WorldFAIR

Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D10.1 Agriculture-related pollinator data standards use cases report

Work Package	WP10 - Agricultural Biodiversity
Lead Author (Org)	Maarten Trekels (MeiseBG)
Contributing Author(s) (Org)	Debora Pignatari Drucker (Embrapa), José Augusto Salim (Unicamp), Jeff Ollerton (UoN/KIB), Jorrit Poelen (Global Biotic Interactions, Ronin Institute, UC Santa Barbara), Filipi Miranda Soares (USP & UTwente), Max Rünzel (HiveTracks), Muo Kasina (KALRO), Quentin Groom (MeiseBG), Mariano Devoto (UBA)
Due Date	31.07.2023

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

Date	30.07.2023
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.8176978

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.1	24.07.2023	All authors	Draft for internal review
0.2	28.07.2023	Christine Taliga, Joe Miller	Reviewed draft
0.3	29.07.2023	All authors	Version incorporating reviewer feedback
1.0	30.07.2023	Laura Molloy, Debora Drucker	Proof read and final edit

Disclaimer

WorldFAIR has received funding from the European Commission's WIDERA coordination and support programme under the Grant Agreement no. 101058393. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Abbreviations and Acronyms

CDIF	Cross-Domain Interoperability Framework
CORDIS	Community Research and Development Information Service
DDI	Data Documentation Initiative
DwC	Darwin Core
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile
GBIF	Global Biodiversity Information Facility
HEIs	Higher Education Institutions
IGAD	Improving Global Agricultural Data Community of Practice
URI	Uniform Resource Identifier
URN	Uniform Resource Name
REBIPP	Brazilian Network of Plant-Pollinator Interactions
RDA	Research Data Alliance
TDWG	Biodiversity Information Standards

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Executive Summary

Although pollination is an essential ecosystem service that sustains life on Earth, data on this vital process is largely scattered or unavailable, limiting our understanding of the current state of pollinators and hindering effective actions for their conservation and sustainable management. In addition to the well-known challenges of biodiversity data management, such as taxonomic accuracy, the recording of biotic interactions like pollination presents further difficulties in proper representation and sharing. Currently, the widely-used standard for representing biodiversity data, Darwin Core, lacks properties that allow for adequately handling biotic interaction data, and there is a need for FAIR vocabularies for properly representing plant-pollinator interactions. Given the importance of mobilising plant-pollinator interaction data also for food production and security, the Research Data Alliance Improving Global Agricultural Data Community of Practice[1] has brought together partners from representative groups to address the challenges of advancing interoperability and mobilising plant-pollinator data for reuse. This report presents an overview of projects, good practices, tools, and examples for creating, managing and sharing data related to plant-pollinator interactions, along with a work plan for conducting pilots in the next phase of the project.

We present the main existing data indexing systems and aggregators for plant-pollinator interaction data, as well as citizen science and community-based sourcing initiatives. We also describe current challenges for taxonomic knowledge and present two data models and one semantic tool that will be explored in the next phase. In preparation for the next phase, which will provide best practices and FAIR-aligned guidelines for documenting and sharing plant-pollinator interactions based on pilot efforts with data, this Case Study comprehensively examined the methods and platforms used to create and share such data. By understanding the nature of data from various sources and authors, the alignment of the retrieved datasets with the FAIR principles was also taken into consideration. We discovered that a large amount of data on plant-pollinator interaction is made available as supplementary files of research articles in a diversity of formats and that there are opportunities for improving current practices for data mobilisation in this domain. The diversity of approaches and the absence of appropriate data vocabularies causes confusion, information loss, and the need for complex data interpretation and transformation. Our explorations and analyses provided valuable insights for structuring the next phase of the project, including the selection of the pilot use cases and the development of a 'FAIR best practices' guide for sharing plant-pollinator interaction data. This work primarily focuses on enhancing the interoperability of data on plant-pollinator interactions, envisioning its connection with the effort WorldFAIR is undertaking to develop a Cross-Domain Interoperability Framework.

[1] https://www.rd-alliance.org/groups/igad-community-practice

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Table of contents

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

1. Introduction

As human demands on food, energy and other resources are increasing, global challenges related to climate change and biodiversity loss are becoming more complex to overcome. While the nature and the urgency of these challenges are increasingly acknowledged by societies and governments, their complexity is often daunting. To address such problems, it is imperative for scientists, decision-makers and societies to have access to data that can drive science-based decision-making. As an example, initiatives like the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) report (IPBES 2016) have demanded quick and reliable access to high-quality spatial and temporal data of species occurrences, their interspecific relations and the effects of the environment on biotic interactions.

Plant-pollinator interactions are recognised for their pivotal role in ecosystem functioning and sustainable agriculture. Within this context, plant-pollinator data becomes essential in addressing relevant questions such as the impact of domesticated bees on wild ecosystems, the contribution of wild pollinators to crop pollination, and the reciprocal effects of crops on these same wild pollinators. Understanding the behaviour of these organisms and its influence on their effectiveness as pollinators is also crucial. However, at present plant-pollinator data is scattered across various networks and country-specific initiatives, stored in isolated silos. Integrating this data at regional and global levels is crucial to enable pattern analysis and comprehension at biologically-relevant scales. Therefore, the adoption of community-accepted data standards on plant-animal interactions and the implementation of good practices, aligned with the FAIR principles, is urgently required.

It should be noted that much of the available data on plant-animal interactions pertains to whether the animal merely visits the flower. However, determining whether an organism qualifies as an effective pollinator entails painstaking and intricate work, often requiring field and laboratory studies. Consequently, our scope encompasses all interactions between animals and flowers. It is worth emphasising that even if an organism is not an effective pollinator, its visits to flowers may still impact pollination positively or negatively. Only by comprehending the full range of flower visitors can one hope to unravel the complexity of the plant-pollinator community.

The development and adoption of standards for biodiversity data and metadata have facilitated significant advances in biological data sharing and aggregation, supporting large-scale studies and science-based public policies (Wieczorek et al., 2012). Although large datasets of plant–pollinator interactions have become available worldwide, substantial challenges persist concerning data storage and standardisation. Presently, the existing standards are not entirely suitable to fully support the sharing of biotic interactions data. Addressing these issues is imperative to enable the development of integrative studies that broaden our understanding of species biology, phenology, and evolution. Moreover, it is vital to support the decision-making process for pollinator conservation (Salim et al, 2022a).

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Recognising the significance of enabling data sharing and reuse for this critical ecosystem service, the WorldFAIR Case Study on Agricultural Biodiversity[2] is committed to promoting the adoption of standards for plant-pollinator data and their improvement. This report presents the discovery phase of the Case Study, offering an overview of data standards in studies of pollinators and their interactions with agricultural crops and plants in the broader ecosystem. Additionally, the report includes a work plan for the pilots to be adopted in the project's next phase. This effort will establish concrete guidelines, which are currently lacking, on standards for pollination data, metadata and other digital objects. Furthermore, it will contribute to their adoption by active or planned public and private initiatives, while also connecting with the Cross-Domain Interoperability Framework (CDIF)[3] and the Unified Data Model proposed by the Global Biodiversity Information Facility (GBIF), which leads the Case Study on Biodiversity.

1.1. Partner overview and the Research Data Alliance Improving Global Agricultural Data Community of Practice (IGAD CoP)

The Agricultural Biodiversity Case Study was conceived within the RDA Improving Global Agricultural Data Community of Practice (IGAD CoP). IGAD identified initiatives related to plant-pollinator data standardisation being carried out by different groups within the community and assembled a representative team to collaborate on enabling FAIR data for this key ecosystem service vital for agriculture and many other mechanisms that sustain life on Earth. The team is led by Embrapa (Brazilian Agricultural Research Corporation) and gathers several initiatives and institutions, namely the Biodiversity Information Standards (TDWG) Biological Interaction Data Interest Group, the other four direct beneficiaries institutions: Meise Botanic Garden, HiveTracks, Kenya Agricultural and Livestock Research Organization (KALRO) and the RDA Association, and a large network of organisations contributing in-kind and not receiving funding: University de São Paulo (Brazil), University of Campinas (Brazil), USDA Agricultural Research Service (USA), Agriculture and Agri-food Canada (AAFC - Canada), Federal University of São Carlos (Brazil), University of Buenos Aires (Argentina), Global Biotic Interactions (GloBI), UK Centre for Ecology & Hydrology (UKCEH), and the Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Meise Botanic Garden[4] is located just north of Brussels in Belgium and is a research institution of the Flemish Government. It has large living and preserved collections of plants, and scientists at the Garden investigate many aspects of plant and fungal biology. They are notable for the work they do on the taxonomy and conservation of plants, particularly in Northern Europe and Africa. The Biodiversity Informatics team of the Garden is involved in WorldFAIR; this team works on the digitisation, dissemination, standardisation and analysis of biodiversity data. The key focuses are on data on invasive species and their impacts, on data from citizen science, and the analysis of biodiversity data for policy support. The Garden recently completed a mass digitisation of its collections, making data and images of specimens from several million objects available. The

[2] https://worldfair-project.eu/agricultural-biodiversity/
[3] https://worldfair-project.eu/cross-domain-interoperability-framework/
[4] https://www.plantentuinmeise.be/en/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Biodiversity Informatics team is directly involved with further improvement of these data, linking them with other information and ensuring that they continue to be a valuable resource for researchers globally.

Embrapa[5] is the Brazilian Agricultural Research Corporation, established by the country's federal government in 1973 to develop the technological foundation for a genuinely tropical model of agriculture and animal farming. Embrapa today is one of the largest agricultural research corporations in the world, with half a century of contributions focused on innovation, efficiency, sustainability and social inclusion. Embrapa is part of the Brazilian Network of Plant-Pollinator Interactions (REBIPP) coordination initiative and has a long history of research in this field. The organisation is involved with several initiatives related to FAIR data in agriculture in Brazil and internationally.

KALRO[6] is the Kenyan government agency responsible for agricultural research. The agency is semi-autonomous, having been created in 2014 as a result of merging four agencies previously responsible for agricultural research. The KALRO Apiculture and Beneficial Insects Research Institute is responsible for generation and promotion of technologies, innovations and management practices across apiculture and beneficial insects value chains. Plant-pollinator interactions form a major programme in the institute. KALRO partners with local, regional and global institutions in an effort to deliver its mandates.

HiveTracks[7] is a USA-based startup that works with smallholder beekeepers to crowdsource environmental data collection. Since 2010, its technology has empowered beekeepers to improve the health of their hives across 150 countries. The HiveTracks App allows smallholder beekeepers to keep track of their management practices and send rules-based recommendations based on hive health-related observations. Its web portal allows extension officers to monitor beekeeping practices remotely and to send context-specific recommendations to improve beekeeping practices and over-winter survival rates and honey production capabilities. In addition, beekeepers enter information on the available floral resources around the hive and receive weather information based on their unique hive location. Since the launch of the new HiveTracks App in February 2023, over 2,000 beekeepers have added 7,000 unique management practices, including 4,300 unique hive inspections.

The RDA Association[8] is a non-profit international organisation with the overall goal to support Research Data Alliance (RDA) activities in Europe, and to take a part in global RDA development. The RDA Association consulted Jeff Ollerton for expert input. Jeff is an ecological scientist and author with particular interests in the ecology, evolution and conservation of plant-pollinator interactions. In addition to running his own consultancy, he holds visiting professorships at the University of Northampton (UK) and the Kunming Institute of Botany (China), and is a visiting

[5] https://www.embrapa.br/international
[6] https://www.kalro.org/
[7] https://www.hivetracks.com/
[8] https://www.rd-alliance.org/rda-europe

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

lecturer at Roskilde University (Denmark). Jeff has written more than 150 scientific papers, essays and popular articles, and a major book (*Pollinators & Pollination: Nature and Society*, 2021). His next book - *Birds & Flowers: An Intimate 50 Million Year Relationship* - will be published later in 2023.

