, G. C.; ANTIĆ, D. Z.; ANTONIETTO, L. S.; ARANGO, C.; ARTOIS, T.; ARVANITIDIS, C.; BURGOS, M. A.; ATKINSON, S.; ATWOOD, J. J.; AUFFENBERG, K.; BAGNATORI SARTORI, Â. L.; BAILLY, N.; BAIXERAS, J.; BAKER, E.; BALAN, A.; BAMBER, R.; BANDESHA, F.; BANDYOPADHYAY, S.; BANK, R.; BARBER, A.; BARBER-JAMES, H.; BARBOSA, J. P.; PINTO, R. B.; BARRETT, R.; BARTOLOZZI, L.; BARTSCH, I.; BECCALONI, G. W.; BELLAMY, C.; BELLAN-SANTINI, D.; BELLINGER, P.; BEN-DOV, Y.; BERNOT, J.; BEZERRA, T. N.; BIELER, R.; BITNER, M. A.; BLASCO-COSTA, I.; BOATWRIGHT, J. S.; BOCK, P.; BONATO, L.; BORGES, L. M.; BOTA-SIERRA, C.; BOUCHARD, P.; BOUCHET, P.; BOURGOIN, T.; BOURY-ESNAULT, N.; BOUZAN, R.; BOXSHALL, G.; BOYKO, C.; BRANDÃO, S.; BRAUN, H.; BRAY, R.; BRINDA, J. C.; BROCK, P. D.; BROICH, S. L.; BRONSTEIN, O.; BROWN, J.; BRUCE, N.; BRULLO, S.; BRUNEAU, A.; BUENO-VILLEGAS, J.; BURCKHARDT, D.; BUSH, L.; BÖTTGER-SCHNACK, R.; BÜSCHER, T.; BŁAŻEWICZ-PASZKOWYCZ, M.; CAIRNS, S.; CALONJE, M.; CAMILO DE OLIVEIRA, J. P.; CARBALLO, J. L.; CARDINAL-MCTEAGUE, W.; CARDOSO, D.; CARDOSO, L.; CARRERA-PARRA, L.; CASTILHO, R.; SILVA, I. C. C.; CATALANO, S.; CERVANTES, A.; CHATROU, L.; CHEVILLOTTE, H.; CHOO, L. M.; CHRISTIANSEN, K.; CIANFERONI, F.; CIGLIANO, M. M.; CLARKE, R.; COBRA E MONTEIRO, T.; COLLINS, A.; COLLINS, K.; COMPTON, J.; CONSORTI, L.; COPILAŞ-CIOCIANU, D.; CORBARI, L.; CORDEIRO, R.; CoreoideaSF Team; CORNILS, A.; COSTA CORGOSINHO, P. H.; COSTELLO, M.; CRAMERI, S.; CRUZ-LÓPEZ, J. A.; CULHAM, A.; CÁRDENAS, P.; DALY, M.; DANELIYA, M.; DAUVIN, J.-C.; DAVIE, P.; DAVISON, A. J.; DE BROYER, C.; DE LIMA, H. C.; DE PRINS, J.; DE PRINS, W.; DE LA ESTRELLA, M.; DESALLE, R.; DECKER, P.; DECOCK, W.; DEEM, L. S.; DEFAYE, D.; DEKKER, H.; DELGADO-SALINAS, A.; DELIRY, C.; DELLAPÉ, P. M.; DEMPSEY, D. M.; DEN HEYER, J.; DEPREZ, T.; DESIDERATO, A.; DI CAPUA, I.; DIJKSTRA, K.-D.; DIPPENAAR, S.; DMITRIEV, D.; DOHRMANN, M.; DONER, S.; DORADO, Ó.; DORKELD, F.; DOWNEY, R.; DUAN, L.; DUCARME, F.; DUTILH, B. E.; DÍAZ, M.-C.; EADES, D. C.; EGAN, A. N.; EIBYE-JACOBSEN, D.; EISENDLE, U.; EITEL, M.; EL NAGAR, A.; EMIG, C.; EMIG, C. C.; ENGEL, M. S.; ENGHOFF, H.; EVANS, G.; EVENHUIS, N. L.; FABER, M.; FALCÃO, M.; FARJON, A.; FARRUGGIA, F.; FAUCHALD, K.; FAUTIN, D.; FAVRET, C.; FERNÁNDEZ-RODRÍGUEZ, V.; FIGUEROA, D.; FIŠER, C.; FORRÓ, L.; FORSTNER, M.; FORTUNA-PEREZ, A. P.; FRANCIS, A.; FRITSCH, P.; FROESE, R.; FUCHS, A.; FUJIMOTO, S.; FURUYA, H.; GAGNON, E.; GARCIA-ALVAREZ, O.; GARCÍA, M. L.; GARDNER, M.; GARIC, R.; GARNETT, S.; GASCA, R.; GATTOLLIAT, J.-L.; GAVIRIA-MELO, S.; GERKEN, S.; GIBSON, D.; GIBSON, R.; GIELIS, C.; GILLIGAN, T.; GIRIBET, G.; GITTENBERGER, A.; GIUSSO DEL GALDO, G. P.;

GLASBY, C.; GLOVER, A. G.; GODOY, M. Á.; GOFAS, S.; GONCHAROV, M.; GONDIM, A. I.; GOODWIN, C.; GOVAERTS, R.; GRABOWSKI, M.; GRANADO, A. d. A.; GRAY, A.; GREGÓRIO, B. d. S.; GRETHER, R.; GRIMALDI, D. A.; GROSS, O.; GRUN, T. B.; GUERRA-GARCÍA, J. M.; GUGLIELMONE, A.; GUILBERT, E.; GUSENLEITNER, J.; GÓMEZ-NOGUERA, S. E.; HAAS, F.; HADFIELD, K. A.; HAJDU, E.; HARRACH, B.; HARRIS, L.; HARRISON, R. L.; HASSLER, M.; HAYWARD, B. W.; HEADS, S. W.; HENDRICKSON, R. C.; HENDRYCKS, E.; HENRY, T. J.; HERBERT, D.; HERNANDES, F.; HERNANDEZ, F.; HERNÁNDEZ-CRESPO, J. C.; HERRERA BACHILLER, A.; HINE, A.; HIRSCH, H.; HO, J.-s.; HODDA, M.; HODSON, A.; HOEKSEMA, B.; HOENEMANN, M.; HOLOVACHOV, O.; HOLSTEIN, J.; HOOGE, M.; HOOPER, J.; HOPKINS, H.; HORAK, I.; HORTON, T.; HOSOYA, T.; HOUART, R.; HOŠEK, J.; HUGHES, L.; HUIJBERS, C.; HÄUSER, C.; INIESTA, L. F. M.; IVANENKO, V. S.; JANSSEN, R.; JANSSENS, F.; JAUME, D.; JAVADI, F.; JAZDZEWSKI, K.; JOHNSON, K. P.; JORDÃO, L.; JUNGLEN, S.; JÓŹWIAK, P.; KABAT, A.; KAMIŃSKI, M. J.; KANDA, K.; KANTOR, Y.; KARANOVIC, I.; KARAPUNAR, B.; KATHIRITHAMBY, J.; KELLY, M.; KIM, Y.-H.; KING, R.; KIRK, P.; KITCHING, I.; KLAUTAU, M.; KLITGAARD, B. B.; KNOWLES, N. J.; KOENEMANN, S.; KOROVCHINSKY, N.; KOTOV, A.; KOUWENBERG, J.; KOVÁCS, Z.; KRAMINA, T.; KRAPF, A.; KRAPP-SCHICKEL, T.; KREMENETSKAIA, A.; KRISHNA, K.; KRISHNA, V.; KROH, A.; KROUPA, A.; KRUPOVIC, M.; KUHN, J. H.; KURY, A. B.; KURY, M. S.; KVAČEK, J.; KÖHLER, F.; LACHENAUD, O.; LADO, C.; LAMBERT, A. J.; LAMBERT, G.; LANA C. ATUNES, L.; LAZARUS, D.; COZE, F. L.; ROUX, M. M. L.; LECROY, S.; LINARES, J. L.; LEDUC, D.; LEFKOWITZ, E. J.; LEWIS, G. P.; LI, S.-J.; LI-QIANG, J.; LICHTWARDT, R.; LIM, S.-C.; LOBANOV, A.; LOHRMANN, V.; LONDOÑO-MESA, M.; LONGHORN, S. J.; LORENZ, W.; LOWRY, J.; LOZANO, F.; LUJAN-TORO, B. E.; LUMEN, R.; LYAL, C. H.; LYANGOUZOV, I.; LÖRZ, A.-N.; MACKLIN, J. A.; MADIN, L.; MAEHR, M. D.; MAGNIEN, P.; MAH, C.; MAL, N.; MAMOS, T.; MANCONI, R.; MANSANO, V.; MAREK, P.; MARSHALL, B.; MARTIN, J. H.; MARTIN, P.; MARTIN, S. L.; MARTÍNEZ-MELO, A.; MARTÍNEZ-MUÑOZ, C. A.; MASHEGO, K. S.; MASLIN, B.; MATTAPHA, S.; MCFADDEN, C.; MCKAMEY, S.; MCMURTRY, J.; MEDRANO, M. A.; MEDVEDEV, S.; MEES, J.; MEJÍA-MADRID, H. H.; MENDES, A. C.; MERRIN, K.; MESA, N.; MESSING, C.; MIGEON, A.; MILLER, D. R.; MILLS, C.; MINELLI, A.; MISKELLY, A.; MITCHELL, D.; MOKIEVSKY, V.; MOLODTSOVA, T.; KOCH, N. M.; VALLS, J. F. M.; MOOI, R.; MORANDINI, A.; MOREIRA DA ROCHA, R.; MORROW, C.; MOTEETEE, A.; MURPHY, B.; MUSHEGIAN, A. R.; NARITA, J.; NEALOVA, L.; NERY, D. G.; NEU-BECKER, U.; NEUBAUER, T. A.; NEUBERT, E.; NEUHAUS, B.; NEWTON, A.; NG KEE LIN, P.; NGUYEN, A.; NIBERT, M. L.; NICOLSON, D.; NIELSEN, S.; NIJHOF, A.; NISHIKAWA, T.; NORENBURG, J.; NOYES, J.; O'HARA, T.; OCHOA, R.; OHASHI, H.; OHASHI, K.; OKSANEN, H. M.; OLLERENSHAW, J.; OOSTERBROEK, P.; OPRESKO, D.; ORTON, R. J.; OSBORNE, R.; OSIGUS, H.-J.; OSWALD, J.; OTA, Y.; OTTE, D.; OUVRARD, D.; Paleobiology Database contributors; PANDEY, A.; PAPE, T.; PAULAY, G.; PAULSON, D.; PAULY, D.; PAXTON, H.; PEDRAM, M.; PENNINGTON, R. T.; PEREIRA, J. d. S.; PEREZ-GELABERT, D.; PETRUSEK, A.; SANTIAGO, R. P.; PHILLIPSON, P.; PIASECKI, W.; PICTON, B.; PINHEIRO, U.; PISERA, A.; PITKIN, B.; POORE, G.; POVYDYSH, M.; PRAXEDES, R. A.; PULAWSKI, W.; PYLE, R.; PÁLL-GERGELY,

B.; PÉREZ-GARCÍA, J. A.; PUŽA, V.; RAINER, H.; RAKOTONIRINA, N.; RAMOS, G.; FILARDI, F. R.; RAZ, L.; READ, G.; REES, T.; REICH, M.; REIMER, J. D.; REIN, J. O.; REIP, H.; REUSCHER, M.; REWICZ, T.; REYNOLDS, J.; RICHLING, I.; RIUS, M.; ROBERTSON, D. L.; ROBERTSON, T.; ROBINSON, G.; ROBINSON, G. S.; RODRÍGUEZ, E.; ROMANI, L.; ROSENBERG, G.; RUBINO, L.; RUGGIERO, M.; RÍOS, P.; RüTZLER, K.; SABANADZOVIC, S.; SALAZAR-VALLEJO, S.; SANBORN, A.; SANJAPPA, M.; SANTOS, S. G.; SANTOS-GUERRA, A.; SARAIVA DE OLIVEIRA, J.; SARTORI, M.; SATTLER, K.; SAUCÉDE, T.; SCHIERWATER, B.; SCHILLING, S.; SCHLEY, R.; SCHMID-EGGER, C.; SCHMIDT-RHAESA, A.; SCHNEIDER, S.; SCHOOLMEESTERS, P.; SCHORR, M.; SCHRIRE, B.; SCHUCHERT, P.; SCHUH, R.; SCHÖNBERG, C.; RODRIGUES, R. S.; SCOBLE, M.; SEGERS, H.; SEIJO, G.; SELEME, E. P.; SENNA, A.; SEREJO, C.; SFORZI, A.; SHARMA, J.; SHEAR, W.; SHENKAR, N.; SHORT, M.; SICIŃSKI, J.; SIDDELL, S. G.; SIEGEL, V.; SIERWALD, P.; SIGDA, L.; SILVA, E.; SILVA FLORES, A.; CARVALHO, C. Silva de; SIMMONDS, P.; SIMMONS, E.; SIMON, M. F.; SIMONSEN, T.; SIMPSON, C. E.; SIRICHAMORN, Y.; SMITH, A. D.; SMITH, D. B.; SMITH, V. S.; SMOL, N.; GISSI, D. S.; SOKOLOFF, D.; SOULIER-PERKINS, A.; SOUTH, E. J.; SOUZA-FILHO, J. F.; SPEARMAN, L.; SPELDA, J.; STEGER, J.; STEINER, A.; STEMME, T.; STERRER, W.; STEVENSON, D.; STIEWE, M. B. D.; STIRTON, C. H.; STJERNEGAARD JEPPESEN, T.; STOEV, P.; STRAND, M.; STRAUB, S.; STUEBER, G.; STÖHR, S.; SUBRAMANIAM, S.; SUZUKI, N.; SUÁREZ-MORALES, E.; SWALLA, B.; SWEDO, J.; SZUMIK, C.; SÁNCHEZ-RUIZ, M.; SØRENSEN, M. V.; TAITI, S.; TAKIYA, D.; TANDBERG, A. H.; TANG, D.; TAVAKILIAN, G.; TAYLOR, J.; TAYLOR, K.; TCHESUNOV, A.; THESSEN, A.; THOMAS, J. D.; THOMAS, P.; ThripsWiki; THUESEN, E.; THULIN, M.; THURSTON, M.; THUY, B.; TODARO, A.; TODD, J.; TORKE, B. M.; TURIAULT, M.; TURON, X.; TYLER, S.; UETZ, P.; ULMER, J. M.; URIBE-PALOMINO, J.; VACELET, J.; VACHARD, D.; VADER, W.; VAN DOOERSLAER, K.; VAN DER BURGT, X.; VANDAMME, A.-M.; VANHOORNE, B.; VANREUSEL, A.; VARSANI, A.; VATANPARAST, M.; VENEKEY, V.; VINARSKI, M.; VONK, R.; VOS, C.; VäINÖLÄ, R.; WALKER, P. J.; WALKER-SMITH, G.; WALTER, T. C.; WAMBIJI, N.; WARWICK, S.; WATLING, L.; WEAVER, H.; WEBB, J.; WELBOURN, W.; WESENER, T.; WHIPPS, C.; WHITE, K.; WIENEKE, U.; WILDING, N.; WILSON, A. J.; WILSON, R.; WING, P.; WINITSKY, S.; WIRTH, C. C.; WOJCIECHOWSKI, M.; WOODMAN, S.; World Spider Catalog; XAVIER, J.; YESSON, C.; YI, T.; YODER, M.; YU, D. S. K.; YUNAKOV, N.; ZAHNISER, J.; ZANOL, J.; ZEIDLER, W.; ZERBINI, F. M.; ZHANG, R.; ZHANG, Z.; ZHAO, Z.; ZIEGLER, A.; ZINETTI, F.; ZULLINI, A.; DE MORAES, G.; DE VOOGD, N.; TEN HOVE, H.; TER POORTEN, J. J.; VAN HAAREN, T.; VAN NIEUKERKEN, E.; VAN SOEST, R.; ŁOBOCKA, M.; ŞENTÜRK, O.; ITIS; International Committee on Taxonomy of Viruses (ICTV); Legume Phylogeny Working Group (LPWG). Catalogue of life checklist. **Catalogue of Life**, 2023. ISSN 2405-8858. Disponível em: ⟨doi:doi:10.48580/dfry⟩.

BASKAUF, S. J.; WEBB, C. O. Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF. **Semantic Web**, v. 7, n. 6, p. 629–643, 2016. Disponível em: ⟨doi:doi:10.3233/SW-150203⟩.

BASSON, I.; SIMARD, M.-A.; OUANGRÉ, Z. A.; SUGIMOTO, C. R.; LARIVIèRE, V. The effect of data sources on the measurement of open access: A comparison of dimensions and the web of science. **PLOS ONE**, v. 17, n. 3, p. e0265545, 2023. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0265545⟩.

BECK, J.; BÖLLER, M.; ERHARDT, A.; SCHWANGHART, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. **Ecological Informatics**, v. 19, p. 10–15, 2014. ISSN 1574-9541. Disponível em: ⟨doi:10.1016/j.ecoinf.2013.11.002⟩.

BEGON, M.; TOWNSEND, C. R. **Ecology: From Individuals to Ecosystems**. United Kingdom: John Wiley & Sons, 2021. 868 p. ISBN 978-1-119-27935-8.

BEHESHTI, A.; GHODRATNAMA, S.; ELAHI, M.; FARHOOD, H. **Social Data Analytics**. United Kingdom: CRC Press, 2022. 251 p. ISBN 978-1-00-064460-9.

BERENDSOHN, W.; GÜNTSCH, A.; HOFFMANN, N.; KOHLBECKER, A.; LUTHER, K.; MÜLLER, A. Biodiversity information platforms: From standards to interoperability. **ZooKeys**, v. 150, p. 71–87, 2011. ISSN 1313-2970. Disponível em: ⟨doi:10.3897/zookeys.150.2166⟩.

BERLOW, E. L.; NEUTEL, A.-M.; COHEN, J. E.; RUITER, P. C. D.; EBENMAN, B.; EMMERSON, M.; FOX, J. W.; JANSEN, V. A. A.; JONES, J. I.; KOKKORIS, G. D.; LOGOFET, D. O.; MCKANE, A. J.; MONTOYA, J. M.; PETCHEY, O. Interaction Strengths in Food Webs: Issues and Opportunities. **Journal of Animal Ecology**, v. 73, n. 3, p. 585–598, 2004. ISSN 0021-8790. Disponível em: ⟨https://www.jstor.org/stable/3505669⟩.

BERMAN, F.; WILKINSON, R.; WOOD, J. Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance. **D-Lib Magazine**, v. 20, n. 1/2, 2014. ISSN 1082-9873. Disponível em: ⟨doi:10.1045/january2014-berman⟩.

BERNERS-LEE, T. Realising the Full Potential of the Web. **W3C Consortium**, 1997. Disponível em: ⟨http://www.w3.org/1998/02/Potential.html⟩.

Biodiversity Information Standards (TDWG). TDWG Wiki Archive - EnvironmentMeasurementsExtension. 2007. Disponível em: ⟨https://github.com/tdwg/wiki-archive/blob/d77f897a52d96f1bd974d5c438790017b8419fac/twiki/data/DarwinCore/EnvironmentMeasurementsExtension.txt⟩.

Biodiversity Information Standards (TDWG). TDWG Wiki Archive - InteractionExtenstion. 2007. Disponível em: ⟨https://github.com/tdwg/wiki-archive/blob/d77f897a52d96f1bd974d5c438790017b8419fac/twiki/data/DarwinCore/InteractionExtension.txt⟩.

Biodiversity Information Standards (TDWG). TDWG Wiki Archive - PollinationExtension. 2007. Disponível em: ⟨https://github.com/tdwg/wiki-archive/blob/d77f897a52d96f1bd974d5c438790017b8419fac/twiki/data/DarwinCore/PollinationExtension.txt⟩.

BIOGOUDA. **A biodiversity dataset graph archive**. 2022. Disponível em: ⟨hash://sha256/7cd305e9d275763c96e7685847460fcc381b5c97c1460c00441f663c1788800f⟩.

BISBY, F. A. The quiet revolution: Biodiversity informatics and the internet. **Science**, v. 289, n. 5488, p. 2309–2312, 2000. Disponível em: ⟨doi:10.1126/science.289.5488.2309⟩.

BLEI, D. M.; KUCUKELBIR, A.; MCAULIFFE, J. D. Variational inference: A review for statisticians. **Journal of the American statistical Association**, v. 112, n. 518, p. 859–877, 2017.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **The Journal of Machine Learning Research**, v. 3, p. 993–1022, 2003. ISSN 1532-4435.

BOAKES, E. H.; McGowan, P. J. K.; FULLER, R. A.; CHANG-QING, D.; CLARK, N. E.; O'CONNOR, K.; MACE, G. M. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. **PLOS Biology**, v. 8, n. 6, p. e1000385, 2010. ISSN 1545-7885. Disponível em: ⟨doi:10.1371/journal.pbio.1000385⟩.

BOLNICK, D. I.; AMARASEKARE, P.; ARAÚJO, M. S.; BÜRGER, R.; LEVINE, J. M.; NOVAK, M.; RUDOLF, V. H. W.; SCHREIBER, S. J.; URBAN, M. C.; VASSEUR, D. A. Why intraspecific trait variation matters in community ecology. **Trends in Ecology & Evolution**, v. 26, n. 4, p. 183–192, 2011. ISSN 0169-5347. Disponível em: ⟨doi:10.1016/j.tree.2011.01.009⟩.

BRADSHAW, H. D.; SCHEMSKE, D. W. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. **Nature**, v. 426, n. 6963, p. 176–178, 2003. ISSN 1476-4687. Disponível em: ⟨doi:10.1038/nature02106⟩.

