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

D10.3 Agricultural biodiversity FAIR data assessment rubrics

Work Package	WP10 - Agricultural Biodiversity
Lead Author (Org)	Debora Pignatari Drucker (Embrapa)
Contributing Author(s) (Org)	José Augusto Salim (Unicamp), Jorrit Poelen (Ronin Institute, UC Santa Barbara), Filipi Miranda Soares (USP & UTwente), Rocio Ana Gonzalez-Vaquero (UBA), Mariano Devoto (UBA), Jeff Ollerton (University of Northampton), Muo Kasina (KALRO), Luísa Gigante Carvalheiro (UFG), Pedro Joaquim Bergamo (Unesp), Denise Araujo Alves (USP), Isabela Varassin (UFPR), Carla Tinoco

	(UFG), Max Rünzel (HiveTracks), Drew Robinson (HiveTracks), Juliana Cardona-Duque (University CES), Mileidy Idárraga (University CES), M. Camila Agudelo-Zapata (University CES), Esteban Marentes Herrera (SiB Colombia), Christine Taliga (USDA NRCS), Cynthia Parr (USDA ARS), Diana Cox-Foster (USDA ARS), Elizabeth Hill (USDA OCS), Márcia Motta Maués (Embrapa), Kayna Agostini (UFSCar), André Rodrigo Rech (UFVJM), Antonio Saraiva (USP)
Due Date	29.02.2024
Date	26.02.2024
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.10719265

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	12.02.2024	Debora Pignatari Drucker, José Salim, Jorrit Poelen, Filipi Soares	Draft for internal review
0.2	26.02.2024	All authors	Content ready

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.

Abbreviations and Acronyms

CDIF	Cross-Domain Interoperability Framework
DwC	Darwin Core
EMBRAPA	Brazilian Agricultural Research Corporation
EML	Ecological Metadata Language
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile
GBIF	Global Biodiversity Information Facility
GloBI	Global Biotic Interactions
IGAD	Improving Global Agricultural Data Community of Practice
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
KALRO	Kenya Agricultural and Livestock Research Organization
PPI	Plant-Pollinator Interactions Vocabulary
REBIPP	Brazilian Network of Plant-Pollinator Interactions
RDA	Research Data Alliance
SIB/Colombia	Sistema de Información sobre Biodiversidad de Colombia
TDWG	Biodiversity Information Standards
URI	Uniform Resource Identifier
USDA	United States Department of 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.

Executive summary

The WorldFAIR Case Study on Agricultural Biodiversity (WP10) addresses the challenges of advancing interoperability and mobilising plant-pollinator interactions data for reuse. Previous efforts, reported in WorldFAIR Deliverable 10.1, 'Agriculture-related pollinator data standards use cases report' (Trekels *et al.*, 2023), provided an overview of projects, good practices, tools, and examples for creating, managing and sharing data related to plant-pollinator interactions. It also outlined a work plan for conducting pilot studies. Deliverable 10.2 (Drucker et al., 2024) presented Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations. This deliverable presented results from six pilot studies that adopted standards and recommendations from the earlier report. The current report complements the efforts with Agricultural Biodiversity FAIR data assessment rubrics.

We introduce a set of FAIR assessment tools tailored to the plant-pollinator interactions domain. These tools are designed to help researchers and institutions evaluate adherence to the FAIR principles. In the discovery phase, we found that a significant amount of data on plant-pollinator interactions is available as supplementary files of research articles, in a diversity of formats such as PDFs, Excel spreadsheets, and text files. The diversity of approaches and the lack of appropriate data vocabularies lead to confusion, information loss, and the need for complex data interpretation and transformation. Our proposed framework primarily targets researchers in this domain who wish to assess the FAIRness of the data they produce and take action to improve it. However, we believe it can also benefit data reviewers, data stewards, data repository managers and librarians dealing with plant-pollinator data. Our approach focuses on being as familiar as possible with the researcher's practices, language, and jargon. Ultimately, we aim to promote data publishing and reuse in the plant-pollinator interactions domain.

We present a 'Rubric for the assessment of Plant-Pollinator Interactions Data' with examples from the data from the pilots developed in Deliverable 10.2 and in relation to the FAIR Implementation Profile (FIP) created by Work Package 10. We conduct 'dataset assessments' of available data from research projects surveyed in the discovery phase. Additionally, we describe in detail the 'Automated FAIR-enabled Data Reviews' generated by the Global Biotic Interactions (GLoBI) infrastructure, with examples from the pilots.

We believe the tools described in this report will encourage data publishing and reuse in the plant-pollinator interactions domain. Moving from diverse approaches and siloed initiatives to widely available FAIR plant-pollination interactions data for scientists and decision-makers will enable the development of integrative studies that enhance our understanding of species biology, behaviour, ecology, phenology, and evolution.