Many of the partners are connected to the Biodiversity Information Standards (TDWG) Biological Interaction Data Interest Group[9], which has the objective to discuss and formalise biodiversity interaction data aiming at developing a data standard under the TDWG umbrella.

1.2. Why pollination is crucial for biodiversity, agriculture and sustaining life on Earth

Agricultural crops and wild-collected edible fruits and seeds vary considerably in their dependence on pollinators such as bees, flies, and other insects as well as vertebrates like bats and birds. This dependence ranges from zero in the case of wind-pollinated cereals (e.g., rice and wheat) to 100% for cacao and some other tropical crops.

Globally, we know from a review by Klein et al. (2007) that three quarters of the 115 most productive crops require animal pollinators to some extent, accounting for 5% to 8% of global yields, estimated to be worth US$235 to US$577 billion per year (IPBES, 2016). As new crops are developed for the market and there are shifts in consumer tastes for particular foods, reliance on pollinators in agriculture is predicted to increase (Ollerton, 2021).

Even more importantly than overall yield and calorific value, animal-pollinated fruit and seeds make a disproportionate contribution to healthy diets, providing large amounts of essential minerals, vitamins and antioxidants (Ellis et al., 2015). This is especially important in tropical developing countries where a high proportion of animal-pollinated crops are grown by subsistence farmers or collected from local natural areas, such as *Euterpe oleracea* - Açaí (Campbell et al., 2018; Campbell et al., 2023). It is not uncommon for these farmers to include some small-scale beekeeping on their land in order to supplement their incomes with honey and wax production. A decline in pollinators has impacts on seed and fruit production of flowering plants (Rodger et al., 2021) and could have a major impact on the health of people in these parts of the world, pushing populations toward the risk of malnutrition (Eilers et al., 2011; Brittain et al,. 2014; Chaplin-Kramer et al., 2014; Smith et al., 2015). The loss of pollinators could also affect their farm economies. For example, coffee is grown mainly by small-scale tropical farmers; at the moment the global coffee crop, which is worth more than US$80 billion per year, relies on approximately 25 trillion flower visits, mainly by wild and managed bees (Ollerton, 2021).

Open pollination may also help landraces of crops to maintain their genetic diversity potentially making them more resilient to changing conditions, disease and extreme events (Lopes et al., 2015).

In order to fully understand the relationships between pollinators and the food crops that sustain us, we require interoperable data on the networks of interactions between flower-visiting animals

[9] https://github.com/tdwg/interaction

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

and the flowers of the crops, and those of the plants in the wider environment that sustain those pollinators when the crops are not in flower.

1.3. Discovery phase approach

The discovery phase was focused on collecting a broad variety of data sources of plant-pollinator interactions. The data landscape of plant-pollinator information is highly fragmented. Therefore, several approaches were used to collect as many data sources as possible. The main distinction can be made between datasets that are already indexed by data aggregators (such as Zenodo) and data that is still locked up in their silos, or shared as supplementary materials attached to research articles.

To facilitate the discovery of 'hidden' data, within WorldFAIR task 10.1 we organised monthly meetings that involved many of the potential stakeholders (e.g. via the IGAD CoP of the RDA). We attempted to increase the reach of this effort through sessions in relevant conferences such as RDA plenaries[10], SciDataCon[11], TDWG[12], SCAPE[13] and The US National Native Bee Monitoring Research Coordination Network (RCN) workshop[14] in the USA. KALRO conducted a specific survey in the African continent gathering information on plant-pollinator interactions from African sources. This resulted in a list of potential resources (Trekels, 2023).

The analysis of indexed data sources was performed slightly differently. Using a keyword search, a landscape analysis was conducted on the Dimensions, Zenodo, Dryad, Figshare and DataOne portals. Since these datasets have metadata associated with them, it was possible to have a more detailed view of the data content. The results on this search are presented in chapter 3 of this report.

Considering the importance of narrowing the science-industry bridge for the benefit of society, we also included data from HiveTracks in our approach, sourced from a global community of smallholder beekeepers. The standardised data entry process - accessible through the HiveTracks App for beekeepers - has been piloted[15] across Central Asia, the Middle East, East Africa, Europe and North America, and refined throughout the discovery phase to provide an efficient user interface for beekeepers and to determine the criteria for the pilot with the goal of providing an inclusive end-to-end process for community-based sourcing of FAIR managed pollinator data that protects beekeepers' privacy, whose livelihoods depend upon the location and positioning of their hives.

[10] https://www.rd-alliance.org/
[11] Challenges for Plant-Pollinator Data Standardization https://www.scidatacon.org/IDW-2022/sessions/418/
[12] Drucker et al. 2022
[13] https://scape-pollination.org/
[14] http://nativebeemonitoring.org/
[15] AI-Driven Climate-Smart Beekeeping for Women Project Report: https://repo.mel.cgiar.org/handle/20.500.11766/67516

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

1.4. FAIR Implementation Profile (FIP)

Since the FAIR principles were published (Wilkinson et al., 2016), different approaches were developed to perform FAIR assessments of a particular scientific field. The FIP methodology[16] was developed to make standardised assessments of practices within a scientific community and developing the first FAIR Implementation Profile (FIP) for the Agricultural Biodiversity Case Study in September 2022 was an important step to map out best practices for FAIR plant-pollinator data. We based the first FIP on the most prominent initiatives for standardising species interactions, which will be explored in more detail in chapter 2: GloBI (Poelen et al., 2014), Darwin Core (Wieczorek, 2012) and GBIF (2020). We discovered that the majority of the FAIR (sub)principles are enabled by at least one of those three resources, and that the main challenges are related to consistent adoption and convergence. In other words, there are alternative pathways within the existing resources to enable FAIRness, and there are opportunities to explore and propose specific guidelines to encourage the FAIR standards adoption for plant-pollinator data. Lessons learnt were discussed with WorldFAIR partners[17], aiming to contribute to the Cross-Domain Interoperability Framework. In a later phase, the FIP exercise will be repeated using the results of the discovery and pilot phases in the WorldFAIR project.

2. Data and standards

2.1. Existing data indexing systems and aggregators

The Global Biodiversity Information Facility (GBIF) has indexed more than 2 billion occurrence records from 85,387 datasets. While many of these datasets contain valuable biotic interactions data (Salim et al., 2022b), their availability in the GBIF portal or API is limited. This is primarily due to the constraints imposed by the current data model used by GBIF and its indexing system, which does not fully accommodate the representation and integration of biotic interactions data. However, GBIF is currently developing a new data model known as the “Unified Data Model”[18]. This model is specifically designed to address these limitations, enabling the integration of diverse use cases, including biotic interactions, which will be detailed in section 2.4 below. Within the WorldFAIR Project, a detailed overview of the model is provided by Miller et al. (2023).

On the other hand, the Global Biotic Interactions database (GloBI)[19] is the main source for discovering and accessing biotic interactions data. GloBI provides open access to biotic interaction data by combining existing open datasets using open source software. GloBI continuously scans existing data infrastructures and registries, enabling it to track, resolve and integrate biotic interactions data that these sources make available. Instead of functioning as a centralised

[16] https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/
[17] https://codata.org/events/conferences/fair-convergence-symposium-2022/fips-in-worldfair-what-have-we-learnt/
[18] https://www.gbif.org/new-data-model
[19] https://www.globalbioticinteractions.org/process

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

repository, GloBI serves as a search index that facilitates the discovery of species interaction datasets within their respective native cyber-habitats. To effectively handle data in various formats, GloBI employs a mapping mechanism between fields in the dataset to terms in GloBI's vocabulary[20].

Currently, GloBI actively tracks data sources such as GitHub and Zenodo for datasets which use its template for data sharing and annotation. Additionally, GloBI tracks observations with interaction types recorded in the observation fields[21] of the iNaturalist platform. iNaturalist is a worldwide social network of naturalists, citizen scientists and biologists who document and share biodiversity observations and will be further described in section 2.2. GloBI has successfully indexed over 15 million biotic interaction records by November 2022 (GloBI Community, 2022), of which 256,420 (1.6%) are plant-pollinator interactions (i.e. interaction types "visits flowers of" and "pollinates").

2.2. Citizen science and community-based sourcing initiatives

Citizen science is of high importance in the natural sciences, and is generating a huge amount of information on biological organisms and processes (Vohland, 2021). Also, interactions between species are easily recorded by citizens through some of the most common platforms. One of the most important of these platforms is iNaturalist (iNaturalist, 2023).

iNaturalist is an online social network and crowdsourcing platform for biodiversity enthusiasts and citizen scientists. The main purpose of iNaturalist is to facilitate the documentation and sharing of observations of plants, animals, and other organisms. Users can upload photos and other relevant information about the organisms they encounter, and the community helps with identifying and providing information about the species. It serves as a platform for people to connect with fellow nature enthusiasts, learn about different species, and contribute to scientific research and conservation efforts.

iNaturalist has a website and mobile app that make it easy for users to record and submit their observations. The platform utilises image recognition software — Computer Vision Model (Ueda, 2020) — and community expertise to help identify the species in the uploaded photos. It also encourages users to engage in discussions, participate in projects, and collaborate with researchers and organisations.

The web-based version of iNaturalist enables users to record biotic interactions by including specific metadata. However, it's important to note that only observations that have registered these interactions on iNaturalist are tracked by GloBI. It is worth pointing out that interactions data on iNaturalist are a rare source of interaction observations, as opposed to summary data. Such data are invaluable to investigate the phenology and diurnality of interactions, which is necessary to understand the impacts of climate changes on interactions, and the potential influence on flower visitors with each other.

[20] https://github.com/globalbioticinteractions/template-dataset#data-format-and-dictionary
[21] https://www.inaturalist.org/observation_fields

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Initiatives like iNaturalist demonstrate the strengths of a community or crowd-based approach around data collection. While several of these public initiatives exist, there continues to be a lack of private sector or industry-led initiatives. Considering the availability of resources and technological tools being used by thousands of farmers around the world, the private sector has to be incentivised and inspired to keep data FAIR. A lack of knowledge, experience and collaboration further deepens the science-industry gap. Additional case studies have to be developed and made available to showcase how small and large companies can gather FAIR plant-pollinator interaction data in particular, following best practices regarding data sharing and privacy protection.

2.3. Taxonomic data

Taxonomy adds an additional layer of complexity to any biological science, but particularly in species interactions. Interactions occur between individuals or groups of organisms, and humans classify those organisms into taxa with common features. However, these taxa are human constructs and natural systems do not always separate into neat groups, and people do not always agree on where the boundaries of those groups should be. Yet it is nevertheless essential to make predictions about the interactions between organisms to know which species are influencing each other.

Binomial nomenclature was introduced by Linnaeus (Linnaeus, 1758) and the system is still used today to identify and categorise life on Earth. Taxonomic names or Linnean taxonomy are the key to biodiversity knowledge, and help us to make sense of the natural world. Taxonomic names - e.g., *Homo sapiens* (Linneas, 1758) - are references to a description of a species as published by some authority in taxonomy. Over the years, the number of described species in taxonomic treatments has not only grown in volume to many millions of described species, but has also shown to be prone to errors, duplication, and revision. For instance, what is considered a single species today may have been independently described in taxonomic treatments by different scientists, resulting in multiple names for a single species (e.g., *Arius felis* is a synonym of *Ariopsis felis*, a catfish species). These inconsistencies in taxonomic literature are described in taxonomic revisions, in which authors argue for a particular interpretation and describe a way to update the taxonomic record. Also, with advancements in molecular sequencing, phylogenetic studies have advanced tremendously in the last two decades.

As a consequence, whole families and other taxonomic ranks have shifted, split or merged, and that affects taxonomic names as well. In the early 2000s, for example, this led to the integration of two existing families (Apocynaceae *sensu stricto* and Asclepiadaceae) into a single family Apocynaceae *sensu lato*. This newly-defined family is much larger, and with over 5300 species is in the top 12 of angiosperm families by size. At lower taxonomic levels there have also been major changes and reclassifications, for example within the genera *Cynanchum* and *Ceropegia*. This is especially relevant in relation to this Case Study, as Apocynaceae has now become a model taxon for understanding the diversity and evolution of pollination systems, backed up by a large and continuously updated database of plant-pollinator interactions (Ollerton et al., 2019). As scientific

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

research continues, further refinements to angiosperm classification may occur to better reflect their evolutionary relationships.