BRAZMA, A.; HINGAMP, P.; QUACKENBUSH, J.; SHERLOCK, G.; SPELLMAN, P.; STOECKERT, C.; AACH, J.; ANSORGE, W.; BALL, C. A.; CAUSTON, H. C.; GAASTERLAND, T.; GLENISSON, P.; HOLSTEGE, F. C. P.; KIM, I. F.; MARKOWITZ, V.; MATESE, J. C.; PARKINSON, H.; ROBINSON, A.; SARKANS, U.; SCHULZE-KREMER, S.; STEWART, J.; TAYLOR, R.; VILO, J.; VINGRON, M. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. **Nature Genetics**, v. 29, n. 4, p. 365–371, 2001. ISSN 1546-1718. Disponível em: ⟨doi:10.1038/ng1201-365⟩.

BRENSKELLE, L.; WIECZOREK, J.; DAVIS, E.; WALLIS, N. J.; EMERY, K.; LEFEBVRE, M. J.; GURALNICK, R. A community-developed extension to Darwin Core for reporting the chronometric age of specimens. **PLoS ONE**, v. 17, n. 9, p. e0261044. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0261044⟩.

BRIMACOMBE, C.; BODNER, K.; MICHALSKA-SMITH, M.; POISOT, T.; FORTIN, M.-J. Shortcomings of reusing species interaction networks created by different sets of researchers. **PLOS Biology**, v. 21, n. 4, p. e3002068, 2023. ISSN 1545-7885. Disponível em: ⟨doi:doi:10.1371/journal.pbio.3002068⟩.

BRONSTEIN, J. L. Conditional outcomes in mutualistic interactions. **Trends in Ecology & Evolution**, v. 9, n. 6, p. 214–217, 1994. ISSN 0169-5347. Disponível em: ⟨doi:10.1016/0169-5347(94)90246-1⟩.

BRONSTEIN, J. L. Mutualisms. In: FOX, C. W.; ROFF, D. A.; FAIRBAIRN, D. J. (Ed.). **Evolutionary Ecology: Concepts and Case Studies**. United Kingdom: Oxford University Press, 2001. ISBN 978-0-19-803013-3.

BROSE, U.; ARCHAMBAULT, P.; BARNES, A. D.; BERSIER, L.-F.; BOY, T.; CANNING-CLODE, J.; CONTI, E.; DIAS, M.; DIGEL, C.; DISSANAYAKE, A.; FLORES, A. A. V.; FUSSMANN, K.; GAUZENS, B.; GRAY, C.; HÄUSSLER, J.; HIRT, M. R.; JACOB, U.; JOCHUM, M.; KÉFI, S.; MCLAUGHLIN, O.; MACPHERSON, M. M.; LATZ, E.; LAYER-DOBRA, K.; LEGAGNEUX, P.; LI, Y.; MADEIRA, C.; MARTINEZ, N. D.; MENDONÇA, V.; MULDER, C.; NAVARRETE, S. A.; O'GORMAN, E. J.; OTT, D.; PAULA, J.; PERKINS, D.; PIECHNIK, D.; POKROVSKY, I.; RAFFAELLI, D.; RALL, B. C.; ROSENBAUM, B.; RYSER, R.; SILVA, A.; SOHLSTRÖM, E. H.; SOKOLOVA, N.; THOMPSON, M. S. A.; THOMPSON, R. M.; VERMANDELE, F.; VINAGRE, C.; WANG, S.; WEFER, J. M.; WILLIAMS, R. J.; WIETERS, E.; WOODWARD, G.; ILES, A. C. Predator traits determine food-web architecture across ecosystems. **Nature Ecology & Evolution**, v. 3, n. 6, p. 919–927, 2019. ISSN 2397-334X. Disponível em: ⟨doi:10.1038/s41559-019-0899-x⟩.

BROUSSEAU, P.-M.; GRAVEL, D.; HANDA, I. T. Trait matching and phylogeny as predictors of predator–prey interactions involving ground beetles. **Functional Ecology**, v. 32, n. 1, p. 192–202, 2018. ISSN 1365-2435. Disponível em: ⟨doi: 10.1111/1365-2435.12943⟩.

BURKEPILE, D. E.; PARKER, J. D. Recent advances in plant-herbivore interactions. **F1000Research**, v. 6, p. 119, 2017. ISSN 2046-1402. Disponível em: ⟨doi: 10.12688/f1000research.10313.1⟩.

BURKHOLDER, P. R. Cooperation and conflict among primitive organisms. **American Scientist**, v. 40, n. 4, p. 600–631, 1952.

BUTTERFIELD, B. J.; CALLAWAY, R. M. A functional comparative approach to facilitation and its context dependence. **Functional Ecology**, v. 27, n. 4, p. 907–917, 2013. ISSN 1365-2435. Disponível em: ⟨doi:10.1111/1365-2435.12019⟩.

BUTTIGIEG, P. L.; MORRISON, N.; SMITH, B.; MUNGALL, C. J.; LEWIS, S. E.; the ENVO Consortium. The environment ontology: Contextualising biological and biomedical entities. **Journal of Biomedical Semantics**, v. 4, n. 1, p. 43, 2013. ISSN 2041-1480. Disponível em: ⟨doi:10.1186/2041-1480-4-43⟩.

CALIŃSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics**, v. 3, n. 1, p. 1–27, 1974. ISSN 0090-3272. Disponível em: ⟨doi:10.1080/03610927408827101⟩.

CANHOS, V. P.; SOUZA, S. d.; GIOVANNI, R. D.; CANHOS, D. A. L. Global biodiversity informatics: setting the scene for a "new world" of ecological forecasting. **Biodiversity Informatics**, v. 1, 2004. ISSN 1546-9735. Disponível em: ⟨doi:10.17161/bi.v1i0.3⟩.

CARSON, R. **Silent Spring**. Boston: Houghton Mifflin Harcourt, 1962. 404 p. ISBN 978-0-618-24906-0.

CARTOLANO, E. A. **Proposta de um sistema de informação orientado a serviços sobre a biodiversidade de abelhas**. Tese (PhD thesis) — Universidade de São Paulo, August 2009. Disponível em: ⟨doi:10.11606/D.3.2009.tde-23092009-151526⟩.

CARVALHEIRO, L. G.; SARAIVA, A. M.; GIANNINI, T. C. Establishing Knowledge Management Systems for Ecological Interactions: The case of crop pollinators. In:

Pollination Services to Agriculture. United Kingdom: Taylor & Francis, 2016. p. 21. ISBN 978-1-315-69635-5.

CHAMBERLAIN, S. A.; BRONSTEIN, J. L.; RUDGERS, J. A. How context dependent are species interactions? **Ecology Letters**, v. 17, n. 7, p. 881–890, 2014. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/ele.12279⟩.

CHAN, L. M.; ZENG, M. L. Metadata Interoperability and Standardization - A Study of Methodology Part I: Achieving Interoperability at the Schema Level. **D-Lib Magazine**, v. 12, n. 6, 2006. ISSN 1082-9873. Disponível em: ⟨doi:doi:10.1045/june2006-chan⟩.

CHOWDHARY, K. R. Natural language processing. In: CHOWDHARY, K. (Ed.). **Fundamentals of Artificial Intelligence**. Springer India, 2020. p. 603–649. ISBN 978-81-322-3972-7. Disponível em: ⟨doi:10.1007/978-81-322-3972-7_19⟩.

CICCARESE, P.; SOILAND-REYES, S.; BELHAJJAME, K.; GRAY, A. J.; GOBLE, C.; CLARK, T. PAV ontology: Provenance, authoring and versioning. **Journal of Biomedical Semantics**, v. 4, n. 1, p. 37, 2013. ISSN 2041-1480. Disponível em: ⟨doi:doi:10.1186/2041-1480-4-37⟩.

CIRTWILL, A. R.; EKLÖF, A. Feeding environment and other traits shape species' roles in marine food webs. **Ecology Letters**, v. 21, n. 6, p. 875–884, 2018. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/ele.12955⟩.

CODATA, Committee on Data of the International Science Council CODATA; CODATA International Data Policy Committee; CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice; HODSON, S.; MONS, B.; UHLIR, P.; ZHANG, L. The Beijing Declaration on Research Data. **Zenodo**, 2019. Disponível em: ⟨doi:10.5281/zenodo.3552330⟩.

CONSORTIUM, T. P. O. The Plant Ontology Consortium and Plant Ontologies. **Comparative and Functional Genomics**, v. 3, n. 2, p. 137–142, 2002. ISSN 1532-6268. Disponível em: ⟨doi:10.1002/cfg.154⟩.

COUX, C. Plant-pollinator pollen transportation network, traits and pollinator abundances. **figshare**, 2016. Dataset. Disponível em: ⟨doi:10.6084/m9.figshare.3154078.v1⟩.

CRYSTAL-ORNELAS, R.; VARADHARAJAN, C.; BOND-LAMBERTY, B.; BOYE, K.; BURRUS, M.; CHOLIA, S.; CROW, M.; DAMEROW, J.; DEVARAKONDA, R.; ELY, K. S.; GOLDMAN, A.; HEINZ, S.; HENDRIX, V.; KAKALIA, Z.; PENNINGTON, S. C.; ROBLES, E.; ROGERS, A.; SIMMONDS, M.; VELLIQUETTE, T.; WEIERBACH, H.; WEISENHORN, P.; WELCH, J. N.; AGARWAL, D. A. A Guide to Using GitHub for Developing and Versioning Data Standards and Reporting Formats. **Earth and Space Science**, v. 8, n. 8, p. e2021EA001797, 2021. ISSN 2333-5084. Disponível em: ⟨doi:10.1029/2021EA001797⟩.

CYGANIAK, R.; WOOD, D.; LANTHALER, M. RDF 1.1 Concepts and Abstract Syntax. **W3C Consortium**, 2014. Disponível em: ⟨http://www.w3.org/TR/rdf11-concepts/⟩.

Darwin Core and RDF/OWL Task Groups. Darwin Core RDF guide. **Biodiversity Information Standards (TDWG)**, 2015. Disponível em: ⟨http://rs.tdwg.org/dwc/terms/guides/rdf/2021-07-15⟩.

Darwin Core Maintenance Group. Darwin Core XML guide. **Biodiversity Information Standards (TDWG)**, 2021. Disponível em: ⟨http://rs.tdwg.org/dwc/terms/guides/xml/2021-07-15⟩.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-1, n. 2, p. 224–227, 1979. ISSN 1939-3539. Disponível em: ⟨doi:10.1109/TPAMI.1979.4766909⟩.

DAVISON, A. Biological Mutualism: A Scientific Survey. **Theology and Science**, v. 18, n. 2, p. 190–210, 2020. ISSN 1474-6700. Disponível em: ⟨doi:10.1080/14746700.2020.1755534⟩.

DCMI Usage Board. DCMI Metadata Terms. **Dublin Core**, 2020. Disponível em: ⟨https://www.dublincore.org/specifications/dublin-core/dcmi-terms/⟩.

DEERWESTER, S.; DUMAIS, S. T.; FURNAS, G. W.; LANDAUER, T. K.; HARSHMAN, R. Indexing by latent semantic analysis. **Journal of the American Society for Information Science**, v. 41, n. 6, p. 391–407, 1990. ISSN 1097-4571. Disponível em: ⟨doi:10.1002/(SICI)1097-4571(199009)41:6⟨391::AID-ASI1⟩3.0.CO;2-9⟩.

DELMAS, E.; BESSON, M.; BRICE, M.-H.; BURKLE, L. A.; RIVA, G. V. D.; FORTIN, M.-J.; GRAVEL, D.; GUIMARÃES JR., P. R.; HEMBRY, D. H.; NEWMAN, E. A.; OLESEN, J. M.; PIRES, M. M.; YEAKEL, J. D.; POISOT, T. Analysing ecological networks of species interactions. **Biological Reviews**, v. 94, n. 1, p. 16–36, 2019. ISSN 1469-185X. Disponível em: ⟨doi:10.1111/brv.12433⟩.

DEMBOSKI, J. DMNS Parasite Collection (Arctos). Version 34.73. 2023. Accessed via GBIF.org on 2023-05-10. Disponível em: ⟨doi:10.15468/jijwox⟩.

DIGHTON, J. **Fungi in Ecosystem Processes**. United States: CRC Press, 2018. 434 p. ISBN 978-1-315-36016-4.

Digital Science. **Dimensions**. 2018. Accessed via on 2023-01-12. Disponível em: ⟨https://app.dimensions.ai/⟩.

DINDAL, D. Symbiosis: Nomenclature and proposed classification. **Biologist**, v. 57, p. 129–142, 1975.

DORMANN, C. F.; FRÜND, J.; SCHAEFER, H. M. Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. **Annual Review of Ecology, Evolution, and Systematics**, v. 48, n. 1, p. 559–584, 2017. Disponível em: ⟨doi:10.1146/annurev-ecolsys-110316-022928⟩.

DUVAL, E.; HODGINS, W.; SUTTON, S.; WEIBEL, S. L. Metadata Principles and Practicalities. **D-Lib Magazine**, v. 8, n. 4, 2002. ISSN 1082-9873. Disponível em: ⟨doi:doi:10.1045/april2002-weibel⟩.

ECHELLE, A. A.; ECHELLE, A. F.; HILL, L. G. Interspecific interactions and limiting factors of abundance and distribution in the red river pupfish, cyprinodon rubrofluviatilis. **The American Midland Naturalist**, v. 88, n. 1, p. 109–130, 1972. ISSN 0003-0031. Disponível em: ⟨doi:10.2307/2424492⟩.

ELLIOTT, M. J.; POELEN, J. H.; FORTES, J. A. B. Toward reliable biodiversity dataset references. **Ecological Informatics**, v. 59, p. 101132, 2020. ISSN 1574-9541. Disponível em: ⟨doi:doi:10.1016/j.ecoinf.2020.101132⟩.

ENDRESEN, D.; GAIJI, S.; ROBERTSON, T. Darwin Core Germplasm Extension and deployment in the GBIF infrastructure. In: . [s.n.], 2009. Disponível em: ⟨doi:10.13140/2.1.1207.3923⟩.

FAIRsharing.org: SRAO. **Subject Resource Application Ontology**. 2022. Disponível em: ⟨doi:10.25504/FAIRsharing.b1xD9f⟩.

FAYLE, T. M.; SAM, K.; HUMLOVA, A.; CAGNOLO, L. The LifeWebs project: A call for data describing plant-herbivore interaction networks. **International Biogeography Society**, 2016. ISSN 1948-6596. Disponível em: ⟨doi:10.21425/F58431122⟩.

FEDERHEN, S. The NCBI Taxonomy database. **Nucleic Acids Research**, v. 40, p. D136–D143, 2012. ISSN 0305-1048. Disponível em: ⟨doi:doi:10.1093/nar/gkr1178⟩.

FINSTAD, A. G.; ANDERSSON, A.; BISSETT, A.; FOSSØY, F.; GROSJEAN, M.; HOPE, M.; KÕLJALG, U.; LUNDIN, D.; NILSSON, H.; PRAGER, M.; JEPPESEN, T. S.; SVENNINGSEN, C.; SCHIGEL, D. Publishing sequence-derived data through biodiversity data platforms. **GBIF Secretariat**, 2020. Disponível em: ⟨doi:10.35035/DOC-VF1A-NR22⟩.

FORTUNA, M. A.; ORTEGA, R.; BASCOMPTE, J. The Web of Life. 2014. Disponível em: ⟨doi:10.48550/arXiv.1403.2575⟩.

FREDERICKSON, M. E. Mutualisms Are Not on the Verge of Breakdown. **Trends in Ecology & Evolution**, v. 32, n. 10, p. 727–734, 2017. ISSN 0169-5347. Disponível em: ⟨doi:10.1016/j.tree.2017.07.001⟩.

GBIF. **Darwin Core Archives – How-to Guide, Version 2.2**. 2021. Disponível em: ⟨https://ipt.gbif.org/manual/en/ipt/2.5/dwca-guide⟩.

GBIF Secretariat. GBIF Backbone Taxonomy. Checklist dataset. 2022. Disponível em: ⟨doi:doi:10.15468/39omei⟩.

GEURTS, E. M.; REYNOLDS, J. D.; STARZOMSKI, B. M. Turning observations into biodiversity data: Broadscale spatial biases in community science. **Ecosphere**, v. 14, n. 6, p. e4582, 2023. ISSN 2150-8925. Disponível em: ⟨doi:10.1002/ecs2.4582⟩.

GitHub. **New Term - relationshipOfResourceID · Issue #283 · Tdwg/Dwc**. 2023. Disponível em: ⟨https://github.com/tdwg/dwc/issues/283⟩.

GÓMEZ, J. M.; IRIONDO, J. M.; TORRES, P. Modeling the continua in the outcomes of biotic interactions. **Ecology**, v. 104, n. 4, p. e3995, 2023. ISSN 1939-9170. Disponível em: ⟨doi:10.1002/ecy.3995⟩.

GRIFFITHS, T. L.; STEYVERS, M. Finding scientific topics. **Proceedings of the National Academy of Sciences**, v. 101, p. 5228–5235, 2004. Disponível em: ⟨doi:10.1073/pnas.0307752101⟩.

GUIMARÃES, P.; RAIMUNDO, R.; CAGNOLO, L. Interaction web database. **Universidade de São Paulo**, 2012. Disponível em: ⟨http://www.ecologia.ib.usp.br/iwdb/⟩.

GUIZZARDI, G.; WAGNER, G.; FALBO, R. de A.; GUIZZARDI, R. S. S.; ALMEIDA, J. P. A. Towards Ontological Foundations for the Conceptual Modeling of Events. In: NG, W.; STOREY, V. C.; TRUJILLO, J. C. (Ed.). **Conceptual Modeling**. Springer, 2013. (Lecture Notes in Computer Science), p. 327–341. ISBN 978-3-642-41924-9. Disponível em: ⟨doi:10.1007/978-3-642-41924-9_27⟩.

HAENDEL, M.; GKOUTOS, G.; LEWIS, S.; MUNGALL, C. Uberon: Towards a comprehensive multi-species anatomy ontology. **Nature Precedings**, p. 1–1, 2009. ISSN 1756-0357. Disponível em: ⟨doi:10.1038/npre.2009.3592.1⟩.

HARDISTY, A.; ROBERTS, D.; The Biodiversity Informatics Community. A decadal view of biodiversity informatics: Challenges and priorities. **BMC Ecology**, v. 13, n. 1, p. 16, 2013. ISSN 1472-6785. Disponível em: ⟨doi:10.1186/1472-6785-13-16⟩.

HASKELL, E. F. A natural classification of societies. **Transactions of the New York Academy of Sciences**, v. 9, n. 5, p. 186–196, 1947. ISSN 0028-7113.

HASKELL, E. F. A clarification of social science. **Main currents in modern thought**, v. 7, n. 2, p. 45–51, 1949.

HEIDORN, P. B. Biodiversity informatics. **Bulletin of the American Society for Information Science and Technology**, v. 37, n. 6, p. 38–44, 2011. ISSN 1550-8366. Disponível em: ⟨doi:10.1002/bult.2011.1720370612⟩.

HERNÁNDEZ, C. X. P.; HERNÁNDEZ-ROBLES, D. R.; CORONA-LÓPEZ, A. M.; TOLEDO-HERNÁNDEZ, V. M.; DEL VAL, E. Dataset of the plant-buprestidae (coleoptera) interactions from mexico. 2021. Accessed via GBIF.org on 2023-05-doi:10. Disponível em: ⟨doi:10.15468/bck627⟩.

HERRANDO-PÉREZ, S.; BROOK, B. W.; BRADSHAW, C. J. A. Ecology Needs a Convention of Nomenclature. **BioScience**, v. 64, n. 4, p. 311–321, 2014. ISSN 0006-3568. Disponível em: ⟨doi:10.1093/biosci/biu013⟩.

HOEHNDORF, R.; ALSHAHRANI, M.; GKOUTOS, G. V.; GOSLINE, G.; GROOM, Q.; HAMANN, T.; KATTGE, J.; DE OLIVEIRA, S. M.; SCHMIDT, M.; SIERRA, S.; SMETS, E.; VOS, R. A.; WEILAND, C. The flora phenotype ontology (FLOPO): Tool for integrating morphological traits and phenotypes of vascular plants. **Journal of Biomedical Semantics**, v. 7, n. 1, p. 65, 2016. ISSN 2041-1480. Disponível em: ⟨doi:10.1186/s13326-016-0107-8⟩.

HOEKSEMA, J. D.; BRUNA, E. M. Context-dependent outcomes of mutualistic interactions. In: BRONSTEIN, J. L. (Ed.). **Mutualism**. United Kingdom: Oxford University Press, 2015. p. 0. ISBN 978-0-19-967565-4. Disponível em: ⟨doi:10.1093/acprof:oso/9780199675654.003.0010⟩.

HOEKSEMA, J. D.; CHAUDHARY, V. B.; GEHRING, C. A.; JOHNSON, N. C.; KARST, J.; KOIDE, R. T.; PRINGLE, A.; ZABINSKI, C.; BEVER, J. D.; MOORE, J. C.; WILSON, G. W. T.; KLIRONOMOS, J. N.; UMBANHOWAR, J. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. **Ecology Letters**, v. 13, n. 3, p. 394–407, 2010. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/j.1461-0248.2009.01430.x⟩.

HOOK, D. W.; PORTER, S. J.; HERZOG, C. Dimensions: Building context for search and evaluation. **Frontiers in Research Metrics and Analytics**, v. 3, 2018. ISSN 2504-0537.

HORTAL, J.; BELLO, F.; DINIZ-FILHO, J. A. F.; LEWINSOHN, T. M.; LOBO, J. M.; LADLE, R. J. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. **Annual Review of Ecology, Evolution, and Systematics**, v. 46, n. 1, p. 523–549, 2015. Disponível em: ⟨doi:10.1146/annurev-ecolsys-112414-054400⟩.

HOSODA, S.; NISHIJIMA, S.; FUKUNAGA, T.; HATTORI, M.; HAMADA, M. Revealing the microbial assemblage structure in the human gut microbiome using latent dirichlet allocation. **Microbiome**, v. 8, n. 1, p. 95, 2020. ISSN 2049-2618. Disponível em: ⟨doi:10.1186/s40168-020-00864-3⟩.

HOTHO, A.; NÜRNBERGER, A.; PAASS, G. A brief survey of text mining. **Journal for Language Technology and Computational Linguistics**, v. 20, n. 1, p. 19–62, 2005.