'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

Plant-pollinator interactions play a pivotal role in ecosystem functioning and sustainable agriculture. Understanding these interactions is essential for addressing key questions such as the impact of managed bees on wild ecosystems, the contribution of wild and managed pollinators to crop pollination, and the reciprocal effects of crops on pollinators. The WorldFAIR Agricultural Biodiversity Case Study (WP10) aims to ensure that plant-pollinator interactions data are FAIR (Findable, Accessible, Interoperable, and Reusable, Wilkinson *et al.,* 2016) for understanding these interactions at biologically relevant scales for crops and associated habitats. By promoting the adoption of FAIR data standards by multiple initiatives worldwide, we are working to transform the current scenario where data is scattered across various networks and country-specific initiatives, stored in isolated silos, into a scenario where plant-pollinator interactions FAIR data is widely available to scientists and decision-makers. This transformation enables the development of integrative studies that enhance our understanding of species biology, ecology, phenology, and evolution, and provides useful baseline figures for pollinator management practices and conservation efforts.

Following up on previous efforts undertaken by this Case Study, described in Deliverable 10.1, “Agriculture-related pollinator data standards use cases report” (Trekels et al. 2023), and Deliverable 10.2, "Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations" (Drucker et al. 2024), this deliverable (D10.3) presents a FAIR assessment framework for plant-pollinator data, following the strategy illustrated in Figure 1.

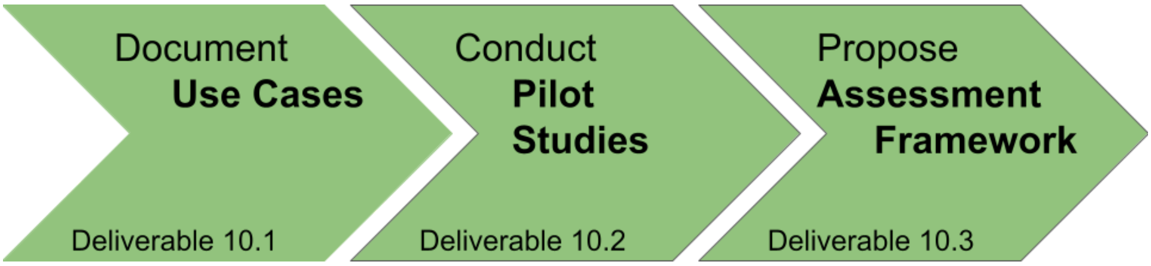


Figure 1. This document represents the third deliverable of the WorldFAIR Agricultural Biodiversity work package, focusing on assessment strategies for plant-pollinator data. The strategies outlined in this document are based on documented use cases (D10.1) and utilise examples from WorldFAIR pilot studies (D10.2).

We introduce a set of FAIR assessment tools tailored to the plant-pollinator interactions domain, designed to assist researchers and institutions in evaluating adherence to the FAIR principles. Through our work in Deliverable 10.1, we discovered that a significant amount of data on plant-pollinator interactions is provided as supplementary files of research articles, often in diverse formats, such as PDFs, Excel spreadsheets, and text files. This variety of approaches, coupled with the absence of appropriate data vocabularies, can lead to confusion, information loss, and the need

‘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 complex data interpretation and transformation. Therefore, our proposed framework primarily targets researchers in this domain who seek to assess and improve the FAIRness of the data they produce. However, we believe that our tools can also be valuable to data reviewers, data stewards, data repository managers and librarians dealing with plant-pollinator data. They may use these tools to ensure that the datasets they manage align with the FAIR principles.

It is worth noting that there are several good quality FAIR assessment tools available, many of which are domain-agnostic. Examples include the FAIR Data Maturity Model (2020) and the F-UJI - An Automated FAIR Data Assessment Tool (Devaraju & Huber, 2020). We recommend these tools for those interested in the FAIR principles in general and for gaining insights into the relative importance of each sub-principle. However, our approach in this work is tailored specifically to the domain of plant-pollinator interactions aimed at those who are familiar with the practices, language, and jargon used by researchers in this field. Ultimately, our goal is to encourage data publishing and reuse in the domain of plant-pollinator interactions.

This report is structured into three main sections: in section 2, we present a “Rubric for Assessment of Plant-Pollinator Interactions Data”, featuring examples from the pilots developed in Deliverable 10.2. Section 3 includes “Dataset Assessments'' of available data from research projects surveyed in the discovery phase (Deliverable 10.1). In section 4, titled “Automated FAIR-enabled Data Reviews”, we provide a detailed description of the data review generated by the Global Biotic Interactions infrastructure (GloBI, Poelen *et al.,* 2014), as introduced in Deliverable 10.2 (Drucker et al. 2024), with examples from the pilots. The report is complemented by five appendices: in Appendix I, we present the FAIR Implementation Profile (FIP) created by Work Package 10, with additional comments and examples. The full review reports generated by GloBI to all the pilots are available at Appendix II. The annotated data sets from the pilots are presented in the Appendix III and, lastly, Appendices IV and V present, respectively, GloBI Contribution Guidelines and GloBI Integration and Review Process in details.