In an effort to catalogue taxonomic literature and provide a comprehensive interpretation of available taxonomic descriptions, taxonomic backbones are made available. Examples include: the GBIF backbone taxonomy (GBIF Secretariat, 2022), the Catalogue of Life (CoL) (Bánki et al., 2023), Integrated Taxonomic Information System ITIS (National Museum of Natural History, Smithsonian Institution, 2023), and the World Register of Marine Species WoRMS (WoRMS Editorial Board, 2023). Owing to the complexities of taxonomic literature, many different subjective interpretations exist of our highly dynamic corpus of scientific taxonomic knowledge published in many languages across many scientific journals, books, and websites (Upham et al., 2021). More recently, CoL and GBIF united their efforts to provide a repository to share checklist data (ChecklistBank[22]).

So while taxonomic names help us to understand the world around us, the context in which the names appear (e.g., a publication in the 1980s) and the context in which the names are interpreted (e.g., an ecologist in 2023) both determine how this name is interpreted. For instance, a bat species *Tadarida pumila* found in an article from the 1980s[23] is interpreted as *Mops pumilus* by Nancy Simmons instead. However, when using available catalogues, results were inconsistent in using the Nomer Corpus of Taxonomic Resources (Poelen, 2023). The Mammal Diversity Database, ITIS, NCBI Taxonomy, and Catalogue of Life do not recognize the name *Tadarida pumila*, whereas the GBIF taxonomic backbone considers the *Tadarida pumila* as synonym of both *Chaerephon pumilus* (Cretzschmar, 1826), as well as *Chaerephon pusillus* (Miller, 1902). In other words, care must be taken in interpreting taxonomic names used in literature or datasets.

In addition to temporal changes in the interpretation of taxa and differences between different taxonomic backbones, major regional differences also exist in the interpretation of species names. Information related to species interactions is often focussed on a particular area of study and at best matched to a local checklist of species. Even more often, the information is collected by non-specialists using vernacular names of species. These names are not unique and can point to different species in different languages. Therefore, it is also important to involve local knowledge in the identification of species. At the very least, the verbatim names should be included as well as a description of the taxonomic name alignment process and associated taxonomic authorities. Taxonomic interpretations should be documented when possible (e.g., the name is interpreted as *Mops pumilus* as suggested by Nancy Simmons in 2023), as well as the specific version of the taxonomic resources used. Ideally, names are aligned across various taxonomic resources to get a sense of the variability of the name interpretations. In a Big Bee Name Alignment Workshop (Miller et al., 2023), automated methods were outlined that were capable of automatically aligning across many version-controlled taxonomic resources[24]. This, or similar, name alignment processes can be used to review names in biodiversity datasets to assess how well-observed names link to commonly

[22] https://www.checklistbank.org/
[23] https://github.com/jhpoelen/bat-taxonomic-alignment/issues/20
[24] https://big-bee-network.github.io/name-alignment-workshop

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

used taxonomic name resources. Consequently, the interoperability (or reusability) of a dataset can be quantitatively and automatically assessed.

2.4. Data models and standards

Darwin Core (DwC) stands as a standardised set of terms and definitions, commonly adopted by the biodiversity community, to describe biodiversity observation data (Wieczorek et al., 2012). Observation data consist at their heart of the “what”, “where”, “when” and “who” of an observation, whether that is the observation of a single organism, a group or of an interaction. The “what” is typically the species involved. The “where” is the geographic location (locality name, country, latitude, longitude). “When” is the time and date that the event occurred. And the “who” provides provenance information and is usually the name of the person who made the observation. DwC also effectively manages other associated information giving context to the observation, but their consistent usage within the community remains variable.

DwC is not a static standard; it is under active maintenance by the Biodiversity Information Standards organisation (TDWG)[25]. GitHub is used to maintain the standard and its documentation, including keeping track of, and debating changes to the standard[26].

However, when it comes to pairwise biotic interaction, the complexity increases. Such interactions involve two occurrences (co-occurrence) of taxonomically homogeneous organisms engaging in a coaction at a specific place and time (Lidicker, 1979). This dynamic nature of species interactions data poses challenges to its representation, leading to diverse ways of utilising DwC elements. For instance, the term “associatedTaxa” is commonly used to represent an interaction, yet the “ResourceRelationship” class offers a more detailed approach to interaction representation (Enetwild consortium et al., 2022). This flexibility highlights the need for clear guidelines and a standardised approach to representing biotic interaction data within the DwC framework.

In the pursuit of refining biodiversity data standards, TDWG plays a crucial role. TDWG, through its dedicated task groups and community-driven efforts, aims to establish best practices and guidelines for data representation, sharing, and integration. By fostering collaboration and consensus-building among experts and stakeholders, TDWG contributes to the development and refinement of data standards like DwC, including its extension to address the challenges of representing biotic interactions. Through TDWG's continuous efforts, the biodiversity community can ensure that standardised data formats effectively capture the dynamic and intricate relationships that shape our understanding of plant-pollinator interactions and other critical ecological processes.

While the taxonomic, spatial, and temporal information about the occurrences of such organisms or groups of organisms can be documented following the same approach used for occurrence data, there is no formal or recommended process to express the association of such occurrences and the particularities of their interaction. According to Jordano (2021) any interaction is composed of three

[25] https://www.tdwg.org/
[26] https://github.com/tdwg/dwc

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

basic components: the co-occurrence, the encounter, and the outcome. To allow more efficient data aggregation and analysis, these three components should be properly documented (Salim et al., 2022a).

Biotic interactions, when explored beyond the scope of "tetranomials", i.e. a concatenation of the two Latin binomials (Jordano, 2022), usually include data that cannot be adequately represented by the DwC standard as it currently stands because the standard lacks appropriate terms for documenting data in detail (e.g., organisms' traits, interaction effects and outcomes). Also, currently there is no existing formal data model that adequately captures the important components of the phenomena, such as the type, direction, effects, and outcomes of an interaction.

In Salim et al. (2022a), a data model for the DwC standard was proposed, based on the needs of the Brazilian Network of Plant-Pollinator Interactions[27] (REBIPP) community. This data model, illustrated in Figure 1, takes into account the discussions and considerations within the TDWG Biological Interactions IG[28].

[27] https://www.rebipp.org.br
[28] https://github.com/tdwg/interaction

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

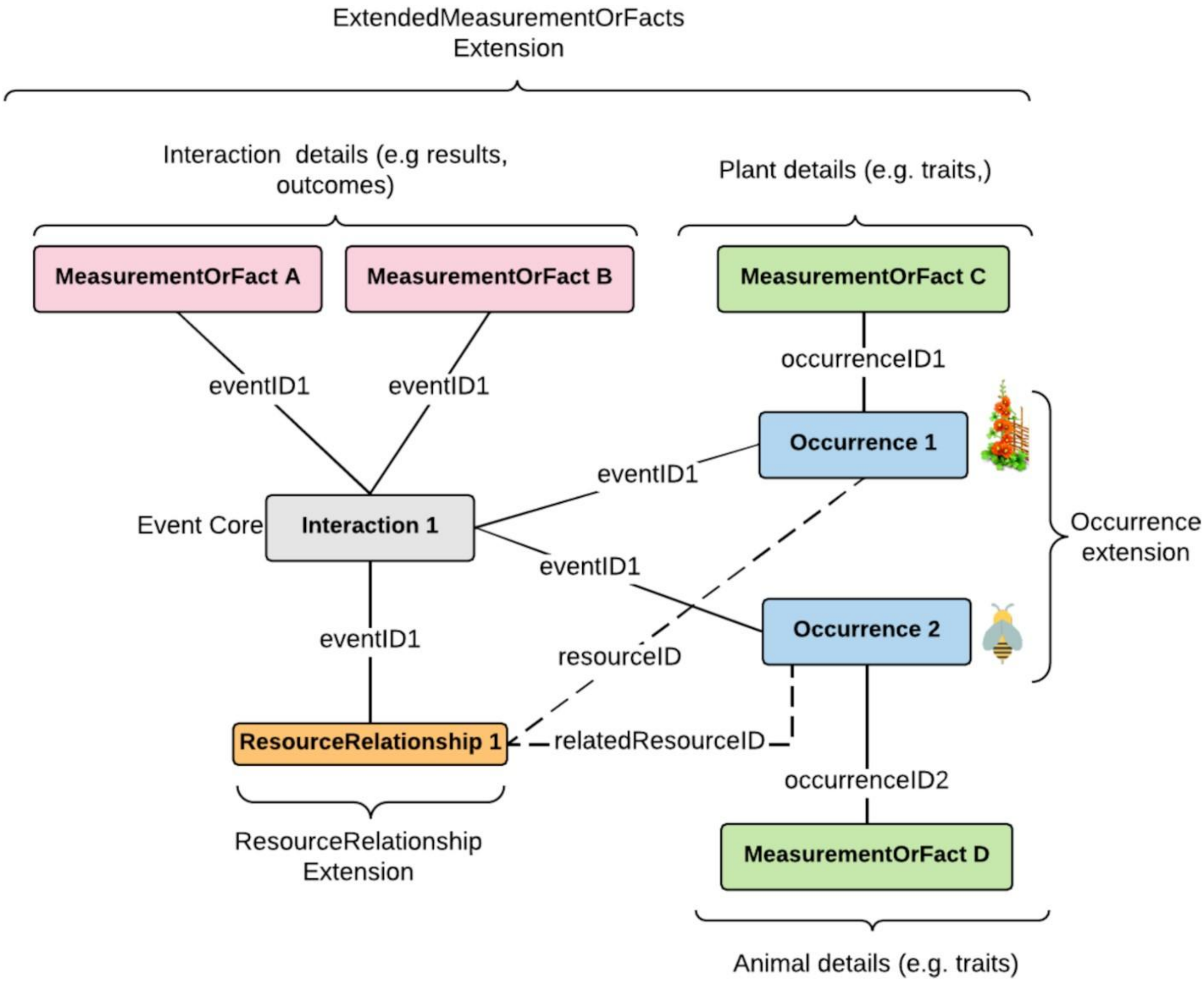

Figure 1. Overview of the data schema to represent plant–pollinator interactions. The plant and animal occurrences (blue boxes) are linked to the interaction (dwc:Event, grey box) using the dwc:eventID (full lines). Measurements related to the interactions (MeasurementOrFact A and MeasurementOrFact B, pink box, e.g., ppi:resourceCollected, ppi:nectarCollectingBodyPart, ppi:numberOfRemovedPollenGrains) are linked directly to the interaction (grey box) using the dwc:eventID. Measurements related to the occurrences (MeasurementOrFact C and MeasurementOrFact D, green boxes, e.g., ppi:flowerColor, ppi:floralAttractants, ppi:caste) are linked to the interactions using the dwc:eventID and the dwc:occurrenceID fields of the obis:ExtendedMeasurementOrFact extension (dashed lines). The direction and the type of the interaction are given by the dwc:ResourceRelationship class (orange box), linked directly to the interaction using dwc:eventID (full lines) and indirectly to the occurrences using dwc:resourceID and dwc:relatedResourceID terms (dashed lines). Source: Salim et al 2022a.

Additionally, Salim et al. (2022a) have developed a plant-pollinator interactions vocabulary (PPI)[29] - a standardised vocabulary maintained by REBIPP. It is designed to be used with DwC data standard,

[29] https://ppi.rebipp.org.br

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

specifically as controlled vocabulary for terms from the dwc:MeasurementOrFact class. PPI also incorporates several controlled vocabularies for the standardisation of measurement values.