HUERTA, M.; DOWNING, G.; HASELTINE, F.; SETO, B.; LIU, Y. Nih working definition of bioinformatics and computational biology. **US National Institute of Health**, p. 1–1, 2000.

IMMELMANN, K. **Introduction to Ethology**. United States: Springer Science & Business Media, 2012. 242 p. ISBN 978-1-4684-1054-9.

IUCN. The IUCN red list of threatened species. version 2022-2. 2022. Downloaded on 2023-05-09. Disponível em: ⟨doi:10.15468/0qnb58⟩.

JOLLIFFE, I. Principal Component Analysis. In: **Wiley StatsRef: Statistics Reference Online**. John Wiley & Sons, Ltd, 2014. ISBN 978-1-118-44511-2. Disponível em: ⟨doi:10.1002/9781118445112.stat06472⟩.

JONES, M.; BERKLEY, C.; BOJILOVA, J.; HIGGINS, D. Metacat: A schema-independent xml database system. In: **Scientific and Statistical Database Management, International Conference on**. Los Alamitos, CA, USA: IEEE Computer Society, 2001. p. 0171. Disponível em: ⟨doi:10.1109/SSDM.2001.938549⟩.

JONES, M.; O'BRIEN, M.; MECUM, B.; BOETTIGER, C.; SCHILDHAUER, M.; MAIER, M.; WHITEAKER, T.; EARL, S.; CHONG, S. Ecological metadata language version 2.2.0. **KNB Data Repository**, 2019. Disponível em: ⟨doi:10.5063/f11834t2⟩.

JONES, M. B.; SCHILDHAUER, M. P.; REICHMAN, O.; BOWERS, S. The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. **Annual Review of Ecology, Evolution, and Systematics**, v. 37, n. 1, p. 519–544, 2006. Disponível em: ⟨doi:10.1146/annurev.ecolsys.37.091305.110031⟩.

JORDANO, P. The Biodiversity of Ecological Interactions: Challenges for recording and documenting the Web of Life. **Biodiversity Information Science and Standards**, v. 5, p. e75564, 2021. ISSN 2535-0897. Disponível em: ⟨doi:10.3897/biss.5.75564⟩.

JORDANO, P.; FORGET, P.-M.; LAMBERT, J. E.; BÖHNING-GAESE, K.; TRAVESET, A.; WRIGHT, S. J. Frugivores and seed dispersal: mechanisms and consequences for biodiversity of a key ecological interaction. **Biology Letters**, v. 7, n. 3, p. 321–323, 2010. Disponível em: ⟨doi:10.1098/rsbl.20doi:10.0986⟩.

KELLER, A.; ANKENBRAND, M. J.; BRUELHEIDE, H.; DEKEYZER, S.; ENQUIST, B. J.; ERFANIAN, M. B.; FALSTER, D. S.; GALLAGHER, R. V.; HAMMOCK, J.; KATTGE, J.; LEONHARDT, S. D.; MADIN, J. S.; MAITNER, B.; NEYRET, M.; ONSTEIN, R. E.; PEARSE, W. D.; POELEN, J. H.; SALGUERO-GOMEZ, R.; SCHNEIDER, F. D.; TÓTH, A. B.; PENONE, C. Ten (mostly) simple rules to future-proof trait data in ecological and evolutionary sciences. **Methods in Ecology and Evolution**, v. 14, n. 2, p. 444–458, 2023. ISSN 2041-210X. Disponível em: ⟨doi:10.1111/2041-210X.14033⟩.

KIERS, E. T.; PALMER, T. M.; IVES, A. R.; BRUNO, J. F.; BRONSTEIN, J. L. Mutualisms in a changing world: An evolutionary perspective. **Ecology Letters**, v. 13, n. 12, p. 1459–1474, 2010. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/j.1461-0248.20doi:10.01538.x⟩.

KIRK, P. M. World catalogue of 340 K fungal names on-line. **Mycological Research**, v. 104, n. 5, p. 516–517, 2000. ISSN 0953-7562. Disponível em: ⟨https://www.cabdirect.org/cabdirect/abstract/20198632168⟩.

KITA, C. A.; FLOREZ-MONTERO, G.; MONTOYA-BUSTAMANTE, S.; MUYLAERT, R. L.; ZAPATA-MESA, N.; MELLO, M. A. R. Ten simple rules for reporting information on species interactions. **PLOS Computational Biology**, v. 18, n. 8, p. e1010362, 2022. ISSN 1553-7358. Disponível em: ⟨doi:10.1371/journal.pcbi.1010362⟩.

KOCH, M.; DRILLER, C.; SCHMIDT, M.; HÖRNSCHEMEYER, T.; WEILAND, C.; MIKO, I.; YODER, M.; HICKLER, T. Current progress in the development of taxonomic and anatomical ontologies within the scope of BIOfid. **Biodiversity Information Science and Standards**, v. 2, p. e25585, 2018. ISSN 2535-0897. Disponível em: ⟨doi:10.3897/biss.2.25585⟩.

KONIETZNY, S. G.; DIETZ, L.; McHardy, A. C. Inferring functional modules of protein families with probabilistic topic models. **BMC Bioinformatics**, v. 12, n. 1, p. 141, 2011. ISSN 1471-2105. Disponível em: ⟨doi:10.1186/1471-2105-12-141⟩.

KURGAN, L. **Close Up at a Distance: Mapping, Technology, and Politics**. Brooklyn, NY: Zone Books, 2013. 232 p. ISBN 978-1-935408-28-4.

LAIGLE, I.; AUBIN, I.; DIGEL, C.; BROSE, U.; BOULANGEAT, I.; GRAVEL, D. Species traits as drivers of food web structure. **Oikos**, v. 127, n. 2, p. 316–326, 2018. ISSN 1600-0706. Disponível em: ⟨doi:10.1111/oik.04712⟩.

LANDI, P.; MINOARIVELO, H. O.; BRÄNNSTRÖM, A.; HUI, C.; DIECKMANN, U. Complexity and stability of ecological networks: A review of the theory. **Population Ecology**, v. 60, n. 4, p. 319–345, 2018. ISSN 1438-390X. Disponível em: ⟨doi:10.1007/s10144-018-0628-3⟩.

LAZIC, S. E.; CLARKE-WILLIAMS, C. J.; MUNAFÒ, M. R. What exactly is 'N' in cell culture and animal experiments? **PLOS Biology**, v. 16, n. 4, p. e2005282, 2018. ISSN 1545-7885. Disponível em: ⟨doi:10.1371/journal.pbio.2005282⟩.

LEARY, R. A. **Interaction Theory in Forest Ecology and Management**. Dordrecht: Springer Netherlands, 1985. v. 19. (Forestry Sciences, v. 19). ISBN 978-94-010-8779-7 978-94-009-5151-8. Disponível em: ⟨doi:10.1007/978-94-009-5151-8⟩.

LEBO, T.; SAHOO, S.; MCGUINNESS, D.; BELHAJJAME, K.; CHENEY, J.; CORSAR, D.; GARIJO, D.; SOILAND-REYES, S.; ZEDNIK, S.; ZHAO, J. Prov-o: The prov ontology. **W3C recommendation**, v. 30, 2013.

LEHRER, K. Knowledge, truth and evidence. **Analysis**, v. 25, n. 5, p. 168–175, 1965. ISSN 0003-2638. Disponível em: ⟨doi:10.2307/3326431⟩.

LIDICKER, W. Z. A Clarification of Interactions in Ecological Systems. **BioScience**, v. 29, n. 8, p. 475–477, 1979. ISSN 0006-3568. Disponível em: ⟨doi:10.2307/1307540⟩.

LIMA, S. L. Putting predators back into behavioral predator–prey interactions. **Trends in Ecology & Evolution**, v. 17, n. 2, p. 70–75, 2002. ISSN 0169-5347. Disponível em: ⟨doi:10.1016/S0169-5347(01)02393-X⟩.

LLOYD, S. Least square quantization in pcm. bell telephone laboratories paper. published in journal much later: Lloyd, sp: Least squares quantization in pcm. **IEEE Trans. Inform. Theor.(1957/1982)**, v. 18, n. 11, 1957.

LOMAKINA, L. S.; RODIONOV, V. B.; SURKOVA, A. S. Hierarchical clustering of text documents. **Automation and Remote Control**, v. 75, n. 7, p. 1309–1315, 2014. ISSN 1608-3032. Disponível em: ⟨doi:10.1134/S000511791407011X⟩.

LOREAU, M.; NAEEM, S.; INCHAUSTI, P.; BENGTSSON, J.; GRIME, J. P.; HECTOR, A.; HOOPER, D. U.; HUSTON, M. A.; RAFFAELLI, D.; SCHMID, B.; TILMAN, D.; WARDLE, D. A. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. **Science**, v. 294, n. 5543, p. 804–808, 2001. Disponível em: ⟨doi:10.1126/science.1064088⟩.

MACQUEEN, J. Classification and analysis of multivariate observations. In: UNIVERSITY OF CALIFORNIA LOS ANGELES LA USA. **5th Berkeley Symp. Math. Statist. Probability**. [S.l.], 1967. p. 281–297.

MARON, J. L.; BAER, K. C.; ANGERT, A. L. Disentangling the drivers of context-dependent plant–animal interactions. **Journal of Ecology**, v. 102, n. 6, p. 1485–1496, 2014. ISSN 1365-2745. Disponível em: ⟨doi:10.1111/1365-2745.12305⟩.

MARSH, G. P. **Man and Nature;or, Physical Geography as Modified by Human Action.** [S.l.]: London, S. Low, son and Marston, 1864. 588 p.

MARTIN, F.; JOHNSON, M. More efficient topic modelling through a noun only approach. In: **Proceedings of the Australasian Language Technology Association Workshop 2015**. Parramatta, Australia: [s.n.], 2015. p. 111–115. Disponível em: ⟨https://aclanthology.org/U15-1013⟩.

MATHIS, K. A.; BRONSTEIN, J. L. Our Current Understanding of Commensalism. **Annual Review of Ecology, Evolution, and Systematics**, v. 51, n. 1, p. 167–189, 2020. Disponível em: ⟨doi:10.1146/annurev-ecolsys-011720-040844⟩.

MATTSON, W. J.; ADDY, N. D. Phytophagous Insects as Regulators of Forest Primary Production. **Science**, American Association for the Advancement of Science, v. 190, n. 4214, p. 515–522. Disponível em: ⟨doi:10.1126/science.190.4214.515⟩.

MICHENER, W.; VIEGLAIS, D.; VISION, T.; KUNZE, J.; CRUSE, P.; JANÉE, G. DataONE: Data Observation Network for Earth Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. **D-Lib Magazine**, v. 17, n. 1/2, 2011. ISSN 1082-9873. Disponível em: ⟨doi:10.1045/january2011-michener⟩.

MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G. S.; DEAN, J. Distributed representations of words and phrases and their compositionality. In: **Advances in Neural Information Processing Systems**. United Kingdom: Curran Associates, Inc., 2013. v. 26.

MITTELBACH, G. G.; MCGILL, B. J. **Community Ecology**. United Kingdom: Oxford University Press, 2019. 430 p. ISBN 978-0-19-257286-8.

MONTERO-CASTAÑO, A.; VILÀ, M. Influence of the honeybee and trait similarity on the effect of a non-native plant on pollination and network rewiring. **Functional Ecology**, v. 31, n. 1, p. 142–152, 2017. Disponível em: ⟨doi:10.1111/1365-2435.12712⟩.

MORITZ, T.; KRISHNAN, S.; ROBERTS, D.; INGWERSEN, P.; AGOSTI, D.; PENEV, L.; COCKERILL, M.; CHAVAN, V. Towards mainstreaming of biodiversity data publishing: Recommendations of the GBIF Data Publishing Framework Task Group. **BMC Bioinformatics**, v. 12, n. 15, p. S1, 2011. ISSN 1471-2105. Disponível em: ⟨doi:doi:10.1186/1471-2105-12-S15-S1⟩.

MUKHIN, V.; VLADYKINA, V. Distribution of xylotrophic fungi of the genus Daedaleopsis in the Asian part of Russia from 1978 to 2019. 2020. Accessed via GBIF.org on 2023-05-10. Disponível em: ⟨doi:10.15468/m4hk49⟩.

MUNGALL, C.; MATENTZOGLU, N.; BALHOFF, J.; OSUMI-SUTHERLAND, D.; pgaudet; TAN, S.; DUNCAN, B.; PILGRIM, C.; OVERTON, J. A.; HOYT, C. T.; LAUREN; HARRIS, N.; MOXON, S.; lschriml; VASILEVSKY, N.; sabrinatoro; BRUSH, M.; TOURÉ, V.; CARON, A.; GOUTTE-GATTAT, D.; SINCLAIR, M.; BRETAUDEAU, A.; CAIN, S.; HAENDEL, M.; diatomsRcool; ZHANG, B.; HAMMOCK, J.; LAPORTE, M.-A. **OBO Relation Ontology**. 2023. Disponível em: ⟨doi:10.5281/zenodo.7665156⟩.

MUNGALL, C. J.; TORNIAI, C.; GKOUTOS, G. V.; LEWIS, S. E.; HAENDEL, M. A. Uberon, an integrative multi-species anatomy ontology. **Genome Biology**, v. 13, n. 1, p. R5, 2012. ISSN 1474-760X. Disponível em: ⟨doi:10.1186/gb-2012-13-1-r5⟩.

MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchical clustering: an overview. **WIREs Data Mining and Knowledge Discovery**, v. 2, n. 1, p. 86–97, 2012. ISSN 1942-4795. Disponível em: ⟨doi:10.1002/widm.53⟩.

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, v. 18, n. 5, p. 544–551, 2011. ISSN 1067-5027. Disponível em: ⟨doi:10.1136/amiajnl-2011-000464⟩.

NAKAZAWA, T. Species interaction: Revisiting its terminology and concept. **Ecological Research**, v. 35, n. 6, p. 1106–1113, 2020. ISSN 1440-1703. Disponível em: ⟨doi:10.1111/1440-1703.12164⟩.

National Museum of Natural History, Smithsonian Institution. Integrated Taxonomic Information System (ITIS). 2023. Disponível em: ⟨doi:doi:10.5066/f7kh0kbk⟩.

Nature. Announcement: Where are the data? **Nature**, v. 537, n. 7619, p. 138–138, 2016. ISSN 1476-4687. Disponível em: ⟨doi:10.1038/537138a⟩.

NILSSON, R. H.; ANDERSSON, A. F.; BISSETT, A.; FINSTAD, A. G.; FOSSøY, F.; GROSJEAN, M.; HOPE, M.; JEPPESEN, T. S.; KÕLJALG, U.; LUNDIN, D.; PRAGER, M.; SUOMINEN, S.; SVENNINGSEN, C. S.; SCHIGEL, D. Introducing guidelines for publishing DNA-derived occurrence data through biodiversity data platforms. **Metabarcoding and Metagenomics**, v. 6, p. e84960, 2022. Disponível em: ⟨doi:10.3897/mbmg.6.84960⟩.

NOSEK, B. A.; ALTER, G.; BANKS, G. C.; BORSBOOM, D.; BOWMAN, S. D.; BRECKLER, S. J.; BUCK, S.; CHAMBERS, C. D.; CHIN, G.; CHRISTENSEN, G.; CONTESTABILE, M.; DAFOE, A.; EICH, E.; FREESE, J.; GLENNERSTER, R.; GOROFF, D.; GREEN, D. P.; HESSE, B.; HUMPHREYS, M.; ISHIYAMA, J.; KARLAN, D.; KRAUT, A.; LUPIA, A.; MABRY, P.; MADON, T.; MALHOTRA, N.; MAYO-WILSON, E.; MCNUTT, M.; MIGUEL, E.; PALUCK, E. L.; SIMONSOHN, U.; SODERBERG, C.; SPELLMAN, B. A.; TURITTO, J.; VANDENBOS, G.; VAZIRE, S.; WAGENMAKERS, E. J.; WILSON, R.; YARKONI, T. Promoting an open research culture. **Science**, v. 348, n. 6242, p. 1422–1425, 2015. Disponível em: ⟨doi:10.1126/science.aab2374⟩.

NOY, N. F.; SHAH, N. H.; WHETZEL, P. L.; DAI, B.; DORF, M.; GRIFFITH, N.; JONQUET, C.; RUBIN, D. L.; STOREY, M.-A.; CHUTE, C. G.; MUSEN, M. A. BioPortal: Ontologies and integrated data resources at the click of a mouse. **Nucleic Acids Research**, v. 37, p. W170–W173, 2009. ISSN 0305-1048. Disponível em: ⟨doi:10.1093/nar/gkp440⟩.

OAKES, M. P.; JI, M. **Quantitative Methods in Corpus-Based Translation Studies: A practical guide to descriptive translation research**. [S.l.]: John Benjamins Publishing, 2012. 373 p. ISBN 978-90-272-7478-6.

ODUM, E. P. **Fundamentals of Ecology**. 2. ed. Philadelphia: Saunders, 1959. 546 p.

OLITO, C.; FOX, J. W. Species traits and abundances predict metrics of plant-pollinator network structure, but not pairwise interactions. **Oikos**, v. 124, n. 4, p. 428–436, 2015. ISSN 00301299. Disponível em: ⟨doi:10.1111/oik.01439⟩.

OLIVAL, K. J.; HOSSEINI, P. R.; ZAMBRANA-TORRELIO, C.; ROSS, N.; BOGICH, T. L.; DASZAK, P. Host and viral traits predict zoonotic spillover from mammals. **Nature**, v. 546, n. 7660, p. 646–650, 2017. ISSN 1476-4687. Disponível em: ⟨doi:10.1038/nature22975⟩.

PAGE, R. Towards a biodiversity knowledge graph. **Research Ideas and Outcomes**, v. 2, p. e8767, 2016. ISSN 2367-7163. Disponível em: ⟨doi:doi:10.3897/rio.2.e8767⟩.

PARK, C. A.; BELLO, S. M.; SMITH, C. L.; HU, Z.-L.; MUNZENMAIER, D. H.; NIGAM, R.; SMITH, J. R.; SHIMOYAMA, M.; EPPIG, J. T.; REECY, J. M. The Vertebrate Trait Ontology: A controlled vocabulary for the annotation of trait data across species. **Journal of Biomedical Semantics**, v. 4, n. 1, p. 13, 2013. ISSN 2041-1480. Disponível em: ⟨doi:10.1186/2041-1480-4-13⟩.

PEARSON, K. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 2, n. 11, p. 559–572, 1901. ISSN 1941-5982. Disponível em: ⟨doi:10.1080/14786440109462720⟩.

PEARSON, K.; ELLWOOD, L.; GILBERT, E.; GURALNICK, R.; MACKLIN, J.; NELSON, G.; SWEENEY, P.; STUCKY, B.; WIECZOREK, J.; YOST, J. Data Standards for the Phenology of Plant Specimens. **Biodiversity Information Science and Standards**, v. 5, p. e74372, 2021. Disponível em: ⟨doi:10.3897/biss.5.74372⟩.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PENEV, L.; DIMITROVA, M.; SENDEROV, V.; ZHELEZOV, G.; GEORGIEV, T.; STOEV, P.; SIMOV, K. OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science. **Publications**, v. 7, n. 2, p. 38, 2019. ISSN 2304-6775. Disponível em: ⟨doi:doi:10.3390/publications7020038⟩.

Plinian Core Task Group. **Plinian Core, a Species-level Data Specification**. Biodiversity Information Standards (TDWG), 2021. Disponível em: ⟨https://github.com/tdwg/PlinianCore⟩.

POELEN, J. H.; SIMONS, J. D.; MUNGALL, C. J. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. **Ecological Informatics**, v. 24, p. 148–159, 2014. ISSN 1574-9541. Disponível em: ⟨doi:10.1016/j.ecoinf.2014.08.005⟩.

POISOT, T.; BAISER, B.; DUNNE, J. A.; KÉFI, S.; MASSOL, F.; MOUQUET, N.; ROMANUK, T. N.; STOUFFER, D. B.; WOOD, S. A.; GRAVEL, D. Mangal – making ecological network analysis simple. **Ecography**, v. 39, n. 4, p. 384–390, 2016. ISSN 1600-0587. Disponível em: ⟨doi:10.1111/ecog.00976⟩.

POISOT, T.; STOUFFER, D. B.; GRAVEL, D. Beyond species: Why ecological interaction networks vary through space and time. **Oikos**, v. 124, n. 3, p. 243–251. Disponível em: ⟨doi:10.1111/oik.01719⟩.

POMERANTZ, J. **Metadata**. United Kingdom: MIT Press, 2015. 252 p. ISBN 978-0-262-33120-3.

POOTER, D. D.; APPELTANS, W.; BAILLY, N.; BRISTOL, S.; DENEUDT, K.; ELIEZER, M.; FUJIOKA, E.; GIORGETTI, A.; GOLDSTEIN, P.; LEWIS, M.; LIPIZER, M.; MACKAY, K.; MARIN, M.; MONCOIFFÉ, G.; NIKOLOPOULOU, S.; PROVOOST, P.; RAUCH, S.; ROUBICEK, A.; TORRES, C.; VAN DE PUTTE, A.; VANDEPITTE, L.; VANHOORNE, B.; VINCI, M.; WAMBIJI, N.; WATTS, D.; SALAS, E. K.; HERNANDEZ, F. Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. **Biodiversity Data Journal**, v. 5, p. e10989, 2017. ISSN 1314-2828. Disponível em: ⟨doi:doi:10.3897/BDJ.5.e10989⟩.

PRATANWANICH, N.; LIO, P. Exploring the complexity of pathway–drug relationships using latent dirichlet allocation. **Computational Biology and Chemistry**, v. 53, p. 144–152, 2014. ISSN 1476-9271. Disponível em: ⟨doi: 10.1016/j.compbiolchem.2014.08.019⟩.

PRICE, P. W. **Insect Ecology**. United Kingdom: Wiley, 1984. 632 p. ISBN 978-0-471-07892-0.

PRINGLE, E. G. Orienting the Interaction Compass: Resource Availability as a Major Driver of Context Dependence. **PLOS Biology**, v. 14, n. 10, p. e2000891, 2016. ISSN 1545-7885. Disponível em: ⟨doi:10.1371/journal.pbio.2000891⟩.