‘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.

2. Rubric for assessment of plant-pollinator interactions data

Promoting data reuse is the ultimate goal of the FAIR principles. This rubric is designed to standardise the review process of datasets on plant-pollinator interactions and to qualitatively assess their reusability. In this assessment tool, we provide a list of 11 rubric items, stated as questions, to facilitate the reuse of plant-pollinator interactions data. Each specific plant-pollinator rubric item is linked to one or more WP10 FAIR Implementation Profile (FIP) items and associated FAIR principles or sub-principles. Users can find more information about the associated principles or sub-principles by clicking on the hyperlink in the fourth column, which will guide them to the detailed FIP presented in Appendix I.

The plant-pollinator rubric serves as a comprehensive tool for introducing strategies aimed at facilitating the reuse of plant-pollinator data, and also to seamlessly introduce global initiatives like FIPS, FAIR and related activities. For each rubric question, to facilitate interpretation, we provide examples of potential answers, while acknowledging that responses may indicate a work in progress.

Note: this rubric is a collection of suggestions to assess the FAIRness of plant-pollinator interaction datasets. However, we assert that ensuring data quality is an integral part of the research process, and it is up to researchers and their collaborators to develop their own guidelines to enhance the accessibility of plant-pollinator datasets for reuse.

Table 1. Rubric for assessment of plant-pollinator interactions data

Guiding Questions	**"Yes!" Example**	**"Not yet" Example**	**FAIR principle and Related FIP**
Q1. Is the data under review intended to be reused?[1] ♻️	Data include metadata that clearly outlines their purpose and potential reuse. E.g. SIB/Colombia, a member node of GBIF network, published their dataset documented using EML and DwC standards, so the community can reuse it.	Dataset containing raw data without any accompanying metadata or information on its context. E.g. USDA are mandated by US Congress to share data openly and are working to redesign their datasets to help do so. This takes time.	R.

[1] If the authors of the data under review have yet to consider the reusability of their data, we suggest that they consider this before proceeding with Q2-Q11.

Q2. Is the data under review digitised? 🔓	Data consists of digitised records or files accessible in electronic format. E.g. KALRO (Kasina et al. 2024) transcribed plant-pollinators records from literature and shared them in an electronic spreadsheet.	Data remains in non-digital formats such as handwritten notes or physical specimens. E.g. Herbarium or museum specimens with interaction information not yet digitised.	F, A.
Q3. Does the dataset under review use an already existing (meta)data standard (e.g. DwC, EML)? 🔍🔄♻️	The dataset adheres to the DwC and EML standards, ensuring interoperability with existing data systems. E.g. Wolowski et al. (2024) documented their flower visitation dataset using metadata terms from EML and labelled columns of the data table using DwC terms.	The dataset does not adopt any established (meta)data standard. E.g. The original data used in Carvalheiro et al., 2008 did not include metadata in EML or any other standard. Before publishing, it was necessary to create an EML/XML file with metadata extracted from the dataset spreadsheet.	F2, I1 Metadata, I1 Data, I1 Metadata, I1 Data, I2 Metadata, I2 Data, I3 Metadata, I3 Data, R1.3.
Q4. Does the metadata include at least the following information: dataset title, authors, licence, sampling methods and efforts, geographic, temporal and taxonomic coverage? 🔍♻️	The metadata includes all essential information to allow citation and reuse of the dataset. E.g. González-Vaquero and Devoto (2024) provided a complete metadata description about the data they are sharing in EML format[2].	The metadata lacks some essential information. Incomplete metadata hinder the dataset's reusability and interpretation. E.g. Alves et al. (2024) use the EML standard for documenting metadata, but it does not include the description of sampling methods and efforts, neither geographic, temporal and taxonomic coverage [3].	F2, R1.2 Metadata, R1.2 Data.
Q5. Is the data under	The data are published	Data are not published in a	A1.1 Metadata, A1.1

[2] See eml.xml published in: Nomer, & Elton. (2024h). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Plant-flower visitor interactions recorded in 49 sites in Argentina (Buenos Aires: Carlos Casares county) by Marcos Monasterolo (2013-15) and Antonio López Carretero (2016). [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10648048.

[3] See eml.xml published in Nomer, & Elton. (2024b). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Alves, Denise A. et al. 2023. Serviço ecossistêmico de polinização agrícola na cultura da laranja. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647830.