The authors suggest that the data model presented in Figure 1 has the potential to be expanded to accommodate a broader range of use cases involving general biotic interactions, aligned with the concepts being developed in the GBIF "Unified Data Model". Figure 2 shows the current state of the conceptual model proposed by GBIF for Biotic Interactions.

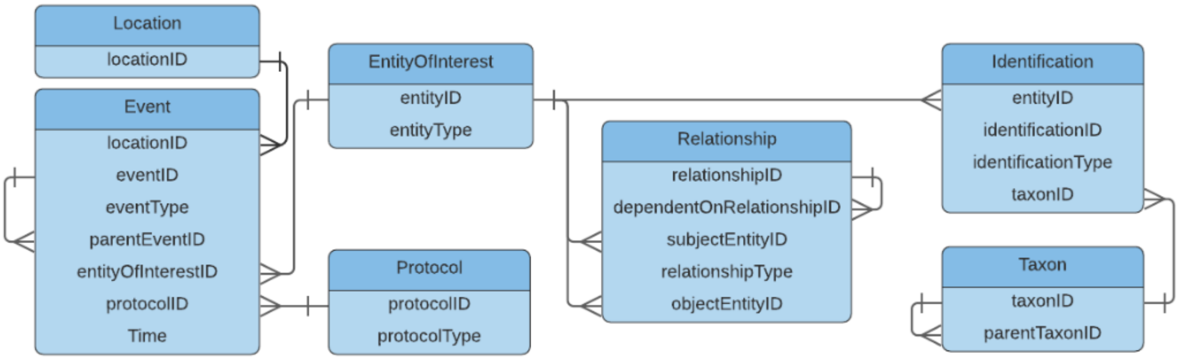

Figure 2. Conceptual model proposed by GBIF for biotic interactions[30] (GBIF, 2023).

Although Figure 2 depicts the conceptual model for biotic interactions data, in contrast to the implementation model shown in Figure 1 for the plant-pollinator interactions data, some similarities can be observed between the two. First, both models are centred at the dwc:Event class to represent an interaction. The recently released version of DwC includes a new term called dwc:eventType, specially designed to handle such scenarios. In this particular case, the dwc:eventType can be filled with a value such as "Interaction" to indicate that the event represents a biotic interaction. In addition, in both models, the interacting organism (represented in the "Unified Data Model" as EntityOfInterest and in the PPI as dwc:Occurrence) are linked using a Relationship entity, which in the PPI translates to the dwc:ResourceRelationship class. This enables the representation of both the direction and type of the interactions.

One crucial element which is currently missing in the "Unified Data Model" is the documentation of interaction outcomes and effects. In the PPI, this is addressed through the adoption of dwc:MeasurementOrFact and its extensions for documenting interaction outcomes and effects, as well as the inclusion of organisms' traits. Hopefully, this work and others (e.g. Salim et al., 2022a; Kita et al., 2022; Keller et al., 2023; Schneider et al., 2019) will contribute to the improvement of the "Unified Data Model" toward a more realistic and context-dependent representation of biotic interactions (Gomez et al., 2023; Hoeksema et al., 2010; Butterfield et al., 2013; Chamberlain et al., 2014; Maron et al., 2014; Hoeksema et al., 2015; Frederickson, 2017; Roy et al., 2016). The other

[30] https://docs.google.com/document/d/1jzb54GbAkB_TOFIjWof5BW6gn1ujSnXXJQu6aZgFAB4/edit

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

elements of the models (e.g., Location, Taxon) are common to all use cases, not only those related to biotic interactions.

In the “Unified Data Model”, an “Event” such as a specimen from a particular time and place can be linked to an “Assertion” about that specimen. This Assertion could be a measurement of the specimen, or the use of the specimen in a publication, or its interaction(s) with other species. An example from the family *Apocynaceae* mentioned above is pollinators that were extracted from the flowers of *Ceropegia* species preserved in spirit in the Herbarium of the Royal Botanic Gardens, Kew, England. These pollinators have been identified and could be linked back to the original plant specimens via an Assertion, which is the publication of this information. Such cases are rare, however, and in this case predicated on the fact that *Ceropegia* flowers temporarily trap their pollinators. Most studies of plant-pollinator interactions are not backed up by georeferenced specimens in GBIF, as either voucher specimens are not collected or, if they are, they remain in the possession of the researchers and may not be added to GBIF for decades, if ever.

Currently, the DwC standard does not offer controlled vocabularies for all the terms defined in its vocabulary, except for establishmentMeans, degreeOfEstablishment and pathway (Groom et al., 2019). Nevertheless, the DwC recommends adopting the OBO Relation Ontology (Mungall, 2023) as a controlled vocabulary for the relationshipOfResourceID term. This term was introduced in the previous version of DwC, following a request primarily initiated by GloBI, Arctos, iNaturalist and TrIAS[31] which had been using RO for a considerable period before its recommendation in DwC.

Despite its maturity, the usage of OBO Relation Ontology in biotic interactions datasets has been limited. The lack of common guidelines for documenting biotic interactions data could be contributing to the limited awareness of its existence. Despite the DwC recommendation on adopting the OBO Relation Ontology, it is the only part of the standard that explicitly mentions biotic interactions.

Another important resource for the standardisation of agriculture data is the FAO AGROVOC Multilingual Thesaurus[32]. AGROVOC consists of more than 41,000 concepts and more than 988,000 terms in up to 42 languages. It provides a well-organised collection of agricultural concepts, terms, definitions, and relationships. This structured resource enables the unambiguous identification of agricultural resources, facilitating standardisation and interoperability of agriculture datasets. In the plant-pollinator interactions context, the AGROVOC plays an important role in the characterisation of crop plants’ traits and interaction outcomes. In the PPI vocabulary, AGROVOC terms such as “ornamental plants”[33] and “fuel crops”[34] are used as controlled vocabulary for the ppi:humanUse[35]

[31] https://github.com/tdwg/dwc/issues/283
[32] http://aims.fao.org/aos/agrovoc
[33] http://aims.fao.org/aos/agrovoc/c_5417
[34] http://aims.fao.org/aos/agrovoc/c_15583
[35] http://rs.rebipp.org.br/ppi/terms/humanUse

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

term. Other examples include AGROVOC terms such as “seed set”, “flowering time”, “flowering stage”, and “pollen germinability”.

3. Discovery phase results

3.1. Manual data search

In order to make our search for data as wide as possible, the choice was made to also include data sources that don’t show up in typical automated procedures. The main difficulty in this approach is the fact that the discovery relies on (expert) knowledge of the existence of this data. Through presentations at different related events and community calls, we collected as many datasets as possible. By using this approach, the aim was to provide an easy tool to share information with us about potential interesting sources. The results of this search can be found in the data section of the Zenodo repository (Trekels, 2023). The repository contains a detailed overview of all datasets gathered during the discovery phase.

Within this exercise, participants were specifically requested to look for data that are particularly relevant for the region of the contributor. KALRO performed an extensive search on plant-pollinator data available on the African continent. Using the list of sources that were collected, KALRO went beyond the collection of data and extracted plant-pollinator data in a standardised format according to the proposed format of GloBI (Poelen 2018), which will be one of the pilots of the next phase of project activity.

Using a hybrid approach between manual and automated searches, the HiveTracks team searched for individual datasets relating to plant-pollinator interactions and for data repositories that may contain multiple datasets pertinent to the research. The focus here was finding datasets directly covering plant-pollinator interactions, and preference was given to datasets stemming from peer-reviewed research. However, datasets relating to blooming information for plants or pollination event information for pollinators were also included in this research since they could be used to enrich plant-pollinator interaction datasets.

In addition, independent datasets and datasets from non-peer-reviewed research were included in our research to better understand the full landscape of plant-pollinator interaction data. By the end of the exploration, we found fourteen datasets directly pertaining to plant-pollinator interactions and fifteen additional datasets focusing on either the plants or the pollinators but not their interactions. Specifically, DataONE proved to be the best repository for finding relevant datasets by providing fifteen of the twenty-nine datasets identified during this research, including nine of the fourteen directly addressing plant-pollinator interactions.

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3.2. Automated exploration of datasets

Biotic interactions data have been made available from various formats and sources for many decades. The scientific literature has traditionally served as the primary source for such data, commonly found as supplementary materials or summarised in tables within scientific publications. However, this scenario has been changing in the last few decades with the emergence of open-science and open-data such initiatives, as pointed out by CODATA (2019), along with the introduction of the FAIR principles[36]. These efforts have significantly contributed to increasing the availability of data across different domains, including biotic interactions. In addition, scientific journals have changed their data policies, with a growing emphasis on promoting or even obligating authors to share the primary data associated with their publications (Nature, 2016).

These efforts have contributed to increasing the number of available datasets on plant-pollinator interactions. In order to illustrate that, datasets from four data repositories were retrieved using the search query “plant AND pollinator AND agriculture”. A total of 8,768 unique datasets containing the specified words were found across all repositories. It is important to note that Figshare includes datasets associated with scientific publications, such as supplementary data, article figures, and tables. Given that many scientific journals have utilised Figshare to provide persistent access to their articles' data, it is expected to find a significant number of datasets from this source. The datasets from the Figshare repository account for 97.7% of all datasets discovered, with 49.8% of these datasets being deposited by journals.

Figure 3 shows the accumulated number of datasets of “plant pollinator agriculture” over the years (2006 to 2023) in Dryad, DataONE, Figshare, Zenodo and Digital Science’s Dimensions (Dimensions 2022). The number of datasets by source is presented in Table 1.

[36] https://www.nature.com/articles/sdata201618

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

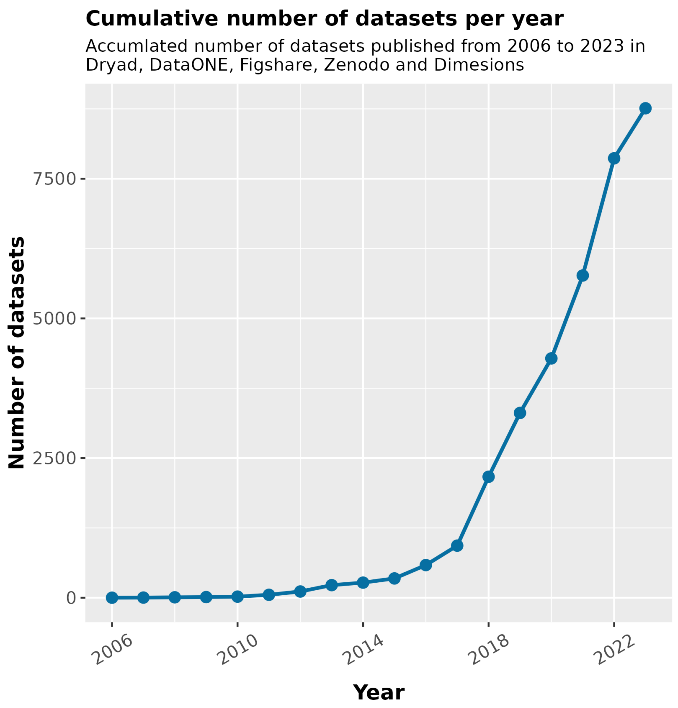

Figure 3. Cumulative number of datasets having the words "plant pollinator agriculture" in their metadata found in Zenodo, Dryad, Figshare,DataONE and Dimensions repositories from 2006 and 2023.

Repository	# Unique datasets
Figshare	8,563
DataONE	89
Dryad	45
Zenodo	47
Dimensions	24
TOTAL	**8,768**

Table 1. Number of unique datasets retrieved from each repository when using the search query "plant AND pollinator AND agriculture".

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3.3. Current practices in documenting and sharing data

Despite the growing availability of data, relatively little attention has been dedicated to the implementation of the FAIR principles concerning plant-pollinator interactions data. When examining datasets retrieved from widely used scientific repositories (Zenodo, Dryad, Figshare), the DataONE network and the Dimensions platform, the analysis reveals a notable lack of standardisation in terms of file formats (Figure 4). Due to the utilisation of Figshare by many scientific journals for publishing images, tables and supplementary materials associated with publications, it is not surprising that image file formats and Microsoft Excel are the most prevalent formats encountered. However, after removing datasets from Figshare, the contribution of textual file formats (e.g. txt, csv, tsv) as well as other open-formats (OpenXML, PDF) becomes more apparent (Figure 5).