QIAN, H.; ZHANG, J.; JIANG, M.-C. Global patterns of fern species diversity: An evaluation of fern data in GBIF. **Plant Diversity**, v. 44, n. 2, p. 135–140, 2022. ISSN 2468-2659. Disponível em: ⟨doi:10.1016/j.pld.2021.doi:10.001⟩.

Queensland Department of Environment and Science. BRI AVH data. Occurrence dataset. 2023. Accessed via GBIF.org on 2023-05-10. Disponível em: ⟨doi:10.15468/jsffsa⟩.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: ⟨https://www.R-project.org/⟩.

RAYNER, T. F.; ROCCA-SERRA, P.; SPELLMAN, P. T.; CAUSTON, H. C.; FARNE, A.; HOLLOWAY, E.; IRIZARRY, R. A.; LIU, J.; MAIER, D. S.; MILLER, M.; PETERSEN, K.; QUACKENBUSH, J.; SHERLOCK, G.; STOECKERT, C. J.; WHITE, J.; WHETZEL, P. L.; WYMORE, F.; PARKINSON, H.; SARKANS, U.; BALL, C. A.; BRAZMA, A. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. **BMC bioinformatics**, v. 7, p. 489, 2006. ISSN 1471-2105. Disponível em: ⟨doi:10.1186/1471-2105-7-489⟩.

REES, J.; CRANSTON, K. Automated assembly of a reference taxonomy for phylogenetic data synthesis. **Biodiversity Data Journal**, v. 5, p. e12581, 2017. ISSN 1314-2828. Disponível em: ⟨doi:doi:10.3897/BDJ.5.e12581⟩.

REHUREK, R.; SOJKA, P. Software Framework for Topic Modelling with Large Corpora. In: **Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks**. Valletta, Malta: ELRA, 2010. p. 45–50.

ROBERTS, D. CHAS Entomology Collection (Arctos). 2023. Accessed via GBIF.org on 2023-05-10. Disponível em: ⟨doi:10.15468/i5oupp⟩.

ROBERTSON, T.; DÖRING, M.; GURALNICK, R.; BLOOM, D.; WIECZOREK, J.; BRAAK, K.; OTEGUI, J.; RUSSELL, L.; DESMET, P. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. **PLOS ONE**, v. 9, n. 8, p. e102623, 2014. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0102623⟩.

ROCCA-SERRA, P.; SALEK, R. M.; ARITA, M.; CORREA, E.; DAYALAN, S.; GONZALEZ-BELTRAN, A.; EBBELS, T.; GOODACRE, R.; HASTINGS, J.; HAUG, K.; KOULMAN, A.; NIKOLSKI, M.; ORESIC, M.; SANSONE, S.-A.; SCHOBER, D.; SMITH, J.; STEINBECK, C.; VIANT, M. R.; NEUMANN, S. Data standards can boost metabolomics research, and if there is a will, there is a way. **Metabolomics**, v. 12, n. 1, p. 14, 2015. ISSN 1573-3890. Disponível em: ⟨doi:10.1007/s11306-015-0879-3⟩.

ROCHA-ORTEGA, M.; RODRIGUEZ, P.; CÓRDOBA-AGUILAR, A. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. **Ecological Entomology**, v. 46, n. 4, p. 718–728, 2021. ISSN 1365-2311. Disponível em: ⟨doi:10.1111/een.13027⟩.

RÖDER, M.; BOTH, A.; HINNEBURG, A. Exploring the space of topic coherence measures. In: **Proceedings of the Eighth ACM International Conference on Web Search and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2015. (WSDM '15), p. 399–408. ISBN 978-1-4503-3317-7. Disponível em: ⟨doi:10.1145/2684822.2685324⟩.

ROGERS, S.; GIROLAMI, M.; CAMPBELL, C.; BREITLING, R. The latent process decomposition of cDNA microarray data sets. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 2, n. 2, p. 143–156, 2005. ISSN 1557-9964. Disponível em: ⟨doi:10.1109/TCBB.2005.29⟩.

ROUSH, S. **Tracking Truth: Knowledge, Evidence, and Science**. United Kingdom: Clarendon Press, 2005. 248 p. ISBN 978-0-19-153448-5.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987. ISSN 0377-0427. Disponível em: ⟨doi:10.1016/0377-0427(87)90125-7⟩.

RUETE, A. Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. **Biodiversity Data Journal**, n. 3, p. e5361, 2015. ISSN 1314-2836. Disponível em: ⟨doi:10.3897/BDJ.3.e5361⟩.

RUMEU, B.; SHEATH, D. J.; HAWES, J. E.; INGS, T. C. Zooming into plant-flower visitor networks: An individual trait-based approach. **PeerJ**, v. 6, p. e5618, 2018. ISSN 2167-8359. Disponível em: ⟨doi:10.7717/peerj.5618⟩.

SACHS, J. L.; WILCOX, T. P. A shift to parasitism in the jellyfish symbiont Symbiodinium microadriaticum. **Proceedings. Biological Sciences**, v. 273, n. 1585, p. 425–429, 2006. ISSN 0962-8452. Disponível em: ⟨doi:10.1098/rspb.2005.3346⟩.

SALIM, J. A. zedomel/thesis_2023: Initial release. **Zenodo**, 2023. Disponível em: ⟨doi:10.5281/zenodo.8106654⟩.

SALIM, J. A.; SARAIVA, A. A Google Sheet Add-on for Biodiversity Data Standardization and Sharing. **Biodiversity Information Science and Standards**, v. 4, p. e59228, 2020. ISSN 2535-0897. Disponível em: ⟨doi:10.3897/biss.4.59228⟩.

SALIM, J. A.; SARAIVA, A. M.; ZERMOGLIO, P. F.; AGOSTINI, K.; WOLOWSKI, M.; DRUCKER, D. P.; SOARES, F. M.; BERGAMO, P. J.; VARASSIN, I. G.; FREITAS, L.; MAUÉS, M. M.; RECH, A. R.; VEIGA, A. K.; ACOSTA, A. L.; ARAUJO, A. C.; NOGUEIRA, A.; BLOCHTEIN, B.; FREITAS, B. M.; ALBERTINI, B. C.; MAIA-SILVA, C.; NUNES, C. E. P.; PIRES, C. S. S.; DOS SANTOS, C. F.; QUEIROZ, E. P.; CARTOLANO, E. A.; DE OLIVEIRA, F. F.; AMORIM, F. W.; FONTÚRBEL, F. E.; DA SILVA, G. V.; CONSOLARO, H.; ALVES-DOS-SANTOS, I.; MACHADO, I. C.; SILVA, J. S.; ALEIXO, K. P.; CARVALHEIRO, L. G.; ROCCA, M. A.; PINHEIRO, M.; HRNCIR, M.; STREHER, N. S.; FERREIRA, P. A.; DE ALBUQUERQUE, P. M. C.; MARUYAMA, P. K.; BORGES, R. C.; GIANNINI, T. C.; BRITO, V. L. G. Data standardization of plant–pollinator interactions. **GigaScience**, v. 11, p. giac043, 2022. ISSN 2047-217X. Disponível em: ⟨doi:10.1093/gigascience/giac043⟩.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513–523, 1988. ISSN 0306-4573. Disponível em: ⟨doi:10.1016/0306-4573(88)90021-0⟩.

SANSONE, S.-A.; MCQUILTON, P.; ROCCA-SERRA, P.; GONZALEZ-BELTRAN, A.; IZZO, M.; LISTER, A. L.; THURSTON, M. FAIRsharing as a community approach to standards, repositories and policies. **Nature Biotechnology**, v. 37, n. 4, p. 358–367, 2019. ISSN 1546-1696. Disponível em: ⟨doi:10.1038/s41587-019-0080-8⟩.

SARAIVA, A. M.; GEMMILL-HERREN, B.; RUGGIERO, M. A COMMON SCHEMA FOR MANAGING PLANT-POLLINATOR INTERACTION DATA. **United Nations Environment Programme (UNEP)**, p. 33–33, 2010.

SARAIVA, A. M.; JUNIOR, E. A. C.; GIOVANNI, R.; GIANNINI, T. C.; CORREA, P. L. P. Exchanging specimen interaction data using Darwin Core. In: **The Proceedings of TDWG 2009**. Biodiversity Information Standards (TDWG), 2009. Disponível em: ⟨https://static.tdwg.org/conferences/2009/tdwg_2009_pre-proceedings.pdf⟩.

SCHEMSKE, D. W.; MITTELBACH, G. G.; CORNELL, H. V.; SOBEL, J. M.; ROY, K. Is There a Latitudinal Gradient in the Importance of Biotic Interactions? **Annual Review of Ecology, Evolution, and Systematics**, v. 40, n. 1, p. 245–269, 2009. Disponível em: ⟨doi:10.1146/annurev.ecolsys.39.110707.173430⟩.

SCHNEIDER, F. D.; FICHTMUELLER, D.; GOSSNER, M. M.; GÜNTSCH, A.; JOCHUM, M.; KÖNIG-RIES, B.; LE PROVOST, G.; MANNING, P.; OSTROWSKI, A.; PENONE, C.; SIMONS, N. K. Towards an ecological trait-data standard. **Methods in Ecology and Evolution**, v. 10, n. 12, p. 2006–2019, 2019. ISSN 2041-210X. Disponível em: ⟨doi:doi:10.1111/2041-210X.13288⟩.

SCHUTZE, H.; MANNING, C. D.; RAGHAVAN, P. **Introduction to information retrieval**. Italy: Cambridge University Press, 2008. 482 p.

SEBASTIÁN-GONZÁLEZ, E.; PIRES, M. M.; DONATTI, C. I.; GUIMARÃES, P. R.; DIRZO, R. Species traits and interaction rules shape a species-rich seed-dispersal interaction network. **Ecology and Evolution**, v. 7, n. 12, p. 4496–4506, 2017. Disponível em: ⟨doi:10.1002/ece3.2865⟩.

SHIVASHANKAR, S.; SRIVATHSAN, S.; RAVINDRAN, B.; TENDULKAR, A. V. Multi-view methods for protein structure comparison using latent dirichlet allocation. **Bioinformatics**, v. 27, n. 13, p. i61–i68, 2011. ISSN 1367-4803. Disponível em: ⟨doi:10.1093/bioinformatics/btr249⟩.

SHUKLA, S.; NAGANNA, S. A review on k-means data clustering approach. **International Journal of Information & Computation Technology**, v. 4, n. 17, p. 1847–1860, 2014.

SIELEMANN, K.; HAFNER, A.; PUCKER, B. The reuse of public datasets in the life sciences: Potential risks and rewards. **PeerJ**, v. 8, p. e9954, 2020. ISSN 2167-8359. Disponível em: ⟨doi:10.7717/peerj.9954⟩.

SLETVOLD, N.; TYE, M.; AGREN, J. Resource- and pollinator-mediated selection on floral traits. **Functional Ecology**, v. 31, n. 1, p. 135–141. ISSN 1365-2435. Disponível em: ⟨doi:10.1111/1365-2435.12757⟩.

SMITH, C. L.; EPPIG, J. T. The mammalian phenotype ontology: Enabling robust annotation and comparative analysis. **WIREs Systems Biology and Medicine**, v. 1, n. 3, p. 390–399, 2009. ISSN 1939-005X. Disponível em: ⟨doi:10.1002/wsbm.44⟩.

SOBERÓN, J.; PETERSON, A. T. Biodiversity informatics: Managing and applying primary biodiversity data. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 359, n. 1444, p. 689–698, 2004. ISSN 0962-8436.

STUCKY, B. J.; GURALNICK, R.; DECK, J.; DENNY, E. G.; BOLMGREN, K.; WALLS, R. The Plant Phenology Ontology: A New Informatics Resource for Large-Scale Integration of Plant Phenology Data. **Frontiers in Plant Science**, v. 9, 2018. ISSN 1664-462X. Disponível em: ⟨doi:10.3389/fpls.2018.00517⟩.

TDWG Ontology. Ontology/ontology/voc/TaxonOccurrenceInteraction.rdf at master · tdwg/ontology · GitHub. **GitHub**, 2015. Disponível em: ⟨https://github.com/tdwg/ontology/blob/master/ontology/voc/TaxonOccurrenceInteraction.rdf⟩.

TDWG Vocabulary Management Task Group. **Report of the TDWG Vocabulary Management Task Group (VoMaG)**. TDWG, 2013. 25 p. Disponível em: ⟨https://www.gbif.org/document/80862⟩.

TEMPLETON, A. R.; GILBERT, L. E. Population genetics and the coevolution of mutualism. In: BOUCHER, D. H. (Ed.). **The Biology of Mutualism: Ecology and Evolution**. New York: Oxford University Press, 1985. p. 128–144. ISBN 978-0-19-505392-0.

THANOS, C. Mediation: The Technological Foundation of Modern Science. **Ubiquity Press**, v. 13, n. 0, p. 88, 2014. ISSN 1683-1470. Disponível em: ⟨doi:10.2481/dsj.14-016⟩.

THANOS, C. Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies. **Publications**, v. 5, n. 1, p. 2, 2017. ISSN 2304-6775. Disponível em: ⟨doi:10.3390/publications5010002⟩.

The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. **Nucleic Acids Research**, v. 47, n. D1, p. D330–D338, 2019. ISSN 0305-1048. Disponível em: ⟨doi:10.1093/nar/gky1055⟩.

THOMPSON, J. N. The Evolution of Species Interactions. **Science**, v. 284, n. 5423, p. 2116–2118, 1999. Disponível em: ⟨doi:10.1126/science.284.5423.2116⟩.

THOMPSON, J. N. **The Geographic Mosaic of Coevolution**. United Kingdom: University of Chicago Press, 2005. ISBN 978-0-226-11869-7. Disponível em: ⟨doi:10.7208/9780226118697⟩.

THOMPSON, M. Ontological Shift or Ontological Drift? Reality Claims, Epistemological Frameworks, and Theory Generation in Organization Studies. **The Academy of Management Review**, v. 36, n. 4, p. 754–773, 2011. ISSN 0363-7425. Disponível em: ⟨https://www.jstor.org/stable/41318094⟩.

THOMPSON, R. M.; BROSE, U.; DUNNE, J. A.; HALL, R. O.; HLADYZ, S.; KITCHING, R. L.; MARTINEZ, N. D.; RANTALA, H.; ROMANUK, T. N.; STOUFFER, D. B.; TYLIANAKIS, J. M. Food webs: Reconciling the structure and function of biodiversity. **Trends in Ecology & Evolution**, v. 27, n. 12, p. 689–697, 2012. ISSN 0169-5347. Disponível em: ⟨doi:10.1016/j.tree.2012.08.005⟩.

TREMBLAY, M. S.; AUBERT, S.; BARNES, J. D.; SAUNDERS, T. J.; CARSON, V.; LATIMER-CHEUNG, A. E.; CHASTIN, S. F.; ALTENBURG, T. M.; CHINAPAW, M. J.; ALTENBURG, T. M.; AMINIAN, S.; ARUNDELL, L.; ATKIN, A. J.; AUBERT, S.; BARNES, J.; GIBBS, B. B.; BASSETT-GUNTER, R.; BELANGER, K.; BIDDLE, S.; BISWAS, A.; CARSON, V.; CHAPUT, J.-P.; CHASTIN, S.; CHAU, J.; CHINAPAW, M.; COLLEY, R.; COPPINGER, T.; CRAVEN, C.; CRISTI-MONTERO, C.; DE ASSIS TELES SANTOS, D.; DEL POZO CRUZ, B.; CRUZ, J. del Pozo; DEMPSEY, P.; DO CARMO SANTOS GONÇALVES, R. F.; EKELUND, U.; ELLINGSON, L.; EZEUGWU, V.; FITZSIMONS, C.; FLOREZ-PREGONERO, A.; FRIEL, C. P.; FRÖBERG, A.; GIANGREGORIO, L.; GODIN, L.; GUNNELL, K.; HALLOWAY, S.; HINKLEY, T.; HNATIUK, J.; HUSU, P.; KADIR, M.; KARAGOUNIS, L. G.; KOSTER, A.; LAKERVELD, J.; LAMB, M.; LAROUCHE, R.; LATIMER-CHEUNG, A.; LEBLANC, A. G.; LEE, E.-Y.; LEE, P.; LOPES, L.; MANNS, T.; MANYANGA, T.; GINIS, K. M.; MCVEIGH, J.; MENEGUCI, J.; MOREIRA, C.; MURTAGH, E.; PATTERSON, F.; SILVA, D. Rodrigues Pereira da; PESOLA, A. J.; PETERSON, N.; PETTITT, C.; PILUTTI, L.; PEREIRA, S. P.; POITRAS, V.; PRINCE, S.; RATHOD, A.; RIVIÈRE, F.; ROSENKRANZ, S.; ROUTHIER, F.; SANTOS, R.; SAUNDERS, T.; SMITH, B.; THEOU, O.; TOMASONE, J.; TREMBLAY, M.; TUCKER, P.; MEYER, R. U.; VAN DER PLOEG, H.; VILLALOBOS, T.; VIREN, T.; WALLMANN-SPERLICH, B.; WIJNDAELE, K.; WONDERGEM, R.; on behalf of SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. **International Journal of Behavioral Nutrition and Physical Activity**, v. 14, n. 1, p. 75, 2017. ISSN 1479-5868. Disponível em: ⟨doi:doi:10.1186/s12966-017-0525-8⟩.

TROIA, M. J.; McManamay, R. A. Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the united states. **Ecology and Evolution**, v. 6, n. 14, p. 4654–4669, 2016. ISSN 2045-7758. Disponível em: ⟨doi:10.1002/ece3.2225⟩.

TROUDET, J.; GRANDCOLAS, P.; BLIN, A.; VIGNES-LEBBE, R.; LEGENDRE, F. Taxonomic bias in biodiversity data and societal preferences. **Scientific Reports**, v. 7, n. 1, p. 9132, 2017. ISSN 2045-2322. Disponível em: ⟨doi:10.1038/s41598-017-09084-6⟩.

TYLIANAKIS, J. M.; DIDHAM, R. K.; BASCOMPTE, J.; WARDLE, D. A. Global change and species interactions in terrestrial ecosystems. **Ecology Letters**, v. 11, n. 12, p. 1351–1363, 2008. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/j.1461-0248.2008.01250.x⟩.

VALLE, D.; ALBUQUERQUE, P.; ZHAO, Q.; BARBERAN, A.; FLETCHER JR., R. J. Extending the latent dirichlet allocation model to presence/absence data: A case study on north american breeding birds and biogeographical shifts expected from climate change. **Global Change Biology**, v. 24, n. 11, p. 5560–5572, 2018. ISSN 1365-2486. Disponível em: ⟨doi:10.1111/gcb.14412⟩.

VALLE, D.; BAISER, B.; WOODALL, C. W.; CHAZDON, R. Decomposing biodiversity data using the latent dirichlet allocation model, a probabilistic multivariate statistical method. **Ecology Letters**, v. 17, n. 12, p. 1591–1601, 2014. ISSN 1461-0248. Disponível em: ⟨doi:10.1111/ele.12380⟩.

VAZQUEZ, D.; MELIAN, C. **Interaction web database**. 2008. Disponível em: ⟨https://www.nceas.ucsb.edu/interactionweb/index.html⟩.

VEIGA, A. K.; SARAIVA, A. M.; CHAPMAN, A. D.; MORRIS, P. J.; GENDREAU, C.; SCHIGEL, D.; ROBERTSON, T. J. A conceptual framework for quality assessment and management of biodiversity data. **PLOS ONE**, v. 12, n. 6, p. e0178731, 2017. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0178731⟩.

VELLEND, M. **The Theory of Ecological Communities (MPB-57)**. United States: Princeton University Press, 2016. ISBN 978-1-4008-8379-0. Disponível em: ⟨doi:10.1515/9781400883790⟩.

VIZENTIN-BUGONI, J.; MARUYAMA, P. K.; SOUZA, C. S.; OLLERTON, J.; RECH, A. R.; SAZIMA, M. Plant-Pollinator Networks in the Tropics: A Review. In: DÁTTILO, W.; RICO-GRAY, V. (Ed.). **Ecological Networks in the Tropics: An Integrative Overview of Species Interactions from Some of the Most Species-Rich Habitats on Earth**. Cham: Springer International Publishing, 2018. p. 73–91. ISBN 978-3-319-68228-0. Disponível em: ⟨doi:10.1007/978-3-319-68228-0_6⟩.

Vocabulary Maintenance Specification Task Group. Standards Documentation Standard. **Biodiversity Information Standards (TDWG)**, 2017. Disponível em: ⟨http://www.tdwg.org/standards/147⟩.

Vocabulary Maintenance Specification Task Group. Vocabulary Maintenance Standard. **Biodiversity Information Standards (TDWG)**, 2017. Disponível em: ⟨http://www.tdwg.org/standards/642⟩.

WANG, S.; SCHLOBACH, S.; KLEIN, M. Concept drift and how to identify it. **Journal of Web Semantics**, v. 9, n. 3, p. 247–265, 2011. ISSN 1570-8268. Disponível em: ⟨doi:10.1016/j.websem.2011.05.003⟩.

WATTS, S.; DORMANN, C. F.; GONZÁLEZ, A. M. M.; OLLERTON, J. The influence of floral traits on specialization and modularity of plant–pollinator networks in a biodiversity hotspot in the Peruvian Andes. **Annals of Botany**, v. 118, n. 3, p. 415–429, 2016. Disponível em: ⟨doi:10.1093/aob/mcw114⟩.

WEISS, C.; KARRAS, P.; BERNSTEIN, A. Hexastore: sextuple indexing for semantic web data management. **Proceedings of the VLDB Endowment**, v. 1, n. 1, p. 1008–1019, 2008. ISSN 2150-8097. Disponível em: ⟨doi:10.14778/1453856.1453965⟩.

WELBERS, W. v. A.; KASPER, C. J. Quantitative analysis of large amounts of journalistic texts using topic modelling. In: SJØVAAG, M. K. (Ed.). **Rethinking Research Methods in an Age of Digital Journalism**. United Kingdom: Routledge, Taylor & Francis Group, 2019. ISBN 978-1-315-11504-7.