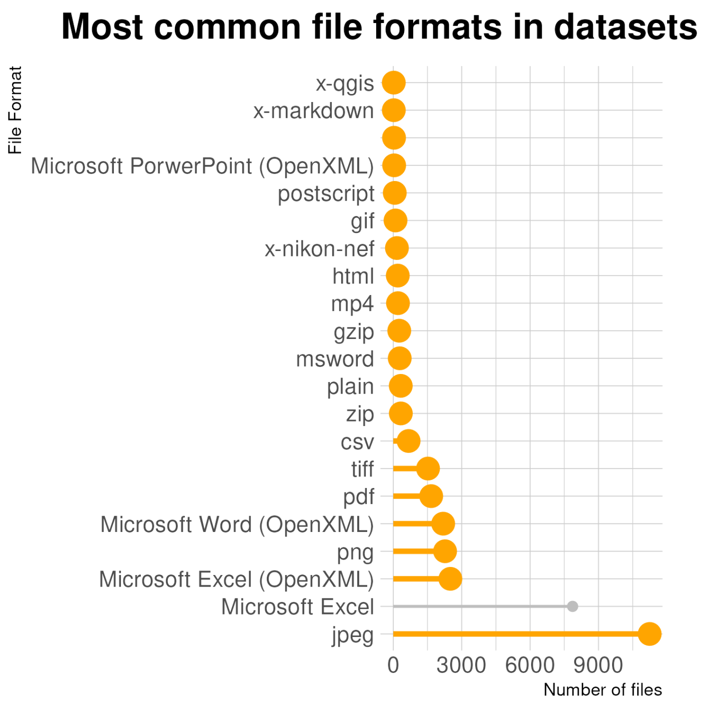

Figure 4. Top 20 most common file formats in datasets containing the words "plant pollinator agriculture" in four data sources (Zenodo, Dryad, Figshare, DataONE and Dimensions).

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

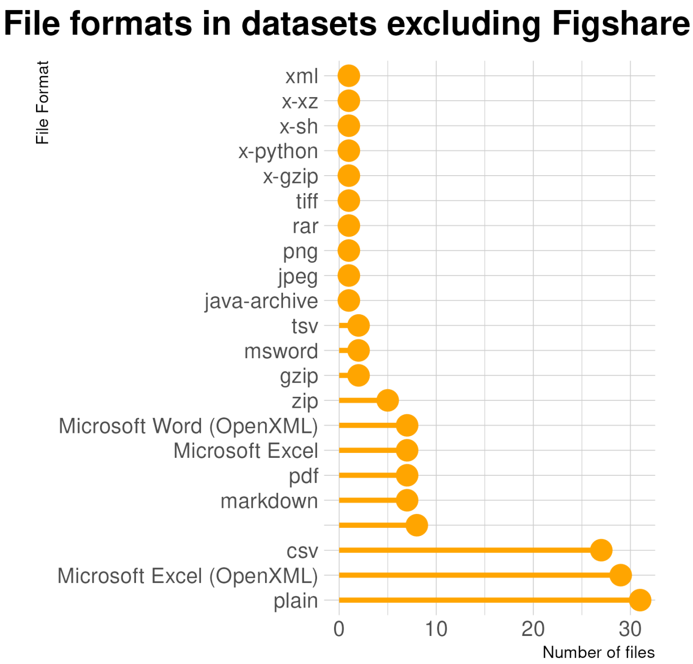

Figure 5. Top 20 most common file formats in datasets containing the words "plant pollinator agriculture" after removing datasets from Figshare repository. Sources: Zenodo, Dryad, DataONE and Dimensions.

In addition, most of the datasets obtained from Figshare seem to be "reusing" the identifier of the corresponding scientific publication to identify the datasets themselves (i.e., using the same DOI for publication and dataset). For instance, the publication identified by the DOI 10.1371/journal.pone.0228305 (Hannah et al., 2020) has the same DOI of the dataset of its supplementary material[37]. According to the FAIR principles, both resources should not share the same identifier, even though one could argue that the publication acts as metadata for the dataset. However, excluding datasets associated with scientific publications, approximately 74% of the datasets utilise DOIs as their identifiers (Figure 6). The remaining 26% use URIs (Uniform Resource Identifiers) or more specifically URNs (Uniform Resource Names) as their identifiers.

37 https://figshare.com/articles/dataset/The_environmental_consequences_of_climate-driven_agricultural_frontiers/11843448

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

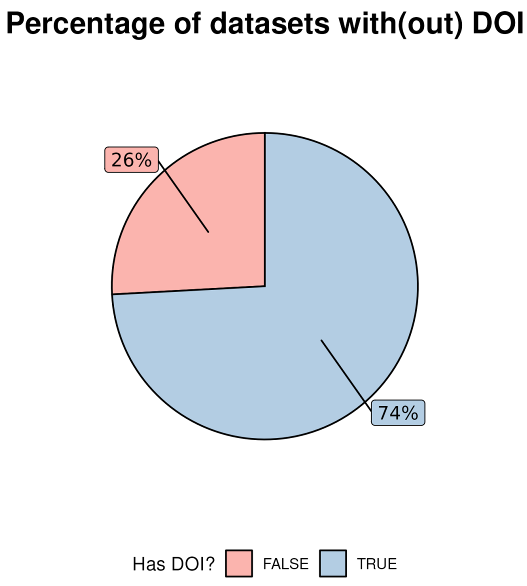

Figure 6. Analysis of the number of datasets being identified with a DOI. Sources: Zenodo, Dryad, DataONE and Dimensions.

Regarding licensing, the CC-BY 4.0 is the most frequently used licence (Figure 7). However, removing Figshare datasets, approximately 55% of the remaining datasets do not have any licence attached to them and 44.4% use CC0-1.0 (public domain) (see Figure 5).

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

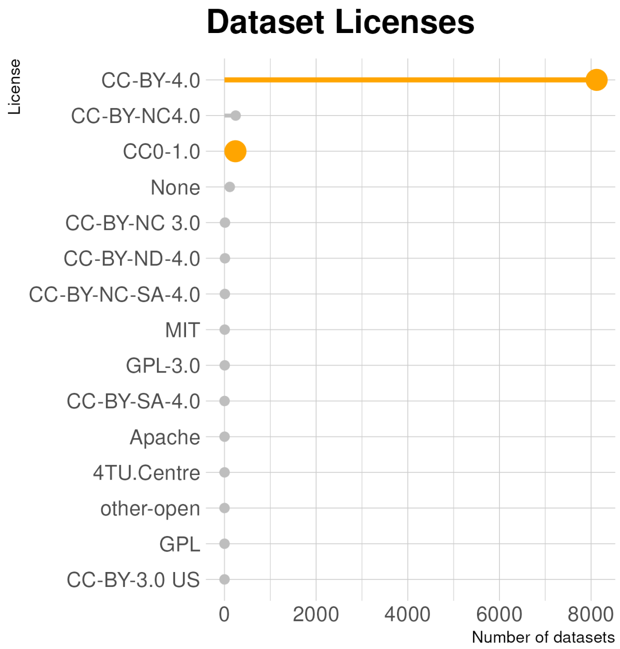

Figure 7. Licences found in all the datasets retrieved from Zenodo, Dryad, Figshare, DataONE and Dimensions.

We also accessed the total number of citations, downloads and views including Figshare datasets ("reusing DOI"): 3025 citations, 620852 downloads and 2328511 views. If we exclude Figshare datasets, the numbers are as follows: 1372 citations, 466882 downloads and 1364105 views. Interestingly, we observed a significant proportion of datasets with zero citations, accounting for 80% of datasets when considering Figshare and 70% when excluding Figshare datasets. This strongly indicates that data publication as we found for plant-pollinator interactions is probably not the best approach to allow for effective reuse (Parsons and Fox, 2013) and that there is a need to advance on data standards and infrastructures in order to mobilise data on this domain to allow for dealing with current grand challenges.

To retrieve relevant datasets about plant-pollinator interactions, it is essential to identify your dataset with appropriate keywords. Figure 8 shows the analysis of the keywords retrieved from the metadata. Although the keywords are definitely relevant for the dataset, there is a clear lack of indicating terms such as "species interactions" or "biotic interactions". To increase the findability of these datasets, it is advisable to add these keywords to the list.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

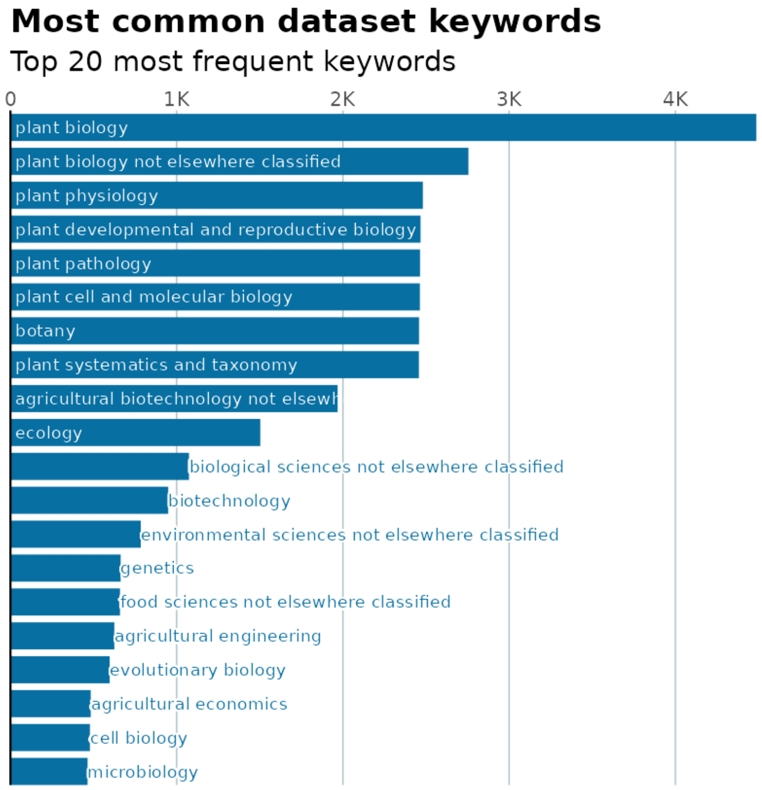

Figure 8. Top 20 most frequent keywords found in datasets from Zenodo, Dryad, Figshare and DataOne.

In this investigation, we also explored the plant-pollinator interaction datasets available in GloBI. A total of 256,420 interaction records across 52 datasets are indexed, representing approximately 1.6% of the overall records in GloBI. Most of these datasets (86%) are in CSV/TSV format, lacking proper metadata, and using GloBI's customised vocabulary to annotate columns within the files. Only three datasets use the DwC standard, while another one originates from iNaturalist. The remaining datasets, although in text file formats, have their own specific data structure. For instance, Web of Life[38] (42k indexed pollination records) exposes their data in a specialised json format. Similarly, the National Database of Plant Pollinators[39] maintained by the Center for Plant Conservation at San Diego Zoo Global uses their own specialised format.

GloBI implements an automated and continuous integration process to include up-to-date versions of datasets. As part of this process, GloBI dataset review reports were generated. These review reports test GloBI's ability to extract species interaction claims from a specific dataset. The review reports include basic statistics such as the number of interactions found, but also include taxonomic alignments of names encountered in the species interaction data across various taxonomies. In

[38] https://www.web-of-life.es
[39] https://saveplants.org/pollinator-search/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

addition, the review report includes indexed species interaction claims in various formats (tsv/csv/nanopubs) as well as reporting any suspicious records in the form of review notes.