WERNER, E. E.; PEACOR, S. D. A Review of Trait-Mediated Indirect Interactions in Ecological Communities. **Ecology**, v. 84, n. 5, p. 1083–1100, 2003. ISSN 1939-9170. Disponível em: ⟨doi:10.1890/0012-9658(2003)084[1083:AROTII]2.0.CO;2⟩.

WESTGATE, M. J.; BARTON, P. S.; PIERSON, J. C.; LINDENMAYER, D. B. Text analysis tools for identification of emerging topics and research gaps in conservation science. **Conservation Biology**, v. 29, n. 6, p. 1606–1614, 2015. ISSN 1523-1739. Disponível em: ⟨doi:10.1111/cobi.12605⟩.

WFO. World Flora Online. **Published on the Internet**, 2023. Disponível em: ⟨http://www.worldfloraonline.org⟩.

WHITE, E. P.; BALDRIDGE, E.; BRYM, Z. T.; LOCEY, K. J.; MCGLINN, D. J.; SUPP, S. R. Nine simple ways to make it easier to (re)use your data. **Ideas in Ecology and Evolution**, v. 6, n. 2, 2013. ISSN 1918-3178. Disponível em: ⟨https://ojs.library.queensu.ca/index.php/IEE/article/view/4608⟩.

WIECZOREK, J.; BÁNKI, O.; BLUM, S.; DECK, J.; DÖRING, M.; DRÖGE, G.; ENDRESEN, D.; GOLDSTEIN, P.; LEARY, P.; KRISHTALKA, L.; TUAMA, e.; ROBBINS, R. J.; ROBERTSON, T.; YILMAZ, P. Meeting Report: GBIF hackathon-workshop on Darwin Core and sample data (22–24 May 2013). **Standards in Genomic Sciences**, v. 9, n. 3, p. 585–598, 2014. ISSN 1944-3277. Disponível em: ⟨doi:doi:10.4056/sigs.4898640⟩.

WIECZOREK, J.; BLOOM, D.; GURALNICK, R.; BLUM, S.; DÖRING, M.; GIOVANNI, R.; ROBERTSON, T.; VIEGLAIS, D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. **PLOS ONE**, v. 7, n. 1, p. e29715, 2012. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0029715⟩.

WIKIPEDIA. Key–value database. **Wikipedia**, 2023. Disponível em: ⟨https://en.wikipedia.org/w/index.php?title=Key%E2%80%93value_database&oldid=1135560734⟩.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; DA SILVA SANTOS, L. B.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; VAN SCHAIK, R.; SANSONE, S.-A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VANMULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, p. 160018, 2016. ISSN 2052-4463. Disponível em: ⟨doi:10.1038/sdata.2016.18⟩.

WOLOWSKI, M.; AGOSTINI, K.; RECH, A. R.; VARASSIN, I. G.; MAUÉS, M.; FREITAS, L.; CARNEIRO, L. T.; BUENO, R. D. O.; CONSOLARO, H.; CARVALHEIRO, L.; SARAIVA, A. M.; SILVA, C. I. D. **Relatório temático sobre polinização, polinizadores e produção de alimentos no Brasil**. 1. ed. Editora Cubo, 2019. ISBN 978-85-60064-83-0. Disponível em: ⟨doi:doi:10.4322/978-85-60064-83-0⟩.

WOOTTON, J. T.; EMMERSON, M. Measurement of Interaction Strength in Nature. **Annual Review of Ecology, Evolution, and Systematics**, v. 36, n. 1, p. 419–444, 2005. Disponível em: ⟨doi:10.1146/annurev.ecolsys.36.091704.175535⟩.

YODER, M. J.; MIKÓ, I.; SELTMANN, K. C.; BERTONE, M. A.; DEANS, A. R. A Gross Anatomy Ontology for Hymenoptera. **PLOS ONE**, v. 5, n. 12, p. e15991, 2010. ISSN 1932-6203. Disponível em: ⟨doi:10.1371/journal.pone.0015991⟩.

ZHAO, W.; ZOU, W.; CHEN, J. J. Topic modeling for cluster analysis of large biological and medical datasets. **BMC Bioinformatics**, v. 15, n. 11, p. S11, 2014. ISSN 1471-2105. Disponível em: ⟨doi:10.1186/1471-2105-15-S11-S11⟩.

APPENDIX A – UNSUPPORTED INTERACTION TYPES FOUND BY *ELTON*

provided interaction type	mapped to interaction type	mapped to interaction type id*
1925. flowers visited	visits flowers of	RO_0002622
aassociated species	ecologically co-occurs with	RO_0008506
accociated species include	ecologically co-occurs with	RO_0008506
accociated taxa	ecologically co-occurs with	RO_0008506
accoc. spp.	ecologically co-occurs with	RO_0008506
accompanying plants	ecologically co-occurs with	RO_0008506
accosiated sp.	ecologically co-occurs with	RO_0008506
accosiated vegetation	ecologically co-occurs with	RO_0008506
alimentación en	eats	RO_0002470
alimentación y percha en	eats	RO_0002470
aliméntandose de	eats	RO_0002470
alimentandose en	eats	RO_0002470
also collected here	ecologically co-occurs with	RO_0008506
also collected nearby	ecologically co-occurs with	RO_0008506
also collected or seen nearby	ecologically co-occurs with	RO_0008506
also flowering	ecologically co-occurs with	RO_0008506
also here	ecologically co-occurs with	RO_0008506
also in area	ecologically co-occurs with	RO_0008506
also present at site	ecologically co-occurs with	RO_0008506
also seen or collected here	ecologically co-occurs with	RO_0008506
also seen or collected nearby	ecologically co-occurs with	RO_0008506
aociado con	ecologically co-occurs with	RO_0008506
aqssociated species	ecologically co-occurs with	RO_0008506

ascociada	ecologically co-occurs with	RO_0008506
asociada a	ecologically co-occurs with	RO_0008506
asociada a spp	ecologically co-occurs with	RO_0008506
asociada con	ecologically co-occurs with	RO_0008506
asociadas	ecologically co-occurs with	RO_0008506
asociado a	ecologically co-occurs with	RO_0008506
asociado	ecologically co-occurs with	RO_0008506
asociados	ecologically co-occurs with	RO_0008506
asociadp con	ecologically co-occurs with	RO_0008506
asociated plants	ecologically co-occurs with	RO_0008506
asociated species	ecologically co-occurs with	RO_0008506
asociate include	ecologically co-occurs with	RO_0008506
asociates	ecologically co-occurs with	RO_0008506
asociates include	ecologically co-occurs with	RO_0008506
asociating species	ecologically co-occurs with	RO_0008506
asociation	ecologically co-occurs with	RO_0008506
asocidas	ecologically co-occurs with	RO_0008506
asoc. sp	ecologically co-occurs with	RO_0008506
asoc. sp. include	ecologically co-occurs with	RO_0008506
asoociates	ecologically co-occurs with	RO_0008506
asooc. plants	ecologically co-occurs with	RO_0008506
asooc. spp.	ecologically co-occurs with	RO_0008506
asosiación de	ecologically co-occurs with	RO_0008506
assciated plants	ecologically co-occurs with	RO_0008506
assciated species	ecologically co-occurs with	RO_0008506
assciations	ecologically co-occurs with	RO_0008506
ass. cirsieto-molinietum accomp.	ecologically co-occurs with	RO_0008506
asscociates include	ecologically co-occurs with	RO_0008506
asscociate trees	ecologically co-occurs with	RO_0008506
assco	ecologically co-occurs with	RO_0008506
asscoiated	ecologically co-occurs with	RO_0008506
asscoiated species	ecologically co-occurs with	RO_0008506
asscoiated spp. include	ecologically co-occurs with	RO_0008506
assc plants	ecologically co-occurs with	RO_0008506
assc. plants	ecologically co-occurs with	RO_0008506

assc pl	ecologically co-occurs with	RO_0008506
assc. pl	ecologically co-occurs with	RO_0008506
assc. species	ecologically co-occurs with	RO_0008506
assc sp	ecologically co-occurs with	RO_0008506
assication	ecologically co-occurs with	RO_0008506
assiciates include	ecologically co-occurs with	RO_0008506
assic species	ecologically co-occurs with	RO_0008506
assicuated plants included	ecologically co-occurs with	RO_0008506
ass. include	ecologically co-occurs with	RO_0008506
assocaited dominant	ecologically co-occurs with	RO_0008506
assocaited	ecologically co-occurs with	RO_0008506
assocaited herbs	ecologically co-occurs with	RO_0008506
assocaited plants included	ecologically co-occurs with	RO_0008506
assocaited species	ecologically co-occurs with	RO_0008506
assocaited species include	ecologically co-occurs with	RO_0008506
assocaited spp.	ecologically co-occurs with	RO_0008506
assocaited taxa	ecologically co-occurs with	RO_0008506
assocaited taxa include	ecologically co-occurs with	RO_0008506
assocaited with	ecologically co-occurs with	RO_0008506
assocaites	ecologically co-occurs with	RO_0008506
assocaite species	ecologically co-occurs with	RO_0008506
assocaites with	ecologically co-occurs with	RO_0008506
assocaition	ecologically co-occurs with	RO_0008506
assocaitions	ecologically co-occurs with	RO_0008506
assoc. apecies	ecologically co-occurs with	RO_0008506
assoc. app	ecologically co-occurs with	RO_0008506
assocated plants	ecologically co-occurs with	RO_0008506
assocated species	ecologically co-occurs with	RO_0008506
assocated spp.	ecologically co-occurs with	RO_0008506
assocated vegetation	ecologically co-occurs with	RO_0008506
assocates	ecologically co-occurs with	RO_0008506
assocates include	ecologically co-occurs with	RO_0008506
assocation	ecologically co-occurs with	RO_0008506
assocation with	ecologically co-occurs with	RO_0008506
assocciates	ecologically co-occurs with	RO_0008506
assocciates include	ecologically co-occurs with	RO_0008506

assoc'd genera	ecologically co-occurs with	RO_0008506
assoc. dominants	ecologically co-occurs with	RO_0008506
. assoc	ecologically co-occurs with	RO_0008506
. assoc.	ecologically co-occurs with	RO_0008506
assoc	ecologically co-occurs with	RO_0008506
assoc.	ecologically co-occurs with	RO_0008506
assoc..	ecologically co-occurs with	RO_0008506
assoc genera	ecologically co-occurs with	RO_0008506
assoc. genera	ecologically co-occurs with	RO_0008506
assoc.genera	ecologically co-occurs with	RO_0008506
assoc. gener	ecologically co-occurs with	RO_0008506
assoc. gerera	ecologically co-occurs with	RO_0008506
assoc grasses	ecologically co-occurs with	RO_0008506
assoc. grasses	ecologically co-occurs with	RO_0008506
assoc. herbaceous species include	ecologically co-occurs with	RO_0008506
assoc herbaceous spp. include	ecologically co-occurs with	RO_0008506
assoc. herbaceous spp. include	ecologically co-occurs with	RO_0008506
assoc herbaceous spp inlcude	ecologically co-occurs with	RO_0008506
assoc. herbaceous spp. inlcude	ecologically co-occurs with	RO_0008506
assoc. herbs	ecologically co-occurs with	RO_0008506
associaates	ecologically co-occurs with	RO_0008506
associaation	ecologically co-occurs with	RO_0008506
associada	ecologically co-occurs with	RO_0008506
associado	ecologically co-occurs with	RO_0008506
associaites include	ecologically co-occurs with	RO_0008506
associaition	ecologically co-occurs with	RO_0008506
associaiton	ecologically co-occurs with	RO_0008506
associatas	ecologically co-occurs with	RO_0008506
associatd species	ecologically co-occurs with	RO_0008506
associated accacias	ecologically co-occurs with	RO_0008506
associated annuals	ecologically co-occurs with	RO_0008506
associated biota	ecologically co-occurs with	RO_0008506
associated canopy vegetation	ecologically co-occurs with	RO_0008506
associated conifers	ecologically co-occurs with	RO_0008506
associated conifers include	ecologically co-occurs with	RO_0008506

associated dominant	ecologically co-occurs with	RO_0008506
associated dominants	ecologically co-occurs with	RO_0008506
associated dominant shrubs included	ecologically co-occurs with	RO_0008506
associated dominats	ecologically co-occurs with	RO_0008506
associated donimants	ecologically co-occurs with	RO_0008506
associated dormants	ecologically co-occurs with	RO_0008506
associated	ecologically co-occurs with	RO_0008506
associated euc. species	ecologically co-occurs with	RO_0008506
associated floa include	ecologically co-occurs with	RO_0008506
associated flora	ecologically co-occurs with	RO_0008506
associated genera and species	ecologically co-occurs with	RO_0008506
associated genera	ecologically co-occurs with	RO_0008506
associated genera include	ecologically co-occurs with	RO_0008506
associated ground cover	ecologically co-occurs with	RO_0008506
associated herbaceous species include	ecologically co-occurs with	RO_0008506
associated herbs	ecologically co-occurs with	RO_0008506
associated in area	ecologically co-occurs with	RO_0008506
associated include	ecologically co-occurs with	RO_0008506
associated in creek are	ecologically co-occurs with	RO_0008506
associated in order of abundance	ecologically co-occurs with	RO_0008506
associated insect	ecologically co-occurs with	RO_0008506
associated lianes	ecologically co-occurs with	RO_0008506
associated lichens	ecologically co-occurs with	RO_0008506
associate dominant species	ecologically co-occurs with	RO_0008506
associated orchids include	ecologically co-occurs with	RO_0008506
associated palnts	ecologically co-occurs with	RO_0008506
associated pct	ecologically co-occurs with	RO_0008506
associated perennials	ecologically co-occurs with	RO_0008506
associated plantas	ecologically co-occurs with	RO_0008506
associated plant	ecologically co-occurs with	RO_0008506
associated plants are	ecologically co-occurs with	RO_0008506
. associated plants	ecologically co-occurs with	RO_0008506
associated plants	ecologically co-occurs with	RO_0008506

associated plants incluced	ecologically co-occurs with	RO_0008506
associated plants included	ecologically co-occurs with	RO_0008506
associated plants included the algae	ecologically co-occurs with	RO_0008506
associated plants included the dominant shrubs	ecologically co-occurs with	RO_0008506
associated plants included the weeds	ecologically co-occurs with	RO_0008506
associated plants include	ecologically co-occurs with	RO_0008506
associated plants in order of abundance	ecologically co-occurs with	RO_0008506
associated plant species	ecologically co-occurs with	RO_0008506
associated plant spp.	ecologically co-occurs with	RO_0008506
associated plants with	ecologically co-occurs with	RO_0008506
associated platns	ecologically co-occurs with	RO_0008506
associated plats	ecologically co-occurs with	RO_0008506
associated secies include	ecologically co-occurs with	RO_0008506
associated sedges include	ecologically co-occurs with	RO_0008506
associated shrub	ecologically co-occurs with	RO_0008506
associated shrubs	ecologically co-occurs with	RO_0008506
associated speceis	ecologically co-occurs with	RO_0008506
associated speces	ecologically co-occurs with	RO_0008506
associated specied	ecologically co-occurs with	RO_0008506
associated specieds	ecologically co-occurs with	RO_0008506
associated specie	ecologically co-occurs with	RO_0008506
associated species are as follows	ecologically co-occurs with	RO_0008506
associated species are	ecologically co-occurs with	RO_0008506
associated species as follows	ecologically co-occurs with	RO_0008506
associated species at this site	ecologically co-occurs with	RO_0008506
associated species collected here	ecologically co-occurs with	RO_0008506
. associated species	ecologically co-occurs with	RO_0008506
associated species	ecologically co-occurs with	RO_0008506
associated species?	ecologically co-occurs with	RO_0008506
associated species}	ecologically co-occurs with	RO_0008506
associated species/genera	ecologically co-occurs with	RO_0008506

. associated species include	ecologically co-occurs with	RO_0008506
... associated species include	ecologically co-occurs with	RO_0008506
associated species include	ecologically co-occurs with	RO_0008506
associated species (in order of abundance)	ecologically co-occurs with	RO_0008506
associated species of wash	ecologically co-occurs with	RO_0008506
associated speciess	ecologically co-occurs with	RO_0008506
associated species written as shown on actual label- several of the names are incorrect/incomplete- accurate names are likely	ecologically co-occurs with	RO_0008506
associated specifies	ecologically co-occurs with	RO_0008506
associated specis	ecologically co-occurs with	RO_0008506
associated speckes	ecologically co-occurs with	RO_0008506
associated sp	ecologically co-occurs with	RO_0008506
associated sp.	ecologically co-occurs with	RO_0008506
associated speices	ecologically co-occurs with	RO_0008506
associated speies	ecologically co-occurs with	RO_0008506
associated spesies	ecologically co-occurs with	RO_0008506
associated sppecies	ecologically co-occurs with	RO_0008506
associated spp	ecologically co-occurs with	RO_0008506
associated spp.	ecologically co-occurs with	RO_0008506
associated spp include	ecologically co-occurs with	RO_0008506
associated spp. include	ecologically co-occurs with	RO_0008506
associated ssp	ecologically co-occurs with	RO_0008506
associated ssp.	ecologically co-occurs with	RO_0008506
associated taca	ecologically co-occurs with	RO_0008506
associated taxa	ecologically co-occurs with	RO_0008506
associated taxa include	ecologically co-occurs with	RO_0008506
associated taxon	ecologically co-occurs with	RO_0008506
associated threatened species	ecologically co-occurs with	RO_0008506
associated to	ecologically co-occurs with	RO_0008506
associated tree	ecologically co-occurs with	RO_0008506
associated trees	ecologically co-occurs with	RO_0008506
associated tree species	ecologically co-occurs with	RO_0008506

associated txa	ecologically co-occurs with	RO_0008506
associated vascular plant species	ecologically co-occurs with	RO_0008506
associated vegation	ecologically co-occurs with	RO_0008506
associated veg	ecologically co-occurs with	RO_0008506
associated veg.	ecologically co-occurs with	RO_0008506
associated vegetaiton	ecologically co-occurs with	RO_0008506
?. associated vegetation	ecologically co-occurs with	RO_0008506
associated vegetation	ecologically co-occurs with	RO_0008506
associated vegetation	ecologically co-occurs with	RO_0008506
associated. vegetation	ecologically co-occurs with	RO_0008506
associated weed species	ecologically co-occurs with	RO_0008506
associated w/ elements of tropical caducifolioius forest	ecologically co-occurs with	RO_0008506
associated wetland plants include	ecologically co-occurs with	RO_0008506
associated with	ecologically co-occurs with	RO_0008506
associated with	ecologically co-occurs with	RO_0008506
associated with species	ecologically co-occurs with	RO_0008506
associated with weedy plants	ecologically co-occurs with	RO_0008506
associated woody plants	ecologically co-occurs with	RO_0008506
associate	ecologically co-occurs with	RO_0008506
associatees include	ecologically co-occurs with	RO_0008506
associate genera	ecologically co-occurs with	RO_0008506
associate grasses	ecologically co-occurs with	RO_0008506
associate herbs	ecologically co-occurs with	RO_0008506
associate include	ecologically co-occurs with	RO_0008506
associate of	ecologically co-occurs with	RO_0008506
associate plant	ecologically co-occurs with	RO_0008506
associate plants	ecologically co-occurs with	RO_0008506
associates by rank	ecologically co-occurs with	RO_0008506
associates collected here	ecologically co-occurs with	RO_0008506
associates/ cover	ecologically co-occurs with	RO_0008506
. associates	ecologically co-occurs with	RO_0008506
associates	ecologically co-occurs with	RO_0008506
associates here	ecologically co-occurs with	RO_0008506

associates inclclude	ecologically co-occurs with	RO_0008506
associates included	ecologically co-occurs with	RO_0008506
associates include	ecologically co-occurs with	RO_0008506
associates includes	ecologically co-occurs with	RO_0008506
associates inclued	ecologically co-occurs with	RO_0008506
associates inculde	ecologically co-occurs with	RO_0008506
associates influde	ecologically co-occurs with	RO_0008506
associate species	ecologically co-occurs with	RO_0008506
associate species include	ecologically co-occurs with	RO_0008506
associates plants	ecologically co-occurs with	RO_0008506
associate spp	ecologically co-occurs with	RO_0008506
associatess	ecologically co-occurs with	RO_0008506
associates species as follows	ecologically co-occurs with	RO_0008506
associates species	ecologically co-occurs with	RO_0008506
associates species include	ecologically co-occurs with	RO_0008506
associate ssp	ecologically co-occurs with	RO_0008506
associate ssp.	ecologically co-occurs with	RO_0008506
associates sp.	ecologically co-occurs with	RO_0008506
associates. sp.	ecologically co-occurs with	RO_0008506
associates spp	ecologically co-occurs with	RO_0008506
associates spp.	ecologically co-occurs with	RO_0008506
associates vegetation	ecologically co-occurs with	RO_0008506
associates were	ecologically co-occurs with	RO_0008506
{associates with	ecologically co-occurs with	RO_0008506
associate taxa	ecologically co-occurs with	RO_0008506
associate trees	ecologically co-occurs with	RO_0008506
associate veg	ecologically co-occurs with	RO_0008506
associate vegetation	ecologically co-occurs with	RO_0008506
associatged vegetation	ecologically co-occurs with	RO_0008506
associatied with	ecologically co-occurs with	RO_0008506
association	ecologically co-occurs with	RO_0008506
associations	ecologically co-occurs with	RO_0008506
association species	ecologically co-occurs with	RO_0008506
associaton	ecologically co-occurs with	RO_0008506
associats	ecologically co-occurs with	RO_0008506
associatse	ecologically co-occurs with	RO_0008506