In other words, the FAIR-ness of an interaction dataset can be measured through these GloBI dataset review reports. For instance, if all names in the dataset align with a taxonomic resource, a quantitative proxy for the "**I**" and **"R"** parts of FAIR is partly assessed. And a review implies that a dataset can be **F**ound and **A**ccessed. This suggests that the GloBI review process outcomes can be used as an indicator of the FAIRness of an interaction dataset. While all datasets indexed via GloBI are Open Access, the same process can be repeated for closed datasets in a compute environment with restricted access. Provided that the citations of the datasets under review are signed, the provenance of the data can be securely traced independent of data access methods or possible data access restrictions (Elliott et al., 2023).

In other words, when using signed citations and provenance tracking, the origin of data can be described regardless of their current availability. Through ten years of GloBI development, it was found that the effort to convert (or translate) these datasets into a more “standard” format takes far less time than it takes to collect, compile and publish the original source data. With a translation service that GloBI provides, a software engineer can add support for a custom data format in a matter of minutes (e.g., a csv file with species interactions by row) to hours (e.g., writing a custom translator for projects like Web-of-life[40]).

Workshops and the proliferation of data publication best practices (Kita et al., 2022; Keller et al., 2022) may help the pollinator data community to improve its ability to collate their digital datasets in meaningful ways. GloBI demonstrates that increased collaborations and concerted efforts lead to an increase in reusable datasets. We expect that introducing a data review process based on FAIR principles will help to build a collective understanding of how to mobilise existing datasets while facilitating the integration of newly published datasets into a (virtual) global collection of a pollination knowledge base.

Note that our estimate is a likely underestimate of the total number of plant-pollination datasets available: anecdotal evidence suggests that natural history collections contain species interaction claims that may indicate flower visitation or pollination events. Professional collectors and natural history collection curators may be more hesitant to make claims like “visits flowers of”, or “pollinates” in their collection records as direct evidence of these two events is hard to prove. Collectors and curators may instead revert to using weaker claims like: “found on” or a more brief “ex.” in their catalogues. With the Parasite Tracker TCN workshop[41], similar behaviour related to "host" or "parasite" was revealed, where many preferred "found on" or "ex." when describing the relation between a parasite and their (likely) host organism (Sullivan et al., 2020).

[40] https://www.web-of-life.es/
[41] https://globalbioticinteractions.org/parasitetracker

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

4. Discussion

The purpose of the discovery phase was to determine the status of plant-pollinator interaction datasets available and accessible by the participating partners. In the previous section, we used several methods to get our overview as complete as possible. In this section we present each of the FAIR principles and how they relate to the analysis performed on the data and point to future developments.

4.1. Findability

One of the key aspects of making data FAIR is to assign a globally unique identifier to the data and metadata ("F1. (Meta)data are assigned a globally unique and persistent identifier"[42]). Moreover, the metadata attached to this identifier needs to describe the dataset as detailed as possible ("F2. Data are described with rich metadata"[43]).

Using the automated data search, it shows that about 75% of all datasets have a DOI attached to them. However, many of them are referring to the published article and not the dataset itself. In order to have a clear picture of the data, it would be essential to assign a GUID/PID to each of the individual datasets.

The metadata connected to the datasets is rather basic, and contains typically no information on the interactions. Authors, creation date and file format are easy to retrieve from the metadata, but it is much harder to find out other essential information on the observed interaction. Even the keywords attached to the datasets are not always straightforward to link to a species interaction (see Figure 6). In order to increase the findability of the relevant datasets, it would be advisable to add at least the taxonomic, geographic and temporal scope to the metadata. These could be provided by using the Bioschemas.org[44] vocabulary. Although currently it is not possible to indicate that the dataset is containing 'interactions', this could be proposed to be added to the standard.

4.2. Accessibility

The 'accessibility' criterion was less investigated in this report. However, all of the discussed datasets that were found in the discovery phase were accessible through common web protocols. A side remark that needs to be made here is that the strategy taken in this report is slightly skewed towards datasets that do meet the accessibility criterion.

[42] https://www.nature.com/articles/sdata201618
[43] Ibid.
[44] https://bioschemas.org

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Since a large fraction of the datasets have a DOI attached to them, there is a good chance that the metadata will persist after the removal of the data. However, this will become useful only when the metadata allows for a more rich description of the data.

4.3. Interoperability

Besides the metadata flaws described above, we found most plant-pollinator interactions data available as supplementary materials attached to articles, and many of those are protected by paywalls. Interpretation of such data for its extraction to standardised formats requires expert knowledge, as no semantic description knowledge representation resources are provided. Moreover, data file formats hamper interoperability, as the preferred formats are Microsoft Excel files or plain text, and a significant amount of the files we found were stored as PDFs or Microsoft Word. We consider there is an opportunity for improvement of current practices and that publishers and journal editors have an important role in promoting cultural change.

4.4. Reusability

Since the goal is to reuse data in different contexts, it is essential that the data is well described and other people are able to use it for their own purposes. In our analysis, we mainly focussed on the reuse of the data (FAIR Principles criterion R1.1, use of data licences) and the description of the data (criterion R1.3, the use of domain-specific standards). It was notable that many of the datasets we encountered did not have any licence attached to them, indicating that the community is not aware of the importance of making clear what can and cannot be done with the data.

Although we discussed the domain-specific standards in section 2 of this document, we see that within the community there are still significant boundaries in the usage of these standards. One of the potential reasons for this could be the lack of knowledge on biodiversity informatics within research teams. The cost-benefit ratio of having the data in a standardised format is in many cases too high for researchers to go through this effort. We see two major improvements that could be made within the community.

First, it would be very beneficial for the standardisation of the data to provide easy tools for researchers to publish interaction data. Data repositories and aggregators could provide these tools to data publishers. Significant efforts are being made by GBIF (the Integrated Publishing Toolkit - IPT [45]) and GloBi (Poelen, 2018) and should be extended in the future.

Secondly, publishers should guide authors toward the domain-specific repositories and avoid the inclusion of data in proprietary formats. Including the data in infrastructures such as GBIF and GloBi will automatically result in higher reusability of the interactions data.

[45] https://www.gbif.org/ipt

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

4.5. Future developments

One of the future developments that we will explore is the semantic annotation of data. Semantic annotation is a process that involves adding meaningful metadata or annotations to data in order to enhance its understanding and interpretation by both humans and machines. In the context of the semantic web, data semantic annotation plays a crucial role in enabling interoperability and facilitating the automated processing of data.

The process of data semantic annotation begins with identifying the relevant concepts and entities within the data. This involves understanding the domain and the specific vocabulary that is appropriate for describing the data. The next step is to assign semantic annotations to the data elements, which involves associating them with standardised and well-defined terms from ontologies or controlled vocabularies.

Semantic annotations are typically expressed using standardised languages and frameworks, such as the Resource Description Framework (RDF)[46] and the Web Ontology Language (OWL)[47], which are fundamental technologies of the semantic web. These technologies provide the means to represent and query the annotated data, enabling advanced knowledge discovery, data integration, and semantic interoperability on the web.

There are several tools available for performing semantic data annotation, including OpenRefine[48] with the RDF transform extension and RDFlib for Python[49]. In this particular Case Study, both tools are being tested, and recommendations on their usage will be provided in the next stage of the project.

A sample triple store is currently available[50]. Even though this is not the final version, it gives an idea of how the data will be when we finish the whole process.

Additionally to the semantic enhancement, it is clear that taxonomic knowledge is essential in recording the species involved in the interactions. The exercise of KALRO in extracting the species interactions showed that this is a tedious process and is highly reliant on expert knowledge. Although we will explore this further in the project, it is definitely worth mentioning that taxonomic knowledge is in decline (European Commission, 2022). More effort is needed to enhance the capacity on taxonomy, e.g. through projects such as the EC funded TeTTRIs project[51] and integrated infrastructure such as ChecklistBank.

46 https://www.w3.org/RDF/
47 https://www.w3.org/OWL/
48 https://openrefine.org/
49 https://github.com/RDFLib/rdflib
50 https://data.pldn.nl/Miranda/PlantPollinatorInteractions
51 https://tettris.eu/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

5. Work plan for pilots

During the next phase of the project, the WorldFAIR Agricultural Biodiversity Case Study will conduct pilots that will serve as targets for standards adoption within the community and allow for the development of guidelines and recommendations, FAIR assessments and estimation of costs of adoption. The pilots will be divided into two groups: plant-pollinators interactions data and pollinators data, as detailed below.

5.1. Plant-pollinator interactions data

Different initiatives on plant-pollinator interactions data will be investigated via pilot projects for standards adoption. We will explore alternative pathways and measure their costs in order to provide an overview of the possibilities. This will enable the generation of guidelines in the next phases of the project.

The first approach will be preparing the datasets for entering the GloBI portal and establishing a data review process to help measure the mobility of knowledge about pollination processes in the datasets under review. We will also explore the preparation of the datasets according to the models detailed in section 2.4 of this report: REBIPP data model and the GBIF Unified Data Model for Biotic Interactions. In both cases, we will use the plant-pollinator interactions vocabulary (PPI), designed to be used with the DwC data standard. We will compare the results and effort for each approach.

The main candidates to go through this pilot phase are:

- The Plant-Pollinator Data Collection by KALRO on the African continent. KALRO performed an extensive search on plant-pollinator data available on the African continent, with 121 datasets, and extracted the data to the GloBI simplified data sheet, which resulted in more than 1,000 records.
- The Brazilian Plant-Pollinator Interactions Network (REBIPP). REBIPP aggregates data from several initiatives in different Brazilian biomes. We will choose representative datasets from the network for the pilot phase.
- Observations of plant-pollinator interactions in the Pampean region of Argentina. This pilot will focus on processing a large dataset of field observations of plant-pollinator interactions recorded at several locations in the Pampean region of Argentina. The dataset comprises ca. 130 plant-flower visitor networks from different locations in the Pampean region, an intensively cropped area in central Argentina. Each network was constructed by sampling interactions between flowering plants in field margins and the insects that visited their flowers to forage for nectar or pollen. When the adjacent crop had flowers that attracted insects (e.g. 78 soybean plots and 15 potato plots) samplings also included flower visitors to the crop.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

This list of pilots is also not exhaustive and could be modified during the pilot phase.