associeated species	ecologically co-occurs with	RO_0008506
associ	ecologically co-occurs with	RO_0008506
associ.	ecologically co-occurs with	RO_0008506
associes	ecologically co-occurs with	RO_0008506
associes.	ecologically co-occurs with	RO_0008506
associeted species	ecologically co-occurs with	RO_0008506
associ. genera	ecologically co-occurs with	RO_0008506
assoc. included	ecologically co-occurs with	RO_0008506
assoc include	ecologically co-occurs with	RO_0008506
assoc. include	ecologically co-occurs with	RO_0008506
assoc.include	ecologically co-occurs with	RO_0008506
associqated species	ecologically co-occurs with	RO_0008506
associqatess	ecologically co-occurs with	RO_0008506
associtaion	ecologically co-occurs with	RO_0008506
associtated dominants	ecologically co-occurs with	RO_0008506
associtated plants	ecologically co-occurs with	RO_0008506
associtated spp. include	ecologically co-occurs with	RO_0008506
associtates	ecologically co-occurs with	RO_0008506
associtations	ecologically co-occurs with	RO_0008506
associted species	ecologically co-occurs with	RO_0008506
associted spp.	ecologically co-occurs with	RO_0008506
associted taxa	ecologically co-occurs with	RO_0008506
associtees include	ecologically co-occurs with	RO_0008506
assocition	ecologically co-occurs with	RO_0008506
assoclates	ecologically co-occurs with	RO_0008506
assoc. lichens	ecologically co-occurs with	RO_0008506
assocl. species	ecologically co-occurs with	RO_0008506
assocl spp.	ecologically co-occurs with	RO_0008506
assocoiates include	ecologically co-occurs with	RO_0008506
assoc. plant	ecologically co-occurs with	RO_0008506
assoc plants	ecologically co-occurs with	RO_0008506
assoc. plants	ecologically co-occurs with	RO_0008506
assoc. plants include	ecologically co-occurs with	RO_0008506
assoc. plants/veg. type	ecologically co-occurs with	RO_0008506
assoc. pl.	ecologically co-occurs with	RO_0008506
assoc. pls/notes	ecologically co-occurs with	RO_0008506

assocs	ecologically co-occurs with	RO_0008506
assocs.	ecologically co-occurs with	RO_0008506
assoc. soecies	ecologically co-occurs with	RO_0008506
assoc spec	ecologically co-occurs with	RO_0008506
assoc. spec	ecologically co-occurs with	RO_0008506
assoc. spec.	ecologically co-occurs with	RO_0008506
assoc. species are	ecologically co-occurs with	RO_0008506
. assoc. species	ecologically co-occurs with	RO_0008506
assoc species	ecologically co-occurs with	RO_0008506
assoc. species	ecologically co-occurs with	RO_0008506
assoc. species.	ecologically co-occurs with	RO_0008506
assoc.species	ecologically co-occurs with	RO_0008506
assoc species include	ecologically co-occurs with	RO_0008506
assoc. species include	ecologically co-occurs with	RO_0008506
assoc. speciesinclude	ecologically co-occurs with	RO_0008506
assoc.species include	ecologically co-occurs with	RO_0008506
. assoc. sp	ecologically co-occurs with	RO_0008506
assoc sp	ecologically co-occurs with	RO_0008506
assoc sp.	ecologically co-occurs with	RO_0008506
assoc. sp	ecologically co-occurs with	RO_0008506
assoc. sp.	ecologically co-occurs with	RO_0008506
assoc.sp	ecologically co-occurs with	RO_0008506
assoc.sp.	ecologically co-occurs with	RO_0008506
assoc. speices	ecologically co-occurs with	RO_0008506
assoc. sp. inc.	ecologically co-occurs with	RO_0008506
assoc. sp. included	ecologically co-occurs with	RO_0008506
assoc sp. include	ecologically co-occurs with	RO_0008506
assoc. sp include	ecologically co-occurs with	RO_0008506
assoc. sp. include	ecologically co-occurs with	RO_0008506
assoc. sp. on face	ecologically co-occurs with	RO_0008506
. assoc. spp	ecologically co-occurs with	RO_0008506
assoc spp	ecologically co-occurs with	RO_0008506
assoc spp.	ecologically co-occurs with	RO_0008506
assoc. spp	ecologically co-occurs with	RO_0008506
assoc. spp.	ecologically co-occurs with	RO_0008506
assoc.spp.	ecologically co-occurs with	RO_0008506

assoc. spp. inc.	ecologically co-occurs with	RO_0008506
assoc. spp. incl	ecologically co-occurs with	RO_0008506
assoc. spp. incl.	ecologically co-occurs with	RO_0008506
assoc spp include	ecologically co-occurs with	RO_0008506
assoc spp. include	ecologically co-occurs with	RO_0008506
assoc. spp include	ecologically co-occurs with	RO_0008506
assoc. spp. include	ecologically co-occurs with	RO_0008506
assoc. spp. inlcude	ecologically co-occurs with	RO_0008506
assoc. spp. inlude	ecologically co-occurs with	RO_0008506
assoc. spp. on verticle rock face	ecologically co-occurs with	RO_0008506
assoc. sps	ecologically co-occurs with	RO_0008506
assoc ssp	ecologically co-occurs with	RO_0008506
assoc. ssp	ecologically co-occurs with	RO_0008506
assoc. ssp.	ecologically co-occurs with	RO_0008506
assoc. ssp. incl.	ecologically co-occurs with	RO_0008506
assoc. taxa	ecologically co-occurs with	RO_0008506
assoc trees	ecologically co-occurs with	RO_0008506
assoc. trees	ecologically co-occurs with	RO_0008506
assoc. type	ecologically co-occurs with	RO_0008506
assoc veg	ecologically co-occurs with	RO_0008506
assoc. veg	ecologically co-occurs with	RO_0008506
assoc. veg.	ecologically co-occurs with	RO_0008506
assoc.veg	ecologically co-occurs with	RO_0008506
assoc.veg.	ecologically co-occurs with	RO_0008506
assoc. vege	ecologically co-occurs with	RO_0008506
assoc. vegetation	ecologically co-occurs with	RO_0008506
assoc vegn	ecologically co-occurs with	RO_0008506
assoc. vegn	ecologically co-occurs with	RO_0008506
assoc. vegn.	ecologically co-occurs with	RO_0008506
assoc.vegn	ecologically co-occurs with	RO_0008506
assoc. w/	ecologically co-occurs with	RO_0008506
assoc. with dominants	ecologically co-occurs with	RO_0008506
assoc with	ecologically co-occurs with	RO_0008506
assoc. with	ecologically co-occurs with	RO_0008506
assocxiated with	ecologically co-occurs with	RO_0008506

assodciated species	ecologically co-occurs with	RO_0008506
asso	ecologically co-occurs with	RO_0008506
asso.	ecologically co-occurs with	RO_0008506
assofciated species	ecologically co-occurs with	RO_0008506
asso. genera	ecologically co-occurs with	RO_0008506
assoiation	ecologically co-occurs with	RO_0008506
assoicaition	ecologically co-occurs with	RO_0008506
assoicaiton	ecologically co-occurs with	RO_0008506
assoicated plants	ecologically co-occurs with	RO_0008506
assoicated species	ecologically co-occurs with	RO_0008506
assoicates	ecologically co-occurs with	RO_0008506
assoication	ecologically co-occurs with	RO_0008506
assoiciated species	ecologically co-occurs with	RO_0008506
assoiciation	ecologically co-occurs with	RO_0008506
assonc.	ecologically co-occurs with	RO_0008506
assosciated species	ecologically co-occurs with	RO_0008506
assosiated species	ecologically co-occurs with	RO_0008506
asso. species	ecologically co-occurs with	RO_0008506
asso. sp	ecologically co-occurs with	RO_0008506
asso. sp.	ecologically co-occurs with	RO_0008506
asso.sp	ecologically co-occurs with	RO_0008506
asso. spp	ecologically co-occurs with	RO_0008506
asso.spp	ecologically co-occurs with	RO_0008506
assoxiated species	ecologically co-occurs with	RO_0008506
asssoc	ecologically co-occurs with	RO_0008506
asssoc. genera	ecologically co-occurs with	RO_0008506
asssociated species	ecologically co-occurs with	RO_0008506
asssociates	ecologically co-occurs with	RO_0008506
asssociation	ecologically co-occurs with	RO_0008506
asssoc. spp.	ecologically co-occurs with	RO_0008506
ass. spec.	ecologically co-occurs with	RO_0008506
ass. species	ecologically co-occurs with	RO_0008506
ass sp.	ecologically co-occurs with	RO_0008506
ass. sp	ecologically co-occurs with	RO_0008506
ass. sp.	ecologically co-occurs with	RO_0008506
ass. spp	ecologically co-occurs with	RO_0008506

ass. spp.	ecologically co-occurs with	RO_0008506
ass. spp. inc.	ecologically co-occurs with	RO_0008506
ass spp include	ecologically co-occurs with	RO_0008506
ass spp. include	ecologically co-occurs with	RO_0008506
ass. spp. include	ecologically co-occurs with	RO_0008506
ass. spp. incude	ecologically co-occurs with	RO_0008506
ass. ssp.	ecologically co-occurs with	RO_0008506
(ate)	eats	RO_0002470
attached to	ecologically co-occurs with	RO_0008506
collectedwith	ecologically co-occurs with	RO_0008506
common asscociates	ecologically co-occurs with	RO_0008506
common assoc	ecologically co-occurs with	RO_0008506
common assoc.	ecologically co-occurs with	RO_0008506
common. assoc	ecologically co-occurs with	RO_0008506
common associated include	ecologically co-occurs with	RO_0008506
common. associated plants	ecologically co-occurs with	RO_0008506
common associated plants included	ecologically co-occurs with	RO_0008506
common associated species	ecologically co-occurs with	RO_0008506
common. associated species	ecologically co-occurs with	RO_0008506
common associated species include	ecologically co-occurs with	RO_0008506
common associated speices include	ecologically co-occurs with	RO_0008506
common associated spescies include	ecologically co-occurs with	RO_0008506
common associated spp.	ecologically co-occurs with	RO_0008506
common associated spp include	ecologically co-occurs with	RO_0008506
common. associated with	ecologically co-occurs with	RO_0008506
common. associate	ecologically co-occurs with	RO_0008506
common associates	ecologically co-occurs with	RO_0008506
common. associates	ecologically co-occurs with	RO_0008506
common. associates	ecologically co-occurs with	RO_0008506
common associates include	ecologically co-occurs with	RO_0008506
common assoc. species	ecologically co-occurs with	RO_0008506
common assoc. species include	ecologically co-occurs with	RO_0008506

common assoc. spp.	ecologically co-occurs with	RO_0008506
common. assoc. spp.	ecologically co-occurs with	RO_0008506
common assoc spp. include	ecologically co-occurs with	RO_0008506
common assoc. spp include	ecologically co-occurs with	RO_0008506
common assoc. spp. include	ecologically co-occurs with	RO_0008506
common assoc. spp. inlcude	ecologically co-occurs with	RO_0008506
common assoicates include	ecologically co-occurs with	RO_0008506
common speces include	ecologically co-occurs with	RO_0008506
common species are	ecologically co-occurs with	RO_0008506
common species	ecologically co-occurs with	RO_0008506
common species include	ecologically co-occurs with	RO_0008506
common species in the area include	ecologically co-occurs with	RO_0008506
common species throughout the canyon include	ecologically co-occurs with	RO_0008506
common spp	ecologically co-occurs with	RO_0008506
common spp.	ecologically co-occurs with	RO_0008506
common spp. include	ecologically co-occurs with	RO_0008506
companian plants	ecologically co-occurs with	RO_0008506
companion plants	ecologically co-occurs with	RO_0008506
companion sp	ecologically co-occurs with	RO_0008506
compartment 98. found in association with	ecologically co-occurs with	RO_0008506
consume flor de	eats	RO_0002470
consume fruto de	eats	RO_0002470
consume néctar de	acquires nutrients from	RO_0002457
consume néctar y polen de	acquires nutrients from	RO_0002457
consume polen de	acquires nutrients from	RO_0002457
consume semillas de	acquires nutrients from	RO_0002457
consume su flor	eats	RO_0002470
consume su fruto	eats	RO_0002470
consumidor de fruto	eats	RO_0002470
depredador	preys on	RO_0002439
direct associates	ecologically co-occurs with	RO_0008506
(eaten by)	is eaten by	RO_0002471
ecologicallyoccurswith	ecologically co-occurs with	RO_0008506

ecology. host plants	has host	RO_0002454
ectoparasite	ectoparasite of	RO_0002632
ectoparasite of	ectoparasite of	RO_0002632
ectoparasito de	ectoparasite of	RO_0002632
ectoparásito de	ectoparasite of	RO_0002632
ectoparásito	ectoparasite of	RO_0002632
ectoparásitos	ectoparasite of	RO_0002632
epibionte	has host	RO_0002454
epibiont	has host	RO_0002454
epífita de	epiphyte of	RO_0008501
epifita en	epiphyte of	RO_0008501
epífita	epiphyte of	RO_0008501
epiparásito	hyperparasite of	RO_0002553
epiphyte on	epiphyte of	RO_0008501
epiphytes	epiphyte of	RO_0008501
epiphytic on	epiphyte of	RO_0008501
epipihyte on	epiphyte of	RO_0008501
especies acompañantes	ecologically co-occurs with	RO_0008506
especies asociadas	ecologically co-occurs with	RO_0008506
especies asociadsa	ecologically co-occurs with	RO_0008506
especies características	ecologically co-occurs with	RO_0008506
especies co-habitantes	ecologically co-occurs with	RO_0008506
flor visitada	visits flower of	RO_0002622
found in association with	ecologically co-occurs with	RO_0008506
found with host species	has host	RO_0002454
fruto consumido	eats	RO_0002470
fruto visitado	visits	RO_0002618
growing with	ecologically co-occurs with	RO_0008506
growingwith	ecologically co-occurs with	RO_0008506
habitat an assoc. spp.	ecologically co-occurs with	RO_0008506
habitat and assoc sp	ecologically co-occurs with	RO_0008506
habitat and assoc. sp	ecologically co-occurs with	RO_0008506
habitat and assoc. sp.	ecologically co-occurs with	RO_0008506
habitat and assoc spp.	ecologically co-occurs with	RO_0008506
habitat and assoc. spp	ecologically co-occurs with	RO_0008506
habitat and assoc. spp.	ecologically co-occurs with	RO_0008506

habitat/associates	ecologically co-occurs with	RO_0008506
habitat & assoc. species	ecologically co-occurs with	RO_0008506
habitat & assoc sp	ecologically co-occurs with	RO_0008506
habitat & assoc. spp.	ecologically co-occurs with	RO_0008506
habitat & asso. species	ecologically co-occurs with	RO_0008506
habitat & ass. species	ecologically co-occurs with	RO_0008506
has host	has host	RO_0002454
hashost	has host	RO_0002454
has hsot	has host	RO_0002454
has parasite	parasitized by	RO_0002445
hemiparásito	parasite of	RO_0002444
hiperparásito	hyperparasite of	RO_0002553
hospedador de	host of	RO_0002453
hospedante	has host	RO_0002454
hospededero	host of	RO_0002453
hospedeiro de	has host	RO_0002454
hospedeiro	has host	RO_0002454
hospedero	has host	RO_0002454
hospederos	has host	RO_0002454
hosped	has host	RO_0002454
hosperdo por infección experimental	has host	RO_0002454
hospes	has host	RO_0002454
hosp	has host	RO_0002454
host/assoc	has host	RO_0002454
host code	has host	RO_0002454
host could be	has host	RO_0002454
host for	host of	RO_0002453
“host”	has host	RO_0002454
host	has host	RO_0002454
host?	has host	RO_0002454
(host of)	host of	RO_0002453
host of	host of	RO_0002453
hostof	host of	RO_0002453
host plant	host of	RO_0002453
hostplant	host of	RO_0002453

host plants	host of	RO_0002453
hosts	has host	RO_0002454
hosts include	host of	RO_0002453
host species	has host	RO_0002454
hostspecies	host of	RO_0002453
host/substrate	host of	RO_0002453
huésped de	host of	RO_0002453
huesped	host of	RO_0002453
huésped	host of	RO_0002453
hyperparasitoid of	parasitoid of	RO_0002208
in assoc.	ecologically co-occurs with	RO_0008506
in associatione	ecologically co-occurs with	RO_0008506
in association with	ecologically co-occurs with	RO_0008506
in assoc. with	ecologically co-occurs with	RO_0008506
infectedby	parasitized by	RO_0002445
interacción con	interactsWith	RO_0002437
interacting taxon	interactsWith	RO_0002437
interacts with	interacts with	RO_0002434
interactswith	interactsWith	RO_0002437
in vicinity	ecologically co-occurs with	RO_0008506
mutualismo	mutualistically interacts with	RO_0002442
nearby associates	ecologically co-occurs with	RO_0008506
nearby plants	ecologically co-occurs with	RO_0008506
nearby plants in area disturbed by trail	ecologically co-occurs with	RO_0008506
nearby plants were	ecologically co-occurs with	RO_0008506
nearby species	ecologically co-occurs with	RO_0008506
nearby species include	ecologically co-occurs with	RO_0008506
nearby speciesinclude	ecologically co-occurs with	RO_0008506
nest	interactsWith	RO_0002437
nestling	interactsWith	RO_0002437
nests in	interactsWith	RO_0002437
nodricismo	creates habitat for	RO_0008505
nodriza	creates habitat for	RO_0008505
observedspecies	interactsWith	RO_0002437
occurring with	ecologically co-occurs with	RO_0008506

ocurring with	ecologically co-occurs with	RO_0008506
on same slide	ecologically co-occurs with	RO_0008506
other associated plants	ecologically co-occurs with	RO_0008506
other associated plants include	ecologically co-occurs with	RO_0008506
other associated species include	ecologically co-occurs with	RO_0008506
other associated spp.	ecologically co-occurs with	RO_0008506
other associated vascular plants include	ecologically co-occurs with	RO_0008506
other associates	ecologically co-occurs with	RO_0008506
other common trees	ecologically co-occurs with	RO_0008506
other companion plants	ecologically co-occurs with	RO_0008506
other plants collected at this location include	ecologically co-occurs with	RO_0008506
other plants collected at this site include	ecologically co-occurs with	RO_0008506
other plants collected here	ecologically co-occurs with	RO_0008506
other plants collected here include	ecologically co-occurs with	RO_0008506
other plants collected here on this date include	ecologically co-occurs with	RO_0008506
other plants collected in this habitat include	ecologically co-occurs with	RO_0008506
other plants include	ecologically co-occurs with	RO_0008506
other plants in flower included	ecologically co-occurs with	RO_0008506
other plants of the area include	ecologically co-occurs with	RO_0008506
other salix spp. nearby. also low annuals	ecologically co-occurs with	RO_0008506
other sedges and rushes collected here include	ecologically co-occurs with	RO_0008506
other sedges collected at this location include	ecologically co-occurs with	RO_0008506
other sedges collected at this site include	ecologically co-occurs with	RO_0008506
other sedges collected here include	ecologically co-occurs with	RO_0008506

other sedges collected here on this date include	ecologically co-occurs with	RO_0008506
other sedges collected in this habitat include	ecologically co-occurs with	RO_0008506
other sedge species collected here include	ecologically co-occurs with	RO_0008506
other sedges seen in this habitat include	ecologically co-occurs with	RO_0008506
other sedges seen or collected here include	ecologically co-occurs with	RO_0008506
other shrubs in vicinity	ecologically co-occurs with	RO_0008506
other shrubs with this species	ecologically co-occurs with	RO_0008506
other species collected at this location include	ecologically co-occurs with	RO_0008506
other species collected at this site include	ecologically co-occurs with	RO_0008506
other species collected here include	ecologically co-occurs with	RO_0008506
other species collected here on this date include	ecologically co-occurs with	RO_0008506
other species collected in this habitat include	ecologically co-occurs with	RO_0008506
other species	ecologically co-occurs with	RO_0008506
other species include	ecologically co-occurs with	RO_0008506
other species lcoally includes	ecologically co-occurs with	RO_0008506
other species present	ecologically co-occurs with	RO_0008506
other species seen or collected here include	ecologically co-occurs with	RO_0008506
other veg doms incl	ecologically co-occurs with	RO_0008506
other vegetation	ecologically co-occurs with	RO_0008506
other weeds	ecologically co-occurs with	RO_0008506
other weed species common here	ecologically co-occurs with	RO_0008506
other willow species seen here include	ecologically co-occurs with	RO_0008506
other woody species present	ecologically co-occurs with	RO_0008506

parasitado por	parasitized by	RO_0002445
(parasite of)	parasite of	RO_0002444
parasite of	parasite of	RO_0002444
parasite	parasite of	RO_0002444
parasites	parasite of	RO_0002444
parasitically found on/in	parasite of	RO_0002444
parasitic on	parasite of	RO_0002444
parasitized by	parasite of	RO_0002444
parásito de	parasite of	RO_0002444
parasitoide de	parasite of	RO_0002444
parasitoide	parasitoide of	RO_0002208
parasitoid of	parasite of	RO_0002444
parasito	parasite of	RO_0002444
parásito	parasitized by	RO_0002445
pathogen	pathogen of	RO_0002556
patógeno de	pathogen of	RO_0002556
plant assoc	ecologically co-occurs with	RO_0008506
plant assoc.	ecologically co-occurs with	RO_0008506
plant associates are	ecologically co-occurs with	RO_0008506
plant associates	ecologically co-occurs with	RO_0008506
plant associates include	ecologically co-occurs with	RO_0008506
plant association	ecologically co-occurs with	RO_0008506
plant associations	ecologically co-occurs with	RO_0008506
plant community	ecologically co-occurs with	RO_0008506
plants associated include	ecologically co-occurs with	RO_0008506
plants include	ecologically co-occurs with	RO_0008506
plant species include	ecologically co-occurs with	RO_0008506
polinizador de	pollinated by	RO_0002456
polinizador	pollinated by	RO_0002456
pollinating	pollinates	RO_0002455
posando en	visits	RO_0002618
posandose en	visits	RO_0002618
possible hosts	has host	RO_0002454
predator of	preyed upon by	RO_0002458
predator	preyed upon by	RO_0002458
predators	preyed upon by	RO_0002458

presa de	preyed upon by	RO_0002458
presa	preyed upon by	RO_0002458
prey of	preyed upon by	RO_0002458
prey	preys on	RO_0002439
rizófago	interactsWith	RO_0002437
same sheet	ecologically co-occurs with	RO_0008506
semiparásito	parasite of	RO_0002444
simbionte	symbiotically interacts with	RO_0002440
species nearby	ecologically co-occurs with	RO_0008506
ssociated species	ecologically co-occurs with	RO_0008506
substrate	interactsWith	RO_0002437
symbiont of	symbiotically interacts with	RO_0002440
symbiont	symbiotically interacts with	RO_0002440
vicinity	ecologically co-occurs with	RO_0008506
visitador	visits	RO_0002618
visitado	visited by	RO_0002619
visitante de flor	visits flower of	RO_0002622
visitante de fruto	is eaten by	RO_0002471
visitante floral de	visits flowers of	RO_0002622
visitante floral	visits flowers of	RO_0002622
visited flower of	visits flowers of	RO_0002622
visiting flower of	visits flowers of	RO_0002622
visiting	visits	RO_0002618
visitng	visits	RO_0002618
visits flowers of	visits flowers of	RO_0002622
visitsflowersof	visits flowers of	RO_0002622
visting flower	visits flowers of	RO_0002622

APPENDIX B – BIOTIC INTERACTIONS BY KINGDOM

Source taxon	Target taxon	Total	%
Plantae	Plantae	4,678,003	56.9
Animalia	Plantae	1,218,355	14.8
Fungi	Plantae	922,207	11.2
Animalia	Viruses	553,900	6.7
Animalia	Animalia	441,204	5.4
Animalia	Bacteria	133,557	1.6
Plantae	Viruses	71,131	0.9
Fungi	Fungi	62,021	0.8
Bacteria	Plantae	58,091	0.7
Animalia	Fungi	31,263	0.4
Animalia	incertae sedis	26,128	0.3
Fungi	Viruses	4,452	0.1
Bacteria	Viruses	4,446	0.1
Bacteria	Fungi	3,142	0.0
incertae sedis	Plantae	2,869	0.0
Bacteria	Bacteria	1,419	0.0
Animalia	Archaea	1,039	0.0
Fungi	incertae sedis	714	0.0
Archaea	Plantae	456	0.0
incertae sedis	incertae sedis	256	0.0
Bacteria	incertae sedis	251	0.0
Archaea	Viruses	164	0.0
Animalia	Protista	144	0.0
incertae sedis	Viruses	95	0.0
Viruses	Viruses	21	0.0
Archaea	Bacteria	14	0.0
Archaea	Archaea	13	0.0
Archaea	Fungi	7	0.0
Archaea	incertae sedis	3	0.0

Table 19: Taxon Interactions

APPENDIX C – INTERACTION TYPES FOR PLANTS

Interaction type	#Records	%
interactsWith	5,424,533	78.0
coOccursWith	1,186,523	17.1
hasHost	310,800	4.4
visitsFlowersOf	17,664	0.3
epiphyteOf	5,395	0.0
createsHabitatFor	2,706	0.0
mutualistOf	1,608	0.0
parasiteOf	1,370	0.0
visits	231	0.0
eats	213	0.0
providesNutrientsFor	25	0.0
pathogenOf	21	0.0
hyperparasiteOf	14	0.0
pollinates	6	0.0
kills	1	0.0
preysOn	1	0.0
symbiontOf	1	0.0

Table 20: Caption

APPENDIX D – INTERACTION TYPES FOR ANIMALS

Interaction type	**#Records**	**%**
interactsWith	5,424,533	78.0
coOccursWith	1,186,523	17.1
hasHost	310,800	4.4
visitsFlowersOf	17,664	0.3
epiphyteOf	5,395	0.0
createsHabitatFor	2,706	0.0
mutualistOf	1,608	0.0
parasiteOf	1,370	0.0
visits	231	0.0
eats	213	0.0
providesNutrientsFor	25	0.0
pathogenOf	21	0.0
hyperparasiteOf	14	0.0
pollinates	6	0.0
kills	1	0.0
preysOn	1	0.0
symbiontOf	1	0.0

Unsupported interaction types mapped to OBO Relations Onotolgy (RO). *RO: ⟨http://purl.obolibrary.org/obo/⟩

APPENDIX E – PLANT-POLLINATOR INTERACTIONS - RDF GRAPH EXAMPLE

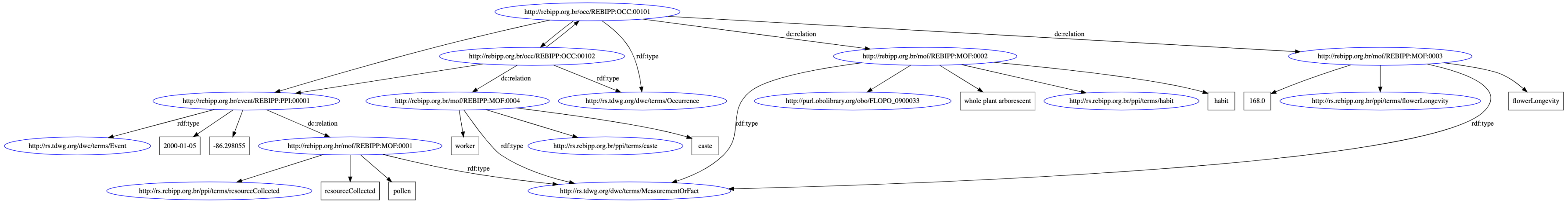

http://rebipp.org.br/occ/REBIPP:OCC:00101
dc:relation
dc:relation
http://rebipp.org.br/occ/REBIPP:OCC:00102
rdf:type
http://rebipp.org.br/mof/REBIPP:MOF:0002
http://rebipp.org.br/mof/REBIPP:MOF:0003
dc:relation
rdf:type
http://rebipp.org.br/event/REBIPP:PPI:00001
http://rebipp.org.br/mof/REBIPP:MOF:0004
http://rs.tdwg.org/dwc/terms/Occurrence
http://purl.obolibrary.org/obo/FLOPO_0900033
whole plant arborescent
http://rs.rebipp.org.br/ppi/terms/habit
habit
168.0
http://rs.rebipp.org.br/ppi/terms/flowerLongevity
flowerLongevity
rdf:type
dc:relation
rdf:type
rdf:type
http://rs.tdwg.org/dwc/terms/Event
2000-01-05
-86.298055
http://rebipp.org.br/mof/REBIPP:MOF:0001
worker
rdf:type
http://rs.rebipp.org.br/ppi/terms/caste
caste
rdf:type
http://rs.rebipp.org.br/ppi/terms/resourceCollected
resourceCollected
pollen
http://rs.tdwg.org/dwc/terms/MeasurementOrFact

Figure 43: Plant-Pollinator Interactions - RDF Graph Example

APPENDIX F – SURVEY

Descritores das Interações Planta-Polinizador

Dados pessoais

* Indicates required question

1. *

____

2. *

____

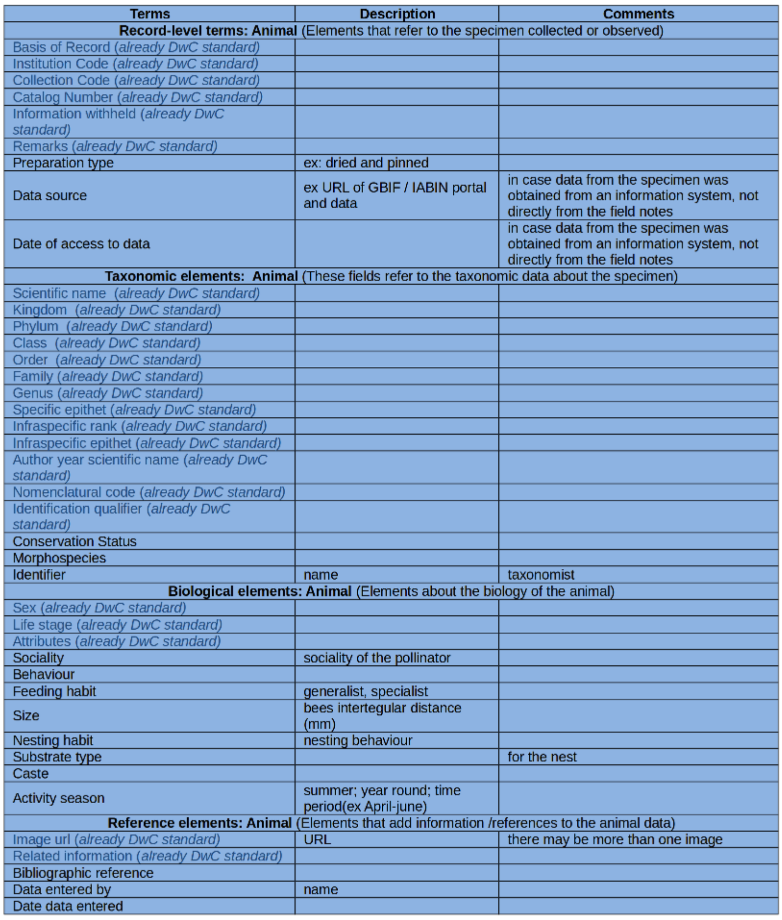
3.

____

4. *

____

Descritores para ANIMAIS



Terms	Description	Comments
Record-level terms: Animal (Elements that refer to the specimen collected or observed)		
Basis of Record (*already DwC standard)*		
Institution Code (*already DwC standard)*		
Collection Code (*already DwC standard)*		
Catalog Number (*already DwC standard)*		
Information withheld (*already DwC standard)*		
Remarks (*already DwC standard)*		
Preparation type	ex: dried and pinned	
Data source	ex URL of GBIF / IABIN portal and data	in case data from the specimen was obtained from an information system, not directly from the field notes
Date of access to data		in case data from the specimen was obtained from an information system, not directly from the field notes
Taxonomic elements: Animal (These fields refer to the taxonomic data about the specimen)		
Scientific name (*already DwC standard)*		
Kingdom (*already DwC standard)*		
Phylum (*already DwC standard)*		
Class (*already DwC standard)*		
Order (*already DwC standard)*		
Family (*already DwC standard)*		
Genus (*already DwC standard)*		
Specific epithet (*already DwC standard)*		
Infraspecific rank (*already DwC standard)*		
Infraspecific epithet (*already DwC standard)*		
Author year scientific name (*already DwC standard)*		
Nomenclatural code (*already DwC standard)*		
Identification qualifier (*already DwC standard)*		
Conservation Status		
Morphospecies		
Identifier	name	taxonomist
Biological elements: Animal (Elements about the biology of the animal)		
Sex (*already DwC standard)*		
Life stage (*already DwC standard)*		
Attributes (*already DwC standard)*		
Sociality	sociality of the pollinator	
Behaviour		
Feeding habit	generalist, specialist	
Size	bees intertegular distance (mm)	
Nesting habit	nesting behaviour	
Substrate type		for the nest
Caste		
Activity season	summer; year round; time period(ex April-june)	
Reference elements: Animal (Elements that add information /references to the animal data)		
Image url (*already DwC standard)*	URL	there may be more than one image
Related information (*already DwC standard)*		
Bibliographic reference		
Data entered by	name	
Date data entered		

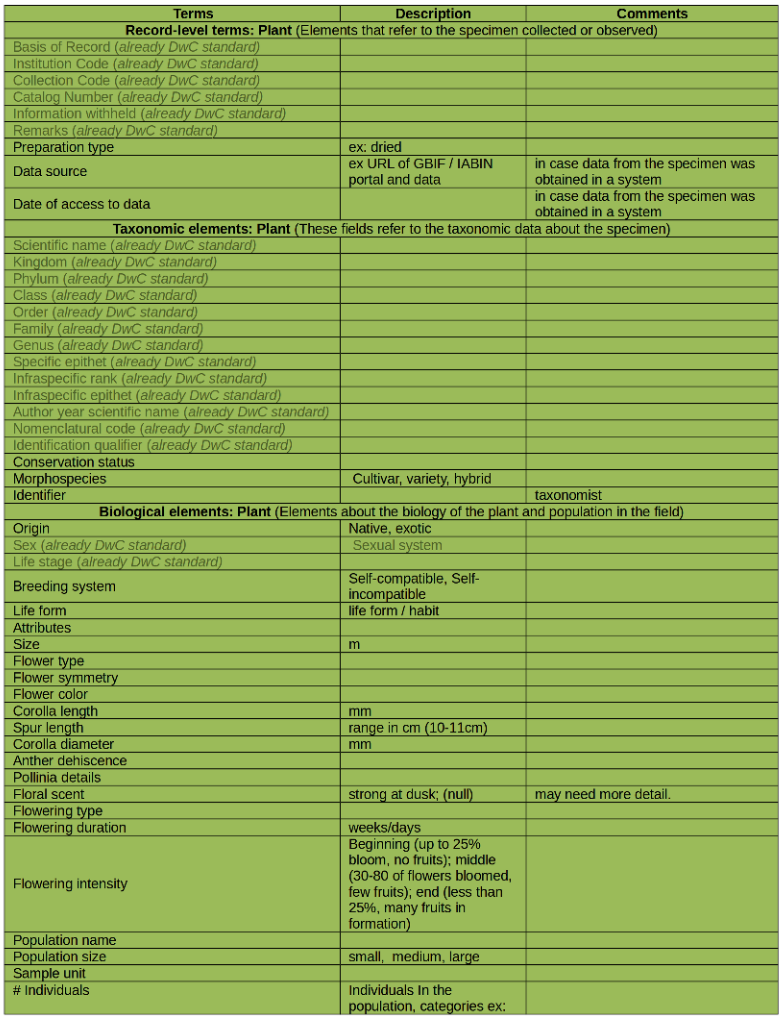
5. Dentre os descritores para ANIMAIS quais aqueles que acredita serem mais adequados às suas necessidades de pesquisa? *

Check all that apply.

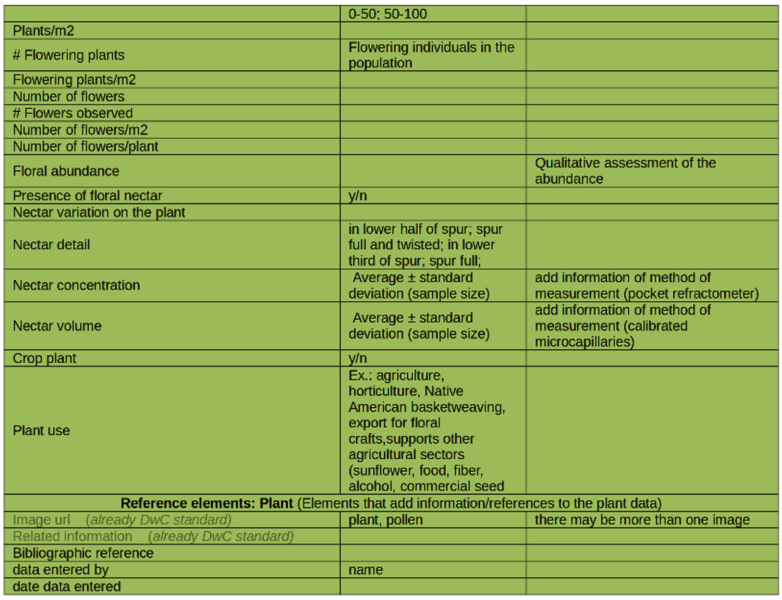
□ Elementos relativos ao espécime coletado ou observado
□ Elementos Taxonômicos
□ Elementos Biológicos (caso selecione este item, selecione também outros elementos abaixo que julgar necessários)
□ Sociabilidade (Sociality)
□ Comportamento
□ Habitat de alimentação (Feeding habitat)
□ Tamanho (Size)
□ Habitat de nidificação
□ Tipo do substrato (para nidificação)
□ Casta
□ Temporada de atividade (verão, todo o ano, Abril-Junho, etc.)
□ Elementos de Referência

□ Other: ____________________

Descritores para PLANTAS



Terms	Description	Comments
Record-level terms: Plant (Elements that refer to the specimen collected or observed)		
Basis of Record (*already DwC standard*)		
Institution Code (*already DwC standard*)		
Collection Code (*already DwC standard*)		
Catalog Number (*already DwC standard*)		
Information withheld (*already DwC standard*)		
Remarks (*already DwC standard*)		
Preparation type	ex: dried	
Data source	ex URL of GBIF / IABIN portal and data	in case data from the specimen was obtained in a system
Date of access to data		in case data from the specimen was obtained in a system
Taxonomic elements: Plant (These fields refer to the taxonomic data about the specimen)		
Scientific name (*already DwC standard*)		
Kingdom (*already DwC standard*)		
Phylum (*already DwC standard*)		
Class (*already DwC standard*)		
Order (*already DwC standard*)		
Family (*already DwC standard*)		
Genus (*already DwC standard*)		
Specific epithet (*already DwC standard*)		
Infraspecific rank (*already DwC standard*)		
Infraspecific epithet (*already DwC standard*)		
Author year scientific name (*already DwC standard*)		
Nomenclatural code (*already DwC standard*)		
Identification qualifier (*already DwC standard*)		
Conservation status		
Morphospecies	Cultivar, variety, hybrid	
Identifier		taxonomist
Biological elements: Plant (Elements about the biology of the plant and population in the field)		
Origin	Native, exotic	
Sex (*already DwC standard*)	Sexual system	
Life stage (*already DwC standard*)		
Breeding system	Self-compatible, Self-incompatible	
Life form	life form / habit	
Attributes		
Size	m	
Flower type		
Flower symmetry		
Flower color		
Corolla length	mm	
Spur length	range in cm (10-11cm)	
Corolla diameter	mm	
Anther dehiscence		
Pollinia details		
Floral scent	strong at dusk; (null)	may need more detail.
Flowering type		
Flowering duration	weeks/days	
Flowering intensity	Beginning (up to 25% bloom, no fruits); middle (30-80 of flowers bloomed, few fruits); end (less than 25%, many fruits in formation)	
Population name		
Population size	small, medium, large	
Sample unit		
# Individuals	Individuals In the population, categories ex:	



	0-50; 50-100	
Plants/m2		
# Flowering plants	Flowering individuals in the population	
Flowering plants/m2		
Number of flowers		
# Flowers observed		
Number of flowers/m2		
Number of flowers/plant		
Floral abundance		Qualitative assessment of the abundance
Presence of floral nectar	y/n	
Nectar variation on the plant		
Nectar detail	in lower half of spur; spur full and twisted; in lower third of spur; spur full;	
Nectar concentration	Average ± standard deviation (sample size)	add information of method of measurement (pocket refractometer)
Nectar volume	Average ± standard deviation (sample size)	add information of method of measurement (calibrated microcapillaries)
Crop plant	y/n	
Plant use	Ex.: agriculture, horticulture, Native American basketweaving, export for floral crafts,supports other agricultural sectors (sunflower, food, fiber, alcohol, commercial seed	
Reference elements: Plant (Elements that add information/references to the plant data)		
Image url *(already DwC standard)*	plant, pollen	there may be more than one image
Related information *(already DwC standard)*		
Bibliographic reference		
data entered by	name	
date data entered		

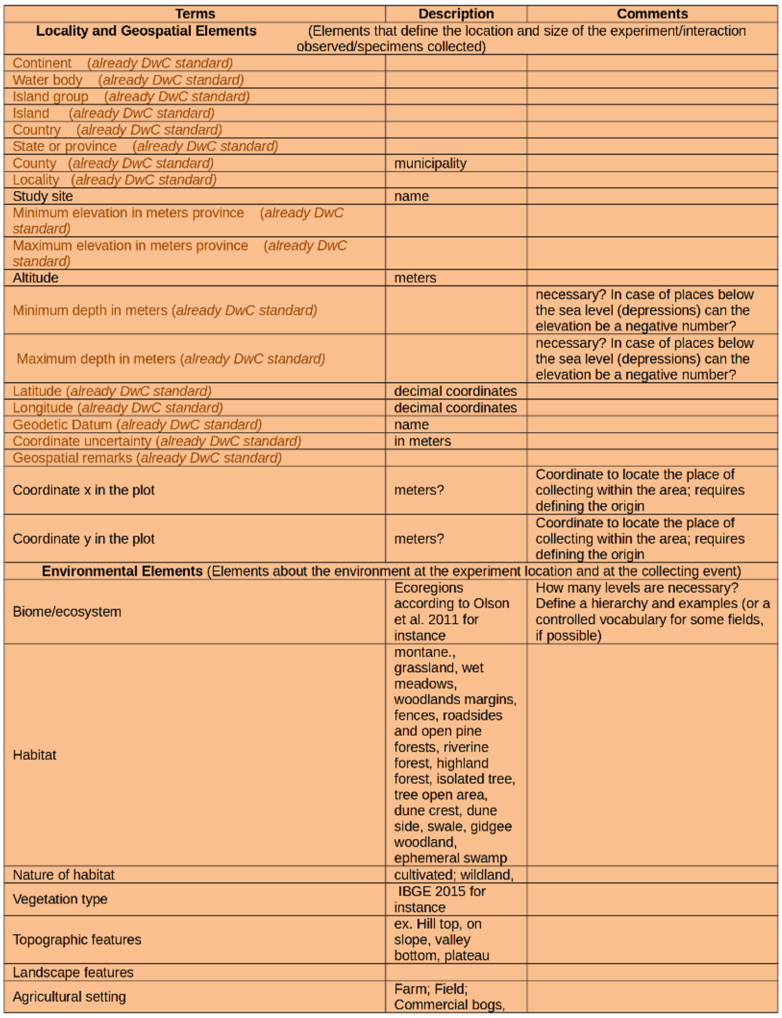
6. Dentre os descritores para PLANTAS quais aqueles que acredita serem mais adequados às suas necessidades de pesquisa? *

Check all that apply.