5.2. Pollinators data (Occurrences)

The pilot led by the startup company HiveTracks will focus on leveraging its standardised data entry process for hive health assessments, beekeeping practices and documentation of available floral resources around a beehive tested during the discovery phase. Based on the data collected through the HiveTracks application (Figure 9) by 2,000 beekeepers since February 2023 (Figure 10), a group of beekeepers will be selected that have shown the highest frequency, quality and depth of data collected across a range of categories:

- Number of apiaries and hives
- Apiary environment and terrain
  - Locations
  - Land-use and land-cover in apiary proximity
- Number of years in beekeeping experience
- Number of hive inspections recorded
  - Hive population size
  - Availability of brood stages
  - Presence of diseases
  - Availability of food stores
- Number of management practices recorded
  - Feeding
  - Treating
- Number of floral resources documented
  - Type and occurrence of floral resource
- Number of rules-based recommendations rated
- Amount of honey harvested

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

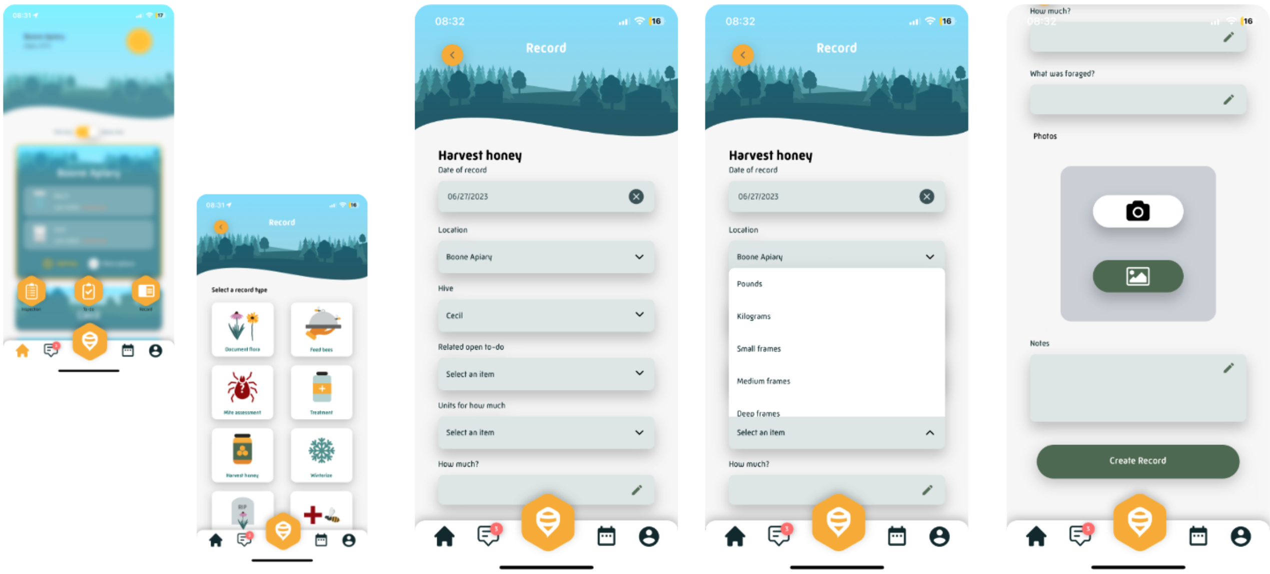

Figure 9. Example of standardised data entry process for a honey harvest record in the HiveTracks App

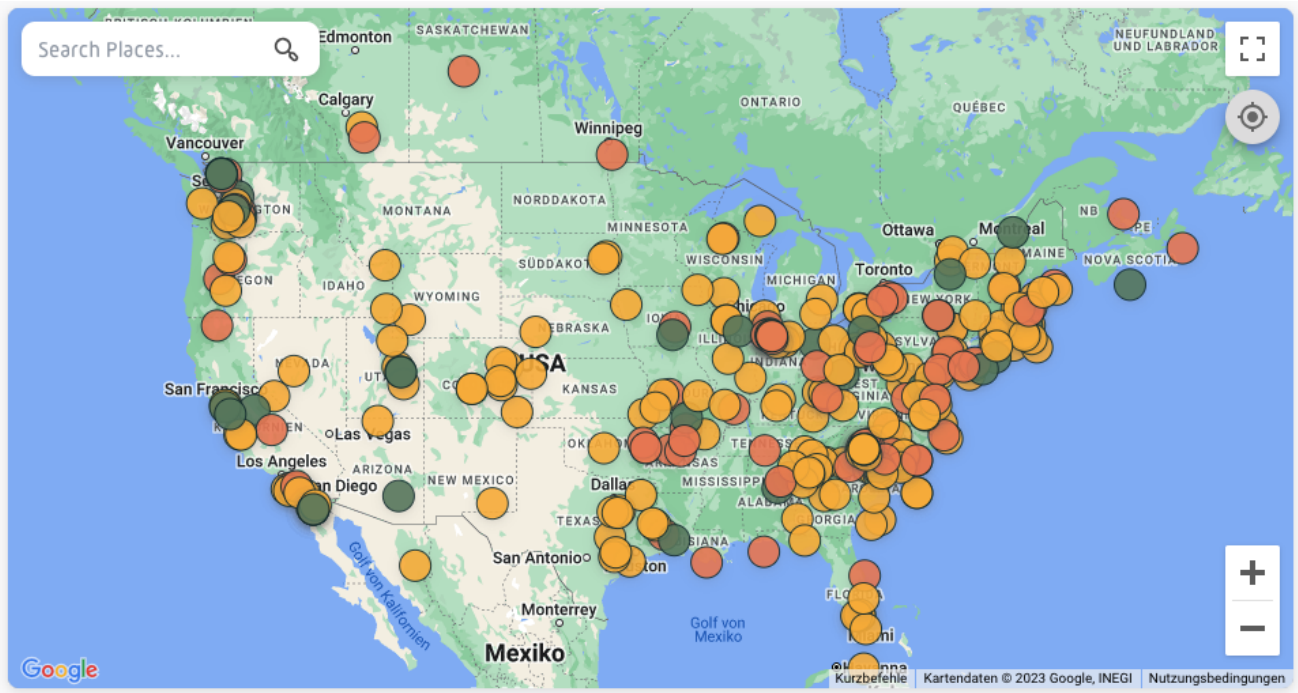

Figure 10. Visualisation of apiary locations set up by beekeepers during the discovery phase

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

The goal of the pilot is to validate an inclusive end-to-end process for community-based sourcing of FAIR managed pollinator data that protect the data privacy of beekeepers, whose livelihoods depend upon the location and positioning of their hives. In addition, recommendations for private-sector organisations to follow FAIR principles will be developed to increase the adoption of FAIR data management. The collected data will then be visualised on a honey harvest-specific profile that follows the design principles of this mock-up.

Considering the importance of promoting standards adoption for occurrences data related to pollination, we will explore standards like DwC, Agrovoc, Hymenoptera and Environment Ontologies for HiveTracks data and measure the costs for adoption. There is also the opportunity for exploring a different type of interaction data in the field “presence of diseases”.

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

6. Recommendations

Our recommendations, resulting from the work described above, are each labelled for clarity with the type of recommendation and the target stakeholders to whom we are recommending action.

Recommendation 1 - Taxonomic data.

- *Type:*
    - ***Technical (data)***
    - ***Technical (metadata)***

- *the stakeholder(s) at which the recommendation is aimed.*

    Journal editors and publishers, biodiversity data infrastructures

Verbatim names should be included with a description of the taxonomic name alignment process and associated taxonomic authorities. Taxonomic interpretations should be documented when possible, as well as the specific version of the taxonomic resources used. Ideally, names are aligned across various taxonomic resources to get a sense of the variability of the name interpretations. Name alignment processes can be used to review names in biodiversity datasets to assess how well-observed names link to commonly used taxonomic name resources. Consequently, the interoperability (or reusability) of a dataset can be quantitatively and automatically assessed.

Recommendation 2 - Infrastructure for biodiversity data publishing.

- *Type:*
    - ***Policy***
    - ***Organisational***

- *the stakeholder(s) at which the recommendation is aimed.*

    Journal editors and publishers, biodiversity data infrastructures, data repositories managers

Shifting from attaching flat files as supplementary materials in papers are urgently needed. Publishers and editors need to guide authors toward domain-specific repositories and avoid the inclusion of data in proprietary formats. Including the data in infrastructures such as GBIF and GloBI

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

will automatically result in a higher adherence of the FAIR principles and enable the reusability of the interactions data.

We also identified the need for a wider conversation with the different data repositories (such as FigShare, Dryad…) from our analysis. It could be beneficial for the objectives of WorldFAIR to engage with these stakeholders. This may improve the FAIRness of data in multiple domains.

Recommendation 3 - Richer metadata schemes and development of controlled vocabularies allied to connection with CDIF.

- *Type:*
    - ***Policy***
    - ***Organisational***

- *the stakeholder(s) at which the recommendation is aimed.*

  Journal editors and publishers, biodiversity data infrastructures, standards organisations and communities

To increase the discoverability and interoperability of the plant-pollinator data we propose to adopt a richer metadata scheme that is attached to the datasets. This metadata scheme should include some essential information on the data:

- taxonomic coverage
- geographic coverage
- an indication of biotic interactions involved in the dataset

One of the most promising metadata schemes according to us, is the adoption and adaptation of Bioschemas.org. This scheme already covers the taxonomic dimension, but could be extended to describe the interactions.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

7. Conclusions

The discovery phase of the Agricultural Biodiversity Case Study showed that there is a significant amount of valuable data still locked up in proprietary formats and images. The reason for this is mainly due to the nature of the recording of these interactions, possibly allied to a lack of knowledge of available resources to foster interoperability. Although major efforts are being made worldwide to publish datasets as valuable scientific outputs, we identified that there is a need to advance data standardisation and infrastructures in order to provide FAIR data and allow for effective reuse. This could be partially solved by creating incentives via the publication process, but also by attributing more value to the data itself.

To facilitate and promote the reuse of data, proper licensing of the datasets is crucial. Potential users will not automatically know whether data are eligible for copyright, and they should not use data without knowing whether they have the right to use them. Even if authors do not claim any rights on their dataset they need to say so in order that users can reuse it (DiSSCo, 2023). Using domain-specific repositories and a revision of the publication process could be instrumental in this.

This report also indicates that the metadata scheme is not sufficient to cover plant-pollinator data and make these data findable outside the pollination ecology community. Especially the taxonomic dimension and the fact that the data dealing with interactions is difficult to cover and not standardised. Improving this will require an active engagement of the community with the non-domain-specific (meta)data standards organisations.

The value of adopting FAIR standards for biodiversity data such as plant-pollinator interactions cannot be over-estimated. In June 2023 a report[52] by Deloitte Access Economics, and commissioned by the Global Biodiversity Information Facility (GBIF) concluded that each €1 that is invested in GBIF resulted in direct benefits of €3 for GBIF users, a three-fold return on investment. Not only that, but there are wider benefits to society of up to €12 per €1 invested. A survey of GBIF users revealed that they “value open access to biodiversity data at €13 million per year” and that the facility “saves users an estimated 845,000 hours of research time annually, valued at €35 million, by providing efficient and open access to required data”. One of the areas highlighted for further work is an analysis of how data standardisation can add significant value to GBIF, which they conclude "is difficult to quantify, but could be in the order of billions of euros"[53].

This makes estimating the costs of adoption even more relevant, which will be explored in the pilot phase of this Case Study. We will also explore the connection with CDIF through the investigation of the metadata standards that are relevant to agricultural biodiversity data. In particular, the role of Bioschemas.org will be investigated.

[52] https://www.gbif.org/news/5WZThcL928vmPnSvrGhZfE/report-reveals-return-on-investments-in-gbif

[53] https://www.deloitte.com/au/en/services/economics/perspectives/total-economic-value-open-access-database-living-world.html

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

8. Bibliography

Bánki, O., Roskov, Y., Döring, M., Ower, G., Hernández Robles, D. R., Plata Corredor, C. A., Stjernegaard Jeppesen, T., Örn, A., Vandepitte, L., Hobern, D., Schalk, P., DeWalt, R. E., Keping, M., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Adlard, R., Adriaenssens, E. M., et al. (2023). Catalogue of Life Checklist (Annual Checklist 2023). Catalogue of Life. https://doi.org/10.48580/dfsr

Brittain, C., Kremen, C., Garber, A. and Klein, A.-M. (2014). Pollination and plant resources change the nutritional quality of almonds for human health. PLoS One 9: e90082. doi: 10.1371/journal.pone.0090082.

Butterfield, B.J. and Callaway, R.M. (2013). A functional comparative approach to facilitation and its context dependence. Funct Ecol, 27: 907-917. https://doi.org/10.1111/1365-2435.12019

Campbell, AJ, Carvalheiro, LG, Maués, MM, et al. (2018). Anthropogenic disturbance of tropical forests threatens pollination services to açaí palm in the Amazon river delta. J Appl Ecol. 2018; 55: 1725– 1736. https://doi.org/10.1111/1365-2664.13086

Campbell, A. J., Silva, F. D. d. S. e., Maués, M. M., Leão, K. L., Carvalheiro, L. G., Moreira, E. F., Mertens, F., Konrad, M. L. d. F., de Queiroz, J. A. L., & Menezes, C. (2023). Forest conservation maximises açaí palm pollination services and yield in the Brazilian Amazon. Journal of Applied Ecology, 00, 1– 13. https://doi.org/10.1111/1365-2664.14460

Chamberlain, S.A., Bronstein, J.L., Rudgers, J.A. (2014). How context dependent are species interactions? Ecology Letters (2014) 17: 881– 890. https://doi.org/10.1111/ele.12279

Chaplin-Kramer, R., Dombeck, E., Gerber, J. et al. (2014). Global malnutrition overlaps with pollinator-dependent micronutrient production. Proceedings of the Royal Society B 281: 20141799. https://doi.org/10.1098/rspb.2014.1799.