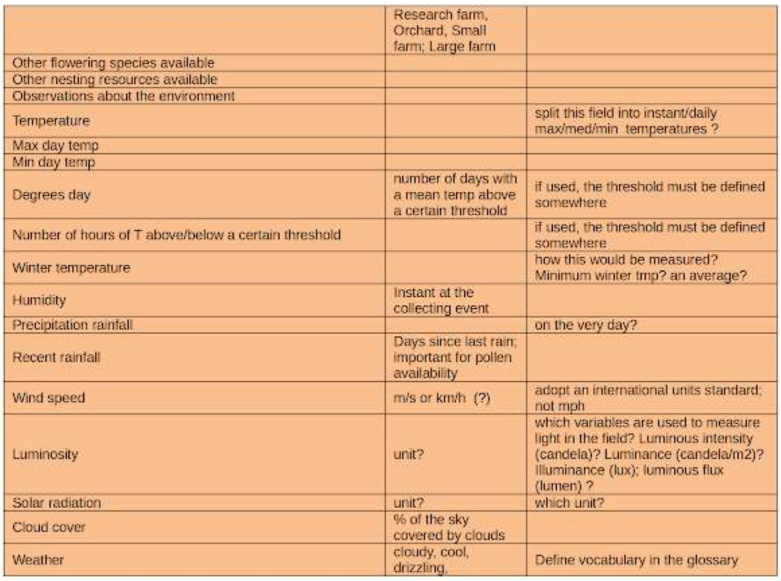
- □ Elementos relativos ao espécime coletado ou observado
- □ Elementos Taxonômicos
- □ Origem (nativa, exótica)
- □ Sistema reprodutivo (Breeding system)
- □ Forma de vida (life form)
- □ Tamanho
- □ Tipo de flor
- □ Simetria da flor
- □ Cor da flor
- □ Extensão da corola
- □ Comprimento do esporão (Spur length)
- □ Diâmetro da corola
- □ Deiscência da antera
- □ Detalhes da polínia
- □ Odor floral
- □ Tipo de floração
- □ Duração da floração
- □ Intensidade de floração
- □ Nome da população (Population name)
- □ Tamanho populacional
- □ Unidade amostral (sample unit)
- □ Número de indíviduos
- □ Densidade populacional (Plants/m2)
- □ Número de indivíduos florescendo na população (# Flowering plants)
- □ Densidade populacional de platnas florescendo (Flowering plants/m2)
- □ Número de flores
- □ Número de flores observadas
- □ Densidade de flores (Flowers/m2)
- □ Flores por planta
- □ Abundância floral
- □ Presença de néctar floral
- □ Variação do néctar na planta
- □ Detalhes do néctar
- □ Concentração do néctar
- □ Volume do néctar
- □ Planta cultivável
- □ Uso da planta
- □ Elementos de Referência (mesmos que da questão anterior)

□ Other: ______________________

Descritores Geográficos e Ambientais



Terms	Description	Comments
Locality and Geospatial Elements (Elements that define the location and size of the experiment/interaction observed/specimens collected)		
Continent *(already DwC standard)*		
Water body *(already DwC standard)*		
Island group *(already DwC standard)*		
Island *(already DwC standard)*		
Country *(already DwC standard)*		
State or province *(already DwC standard)*		
County *(already DwC standard)*	municipality	
Locality *(already DwC standard)*		
Study site	name	
Minimum elevation in meters province *(already DwC standard)*		
Maximum elevation in meters province *(already DwC standard)*		
Altitude	meters	
Minimum depth in meters *(already DwC standard)*		necessary? In case of places below the sea level (depressions) can the elevation be a negative number?
Maximum depth in meters *(already DwC standard)*		necessary? In case of places below the sea level (depressions) can the elevation be a negative number?
Latitude *(already DwC standard)*	decimal coordinates	
Longitude *(already DwC standard)*	decimal coordinates	
Geodetic Datum *(already DwC standard)*	name	
Coordinate uncertainty *(already DwC standard)*	in meters	
Geospatial remarks *(already DwC standard)*		
Coordinate x in the plot	meters?	Coordinate to locate the place of collecting within the area; requires defining the origin
Coordinate y in the plot	meters?	Coordinate to locate the place of collecting within the area; requires defining the origin
Environmental Elements (Elements about the environment at the experiment location and at the collecting event)		
Biome/ecosystem	Ecoregions according to Olson et al. 2011 for instance	How many levels are necessary? Define a hierarchy and examples (or a controlled vocabulary for some fields, if possible)
Habitat	montane., grassland, wet meadows, woodlands margins, fences, roadsides and open pine forests, riverine forest, highland forest, isolated tree, tree open area, dune crest, dune side, swale, gidgee woodland, ephemeral swamp	
Nature of habitat	cultivated; wildland,	
Vegetation type	IBGE 2015 for instance	
Topographic features	ex. Hill top, on slope, valley bottom, plateau	
Landscape features		
Agricultural setting	Farm; Field; Commercial bogs,	



	Research farm, Orchard, Small farm; Large farm	
Other flowering species available		
Other nesting resources available		
Observations about the environment		
Temperature		split this field into instant/daily max/med/min temperatures ?
Max day temp		
Min day temp		
Degrees day	number of days with a mean temp above a certain threshold	if used, the threshold must be defined somewhere
Number of hours of T above/below a certain threshold		if used, the threshold must be defined somewhere
Winter temperature		how this would be measured? Minimum winter tmp? an average?
Humidity	Instant at the collecting event	
Precipitation rainfall		on the very day?
Recent rainfall	Days since last rain; important for pollen availability	
Wind speed	m/s or km/h (?)	adopt an international units standard; not mph
Luminosity	unit?	which variables are used to measure light in the field? Luminous intensity (candela)? Luminance (candela/m2)? Illuminance (lux); luminous flux (lumen) ?
Solar radiation	unit?	which unit?
Cloud cover	% of the sky covered by clouds	
Weather	cloudy, cool, drizzling,	Define vocabulary in the glossary

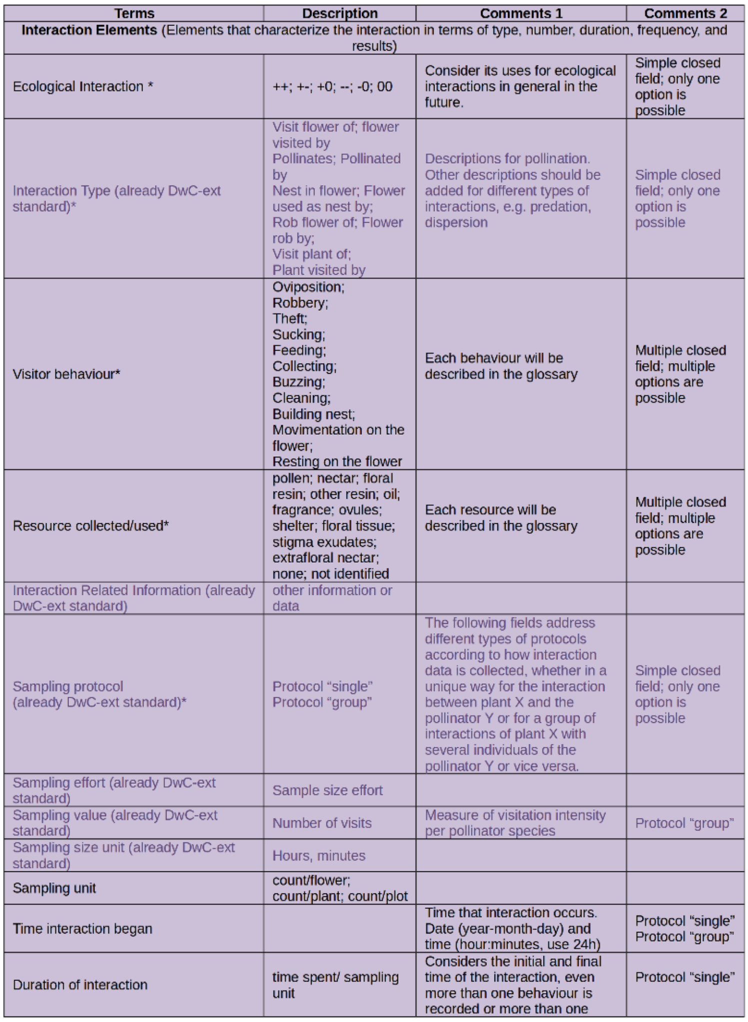
7. Dentre os descritores GEOGRÁFICOS e AMBIENTAIS quais aqueles que acredita serem mais adequados às suas necessidades de pesquisa? *

Check all that apply.

□ Elementos da Localização e Geoespaciais
□ Área de estudo
□ Altitude
□ Coordenadas <x,y> na parcela
□ Elementos Ambientais (considera detalhar melhor sua escolha também escolhendo os itens abaixo)
□ Bioma/Ecossistema
□ Habitat
□ Natureza do habitat
□ Tipo de vegetação
□ Características topográficas
□ Características da paisagem
□ Ambiente agrícola (Agricultural setting)
□ Outras espécies de flores disponíveis
□ Outros recoursos de nidificação disponíveis
□ Temperatura (incluindo máxima e mínima)
□ Degrees day
□ Número de horas em que a Temperatura esteve acima/abaixo de um limiar
□ Temperatura no inverno
□ Umidade
□ Precipitação
□ Chuva recente (Recent rainfall)
□ Velocidade do vento
□ Luminosidade
□ Radição Solar
□ Cobertura de nuvens
□ Condições Climáticas (Tempo/Weather)

□ Other: ____________________

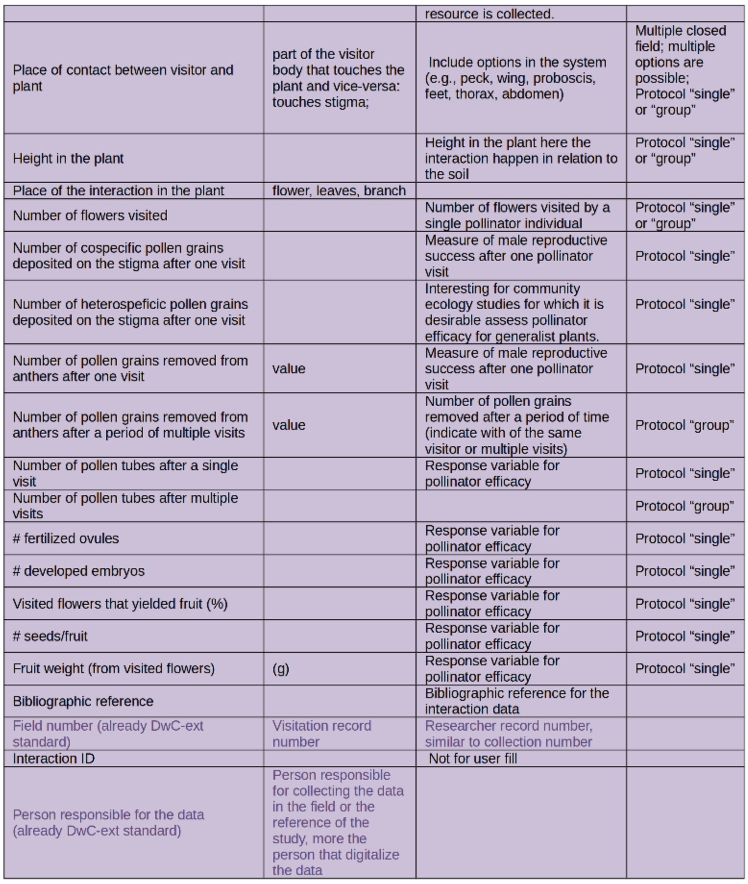
Descritores de Interação Ecológica (Parte 1 de 2)



Terms	Description	Comments 1	Comments 2
Interaction Elements (Elements that characterize the interaction in terms of type, number, duration, frequency, and results)			
Ecological Interaction *	++; +-; +0; --; -0; 00	Consider its uses for ecological interactions in general in the future.	Simple closed field; only one option is possible
Interaction Type (already DwC-ext standard)*	Visit flower of; flower visited by Pollinates; Pollinated by Nest in flower; Flower used as nest by; Rob flower of; Flower rob by; Visit plant of; Plant visited by	Descriptions for pollination. Other descriptions should be added for different types of interactions, e.g. predation, dispersion	Simple closed field; only one option is possible

Visitor behaviour*	Oviposition;
	Robbery;
	Theft;
	Sucking;
	Feeding;
	Collecting;
	Buzzing;
	Cleaning;
	Building nest;
	Movimentation on the flower;
	Resting on the flower
		Each behaviour will be described in the glossary	Multiple closed field; multiple options are possible

Resource collected/used*	pollen; nectar; floral resin; other resin; oil; fragrance; ovules; shelter; floral tissue; stigma exudates; extrafloral nectar; none; not identified	Each resource will be described in the glossary	Multiple closed field; multiple options are possible
Interaction Related Information (already DwC-ext standard)	other information or data		
Sampling protocol (already DwC-ext standard)*	Protocol "single" Protocol "group"	The following fields address different types of protocols according to how interaction data is collected, whether in a unique way for the interaction between plant X and the pollinator Y or for a group of interactions of plant X with several individuals of the pollinator Y or vice versa.	Simple closed field; only one option is possible
Sampling effort (already DwC-ext standard)	Sample size effort		
Sampling value (already DwC-ext standard)	Number of visits	Measure of visitation intensity per pollinator species	Protocol "group"
Sampling size unit (already DwC-ext standard)	Hours, minutes		
Sampling unit	count/flower; count/plant; count/plot		
Time interaction began		Time that interaction occurs. Date (year-month-day) and time (hour:minutes, use 24h)	Protocol "single" Protocol "group"
Duration of interaction	time spent/ sampling unit	Considers the initial and final time of the interaction, even more than one behaviour is recorded or more than one	Protocol "single"



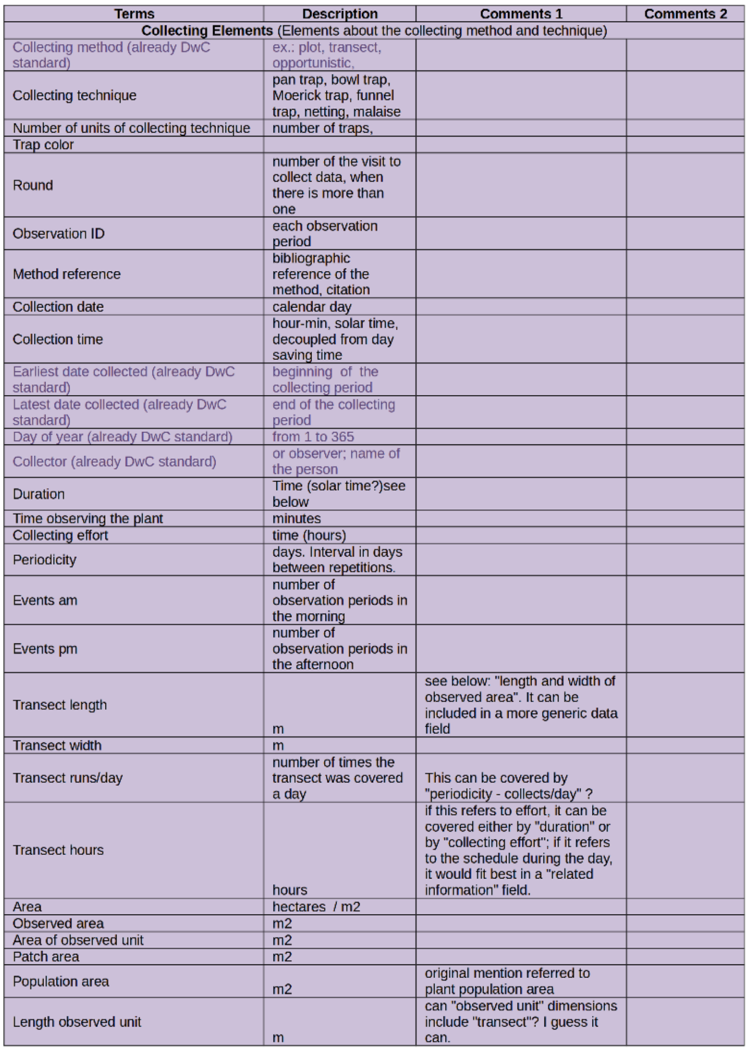
		resource is collected.	
Place of contact between visitor and plant	part of the visitor body that touches the plant and vice-versa: touches stigma;	Include options in the system (e.g., peck, wing, proboscis, feet, thorax, abdomen)	Multiple closed field; multiple options are possible; Protocol "single" or "group"
Height in the plant		Height in the plant here the interaction happen in relation to the soil	Protocol "single" or "group"
Place of the interaction in the plant	flower, leaves, branch		
Number of flowers visited		Number of flowers visited by a single pollinator individual	Protocol "single" or "group"
Number of cospecific pollen grains deposited on the stigma after one visit		Measure of male reproductive success after one pollinator visit	Protocol "single"
Number of heterospeficic pollen grains deposited on the stigma after one visit		Interesting for community ecology studies for which it is desirable assess pollinator efficacy for generalist plants.	Protocol "single"
Number of pollen grains removed from anthers after one visit	value	Measure of male reproductive success after one pollinator visit	Protocol "single"
Number of pollen grains removed from anthers after a period of multiple visits	value	Number of pollen grains removed after a period of time (indicate with of the same visitor or multiple visits)	Protocol "group"
Number of pollen tubes after a single visit		Response variable for pollinator efficacy	Protocol "single"
Number of pollen tubes after multiple visits			Protocol "group"
# fertilized ovules		Response variable for pollinator efficacy	Protocol "single"
# developed embryos		Response variable for pollinator efficacy	Protocol "single"
Visited flowers that yielded fruit (%)		Response variable for pollinator efficacy	Protocol "single"
# seeds/fruit		Response variable for pollinator efficacy	Protocol "single"
Fruit weight (from visited flowers)	(g)	Response variable for pollinator efficacy	Protocol "single"
Bibliographic reference		Bibliographic reference for the interaction data	
Field number (already DwC-ext standard)	Visitation record number	Researcher record number, similar to collection number	
Interaction ID		Not for user fill	
Person responsible for the data (already DwC-ext standard)	Person responsible for collecting the data in the field or the reference of the study, more the person that digitalize the data		

8. Dentre os descritores de INTERAÇÃO ECOLÓGICA quais aqueles que acredita serem mais adequados às suas necessidades de pesquisa? *

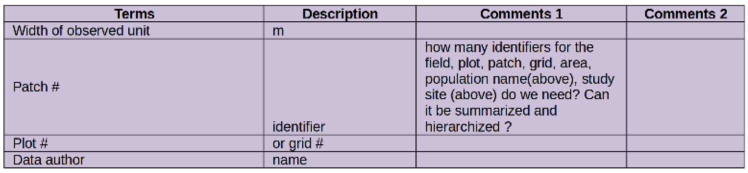
Check all that apply.

□ Interação ecológica (cardinalidade, ++, +-, +0, etc)
□ Tipo da interação
□ Comportamento do visitante
□ Recursos coletados/utilizados
□ Protocolo de amostragem (Sampling protocol)
□ Esforço amostral (Sampling effort)
□ Valor amostral (Sampling value)
□ Unidade do tamanho amostral (Sampling size unit)
□ Unidade amostral
□ Momento em que interação iniciou-se (Time interaction began)
□ Duração da interação
□ Local de contato entre o visitante e a planta
□ Altura da planta
□ Local de interação na planta
□ Número de flores visitadas
□ Número de grãos de pólen coespecíficos/heretoespecíficos depositados no estigma após visitação
□ Número de grãos de pólen removidos das anteras após visitação
□ Número de tubos polínicos após visitação
□ Número de óvulos fertilizados
□ Número de embriões desenvolvidos
□ Porcentagem de flores visitadas que geraram frutos
□ Número de sementes/fruto
□ Peso do fruto
□ Other: ____________________

Descritores de Interação Ecológica (Parte 2 de 2)



Terms	Description	Comments 1	Comments 2
Collecting Elements (Elements about the collecting method and technique)			
Collecting method (already DwC standard)	ex.: plot, transect, opportunistic,		
Collecting technique	pan trap, bowl trap, Moerick trap, funnel trap, netting, malaise		
Number of units of collecting technique	number of traps,		
Trap color			
Round	number of the visit to collect data, when there is more than one		
Observation ID	each observation period		
Method reference	bibliographic reference of the method, citation		
Collection date	calendar day		
Collection time	hour-min, solar time, decoupled from day saving time		
Earliest date collected (already DwC standard)	beginning of the collecting period		
Latest date collected (already DwC standard)	end of the collecting period		
Day of year (already DwC standard)	from 1 to 365		
Collector (already DwC standard)	or observer; name of the person		
Duration	Time (solar time?)see below		
Time observing the plant	minutes		
Collecting effort	time (hours)		
Periodicity	days. Interval in days between repetitions.		
Events am	number of observation periods in the morning		
Events pm	number of observation periods in the afternoon		
Transect length	m	see below: "length and width of observed area". It can be included in a more generic data field	
Transect width	m		
Transect runs/day	number of times the transect was covered a day	This can be covered by "periodicity - collects/day" ?	
Transect hours	hours	if this refers to effort, it can be covered either by "duration" or by "collecting effort"; if it refers to the schedule during the day, it would fit best in a "related information" field.	
Area	hectares / m2		
Observed area	m2		
Area of observed unit	m2		
Patch area	m2		
Population area	m2	original mention referred to plant population area	
Length observed unit	m	can "observed unit" dimensions include "transect"? I guess it can.	



Terms	Description	Comments 1	Comments 2
Width of observed unit	m		
Patch #	identifier	how many identifiers for the field, plot, patch, grid, area, population name(above), study site (above) do we need? Can it be summarized and hierarchized ?	
Plot #	or grid #		
Data author	name		

9. Dentre os descritores de INTERAÇÃO ECOLÓGICA quais aqueles que acredita serem mais adequados às suas necessidades de pesquisa? *

Check all that apply.

□ Elementos da coleta
□ Método de coleta
□ Técnica de coleta
□ Número de unidades da técnica de coleta
□ Round (número da visita que coletou o dado)
□ ID da Observação
□ Referência bibliográfico do método
□ Data/Horário da coleção
□ Datas/Horários de coleta
□ Coletor
□ Duração
□ Tempo observando a planta
□ Esforço de coleta
□ Periodicidade
□ Eventos (am/pm)
□ Dimensões do transecto
□ Área (total, observada, ocupada pela população)
□ Idenficadores de área, plots, patch, grades, etc.
□ Dados do coletor
□ Other: ____________________

Comentários adicionais (opcional)

10. Você considera que os descritores são suficientes para descrever as interações planta-polinizador? *

Mark only one oval.

○ Sim
○ Não

11. Os descritores são adequados para os tipos de dados usualmente coletados no seu grupo de pesquisa? *

Mark only one oval.

○ Sim

○ Não

12. Deixe seus comentários adicionais:

This content is neither created nor endorsed by Google.

Google Forms