CODATA, Committee on Data of the International Science Council, CODATA International Data Policy Committee, CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice, Hodson, Simon, Mons, Barend, Uhlir, Paul, & Zhang, Lili. (2019). The Beijing Declaration on Research Data. Zenodo. https://doi.org/10.5281/zenodo.3552330

Dimensions. (2022). Digital Science. Dimensions [Software] available from https://app.dimensions.ai [Accessed on 26 October 2022, under licence agreement].

DiSSCo. (2023). https://dissco.github.io/ [Accessed 22 July 2023].

Drucker DP, Salim JA, Trekels M, Groom Q, Parr C, Soares FM, Agostini K, Saraiva AM, Molloy L,

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Hodson S, Gregory A. (2022). Plant-pollinator Interaction Data: A case study of the WorldFAIR project. Biodiversity Information Science and Standards 6: e94310. https://doi.org/10.3897/biss.6.94310

Elliott, M.J., Poelen, J.H. & Fortes, J.A.B. (2023). Signing data citations enables data verification and citation persistence. Sci Data 10, 419. https://doi.org/10.1038/s41597-023-02230-y hash://sha256/f849c870565f608899f183ca261365dce9c9f1c5441b1c779e0db49df9c2a19d.

Eilers, E.J., Kremen, C., Greenleaf, S.S., Garber, A.K. and Klein, A.-M. (2011). Contribution of pollinator-mediated crops to nutrients in the human food supply. PLoS One 6: e21363. https://doi.org/10.1371/journal.pone.0021363.

Ellis, A.M., Myers, S.S. and Ricketts, T.H. (2015). Do pollinators contribute to nutritional health? PLoS One 10: e114805. https://doi.org/10.1371/journal.pone.0114805.

Enetwild consortium, Jaroszynska, F., Body, G., Pamerlon, S., & Archambeau, A. S. (2022). Applying the Darwin Core data standard to wildlife disease–advancements toward a new data model. EFSA Supporting Publications, 19(11), 7667E. https://doi.org/10.2903/sp.efsa.2022.EN-7667

European Commission, Directorate-General for Environment, Hochkirch, A., Casino, A., Penev, L., et al. (2022). European Red List of insect taxonomists, Publications Office of the European Union. https://data.europa.eu/doi/10.2779/364246

Frederickson, M.E. (2017). Mutualisms Are Not on the Verge of Breakdown. Trends in Ecology and Evolution 32, ISSUE 10, P727-734. https://doi.org/10.1016/j.tree.2017.07.001

GBIF: The Global Biodiversity Information Facility (2023). What is GBIF? Available from https://www.gbif.org/what-is-gbif [Accessed 13 January 2023].

GBIF Secretariat. (2022). GBIF Backbone Taxonomy. Checklist dataset. https://doi.org/10.15468/39omei [Accessed via GBIF.org on 20 July 2023].

GBIF.org (2023). Use Case: Biotic Interactions - Sottunga Island Melitaea cinxia Population Study. Available from https://docs.google.com/document/d/1jzb54GbAkB_TOFIjWof5BW6gn1ujSnXXJQu6aZgFAB4/edit#heading=h.25ixp1kekb88 [Accessed 20 July 2023].

GloBI Community. (2022). Global Biotic Interactions: Interpreted Data Products (0.5) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7348355

Gómez, José María, Iriondo, José María, and Torres, Pedro. (2023). “Modeling the Continua in the Outcomes of Biotic Interactions.” Ecology 104(4): e3995. https://doi.org/10.1002/ecy.3995

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Groom Q, Desmet P, Reyserhove L, Adriaens T, Oldoni D, Vanderhoeven S, Baskauf SJ, Chapman A, McGeoch M, Walls R, Wieczorek J, Wilson JR.U, Zermoglio PFF, Simpson A. (2019). Improving Darwin Core for research and management of alien species. Biodiversity Information Science and Standards 3: e38084. https://doi.org/10.3897/biss.3.38084

Hannah L, Roehrdanz PR, K. C. KB, Fraser EDG, Donatti CI, Saenz L, et al. (2020). The environmental consequences of climate-driven agricultural frontiers. PLoS ONE 15(2): e0228305. https://doi.org/10.1371/journal.pone.0228305

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. and Umbanhowar, J. (2010). A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. Ecology Letters, 13: 394-407. https://doi.org/10.1111/j.1461-0248.2009.01430.x

Hoeksema, Jason D., and Emilio M. Bruna. (2015). “Context-dependent outcomes of mutualistic interactions”, in Judith L. Bronstein (ed.), Mutualism. Oxford. https://doi.org/10.1093/acprof:oso/9780199675654.003.0010

iNaturalist. (2023). Available from https://www.inaturalist.org [Accessed 21 July 2023].

IPBES. (2016). The Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on Pollinators, Pollination and Food Production. Potts, S.G., Imperatriz-Fonseca, V.L. and Ngo, H.T. (eds). Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.

Jordano, P. (2021). The Biodiversity of Ecological Interactions: Challenges for recording and documenting the Web of Life. Biodivers. Inf. Sci. Stand. 5, e75564. https://doi.org/10.3897/biss.5.75564

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. … Penone, C. (2023). Ten (mostly) simple rules to future-proof trait data in ecological and evolutionary sciences. Methods in Ecology and Evolution, 14, 444– 458. https://doi.org/10.1111/2041-210X.14033

Kita, C.A. et al., (2022). Ten simple rules for reporting information on species interactions. PLoS Computational Biology. https://doi.org/10.1371/journal.pcbi.1010362

Klein, A.-M., Vaissière, B.E., Cane, J.H. et al. (2007). Importance of pollinators in changing landscapes

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

for world crops. Proceedings of the Royal Society of London B. 274: 303–313.

Lidicker, WZ. (1979). A clarification of interactions in ecological systems. BioScience. 29(8):475–7.

Linnaeus, C. (1758). Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, 10th edition. Lars Salvi, Stockholm. https://doi.org/10.5962/bhl.title.542

Lopes, M. S., El-Basyoni, I., Baenziger, P. S., Singh, S., Royo, C., Ozbek, K., ... & Vikram, P. (2015). Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. Journal of experimental botany, 66(12), 3477-3486. https://doi.org/10.1093/jxb/erv122

Magagna, B, et al. (2020). Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. OSF Preprints: https://doi.org/10.31219/osf.io/2p85g

Maron, J.L., Baer, K.C., Angert, A.L. (2014). Disentangling the drivers of context-dependent plant–animal interactions. J Ecol, 102: 1485-1496. https://doi.org/10.1111/1365-2745.12305

Miller, J, Robertson, T, & Wieczorek, J. (2023). WorldFAIR Project (D9.1) Data standard for sharing ecological and environmental monitoring data documented for community review (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7849241

Miller, JT, Poelen, Jorrit, & Seltmann, Katja. (2023). Big Bee Name Alignment Workshop. Zenodo. https://doi.org/10.5281/zenodo.7829969

Mungall, C. et al. (2023). oborel/obo-relations: June Release (v2023-06-28). Zenodo. https://doi.org/10.5281/zenodo.8110608

National Museum of Natural History, Smithsonian Institution. (2023). Integrated Taxonomic Information System (ITIS). Checklist dataset https://doi.org/10.5066/f7kh0kbk accessed via GBIF.org on 2023-07-20.

Nature. (2016). “Announcement: Where Are the Data?” 537 (7619): 138–138. https://doi.org/10.1038/537138a

Ollerton, J. (2021). Pollinators & Pollination: Nature and Society. Pelagic Publishing, Exeter, UK.

Ollerton, J., Liede-Schumann, S., Endress, M. E., Meve, U., et al. (2019). The diversity and evolution of pollination systems in large plant clades: Apocynaceae as a case study. Annals of Botany 123: 311-325

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Parsons, M.A. and Fox, P.A. (2013). Is Data Publication the Right Metaphor? Data Science Journal, 12(0), p.WDS32-WDS46.DOI: https://doi.org/10.2481/dsj.WDS-042

Poelen, J. (2018). globalbioticinteractions/template-dataset: More examples and mappings. (0.0.3) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.1436853

Poelen, J. H. (2023). Nomer Corpus of Taxonomic Resources hash://sha256/0e9bc57bc082b58a2c7a509bb73362b258ec8ddfc6664898e25c639786413fda hash://md5/91dd844e787ffae8f0a2bbb8c1f29192 (0.16) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.8125362

Rodger. J. G. et al. (2021). Widespread vulnerability of flowering plant seed production to pollinator declines.Sci. Adv.7,eabd3524. https://doi.org/10.1126/sciadv.abd3524

Roy H.E., Baxter E., Saunders A., Pocock M.J.O. (2016). Focal Plant Observations as a Standardised Method for Pollinator Monitoring: Opportunities and Limitations for Mass Participation Citizen Science. PLOS ONE 11(3): e0150794. https://doi.org/10.1371/journal.pone.0150794

Smith, M.R., Singh, G.M., Mozaffarian, D. and Myers, S.S. (2015). Effects of decreases of animal pollinators on human nutrition and global health: a modelling analysis. Lancet 386: (15) 61085-6 1964–1972. https://doi.org/10.1016/S0140-6736(15)61085-6 .

Salim J.A. et al. (2022a). Data standardization of plant–pollinator interactions, GigaScience, Volume 11, giac043, https://doi.org/10.1093/gigascience/giac043

Salim J.A., Seltmann K.C., Poelen J.H., Saraiva A.M. (2022b). Indexing Biotic Interactions in GBIF data. Biodiversity Information Science and Standards 6: e93565. https://doi.org/10.3897/biss.6.93565

Schneider, F.D., Fichtmueller, D., Gossner, M.M., et al. (2019). Towards an ecological trait-data standard. Methods Ecol Evol. 10: 2006– 2019. https://doi.org/10.1111/2041-210X.13288

Soares, F. M., Hamanaka, R.Y., Maculan, B.C.M.S. (2021). Interoperabilidade semântica no contexto de dados da biodiversidade: um estudo de caso sobre a utilização de padrões de metadados [Semantic interoperability in the context of biodiversity data: a case study on the use of metadata standards]. Proceedings of Encontro Nacional de Pesquisa em Ciência da Informação, 21. https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/79/164

Sullivan, K., Seltmann, K., Poelen, J., Zaspel, J. (2020). Making Parasite-Host Associations Visible in Terrestrial Parasite Tracker (TPT). Zenodo. https://doi.org/10.5281/zenodo.3780543

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Trekels, M. (2023). AgentschapPlantentuinMeise/WorldFAIR. Zenodo. https://doi.org/10.5281/zenodo.8169964

Ueda, K. (2020). An overview of computer vision in inaturalist. Biodiversity Information Science and Standards, 4, e59133. https://doi.org/10.3897/biss.4.59133

Upham, N. S., Poelen, J., Paul, D., Groom, Q., Simmons, N., Vanhove, M., Bertolino, S., Reeder, D., Bastos-Silveira, C., Sen, A., Sterner, B., Franz, N., Guidoti, M., Penev, L., Agosti, D. (2021). Liberating host–virus knowledge from biological dark data. The Lancet Planetary Health. https://doi.org/10.1016/S2542-5196(21)00196-0.

Vohland, K. et al. (2021). The Science of Citizen Science. Springer. https://doi.org/10.1007/978-3-030-58278-4

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin core: An evolving community-developed biodiversity data standard. PLoS ONE, 7(1), e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

WoRMS Editorial Board (2023). World Register of Marine Species. Available from https://www.marinespecies.org at VLIZ. [Accessed 20 July 2023]. https://doi.org/10.14284/170

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.