'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.

review *published* in a data repository or data journal like GitHub, Zenodo, Dryad, Figshare or Scientific Data, REBIPP? 🔍🔓♻️	in a data repository, ensuring long-term accessibility and citability through a persistent unique identifier. E.g. the original data used in Carvalheiro (2024) was previously published (metadata and data) into a GitHub repository accessible through the URL: https://github.com/globalbioticinteractions/carvalheiro2023.	data repository. Instead, it is only available in supplementary materials of a scientific paper. E.g. Martín Gonzalez et al. (2015) published PDF files with hummingbird–plant networks in the supplementary material of their paper.	Data, A1.2 Metadata, A1.2 Data, F1 Metadata, F1 Data, F2, F3, F4 Metadata, F4 Data, R1.1 Metadata, R1.1 Data.
Q6. Is the data under review *registered* with GBIF, GloBI or other registries? 🔍🔓🔄♻️	The data are accessible in a FAIR enabled registry. E.g. As part of WP10.2, we facilitated the registration of pilot data with GloBI. Note that some of Carvalheiro (2024) data was already available via the Database of Pollinator Interactions (DoPI). SIB/Colombia registered their dataset with GBIF (https://doi.org/10.15472/od8jpn).	The data is not registered in any registry. Without registration, a dataset's FAIR adherence is limited, reducing its potential for discovery and reusability. E.g. Thébault, & Fontaine. (2022) has been published in Zenodo (10.5281/zenodo.6630184) but it was not yet registered with GloBI (see https://github.com/globalbioticinteractions/globalbioticinteractions/issues/619).	F1 Metadata, F1 Data, F2, F3, F4 Metadata, F4 Data, A1.1 Metadata, A1.1 Data, A1.2 Metadata, A1.2, A1.1 Metadata, A1.1 Data, A1.2 Metadata, A1.2, I1 Metadata, I1 Data, I2 Metadata, I2 Data, I3 Metadata, I3 Data, R1.1 Metadata, R1.1 Data, R1.2 Metadata, R1.2 Data, R1.3.
Q7. Can the data under review be cited in a research paper? 🔍♻️	The data can be cited in a research paper, website, other datasets or any resource, and the citation includes a direct link to the (meta)data. E.g. Carvalheiro (2024). Plant-flower visitor network from Avon Gorge, UK. (https://zenodo.org/doi/10.5281/zenodo.10679321).	The data cannot be cited in a clear citation format. Others may struggle to properly acknowledge, reference and access the dataset. E.g. Silva & Ana (2024) Flower Pollination Dataset. Personal Communication[4].	F1 Metadata, F1 Data, R1.2 Metadata, R1.2 Data.

[4] Fictitious example to demonstrate a citation of an unpublished dataset.

Q8. Is the origin of the data under review documented? ♻️	The origin of the data is documented, providing detailed information about its processing and transformations. It includes how, when and where the data was collected/generated, and by whom. E.g. Carvalheiro et al. (2023) documented the origin of the data and published it at https://doi.org/10.5281/zenodo.10530109.	The origin of the data is not adequately documented. It lacks information about its origin, processing steps, or any transformations applied to the data. E.g. Alves et al. (2024) provided insufficient metadata details for anyone interested in reusing the data to access the origins of the dataset.	R1.2 Metadata, R1.2 Data.
Q9. Has the data been described with sufficient precision (using well-defined terms) to enable others to understand and reuse it independently, without needing to contact the owner for clarification on its meaning? 🔍♻️	The data has been described using terms from established (meta)data standards or precisely defined concepts not yet incorporated by existing standards. It ensures the reduction of ambiguity and the need for additional clarification from the data owner. E.g. Wolowski et al. (2024) used EML for documenting metadata; DwC for taxonomic and spatiotemporal data documentation; and PPI vocabulary for documenting the sexual system of recorded plants.	The data contain vague or poorly defined terms, making it challenging to interpret without seeking clarification from the data owner. E.g. Before the FAIRification process presented in Deliverable 10.2, the dataset that originated Carvalheiro (2024) contained variables with ambiguous definitions. After consulting the data owner, the variables were mapped to terms of existing data standards.	F2, R1.1 Metadata, R1.1 Data, R1.2 Metadata, R1.2 Data, R1.3.
Q10. Does the description of the dataset under review clearly outline how others may (or may not) reuse the data?	The dataset description provides detailed information on usage rights, licences and any restrictions imposed by the data owner. Additionally, it may	The dataset description does not (clearly) outline how others may reuse the data. It leaves others uncertain about the legal and ethical implications of data reuse. E.g. Tinoco (2024) adopts EML	R1.1 Metadata, R1.1 Data.

♻️	provide guidance on proper attribution, citation requirements and potential ethical considerations associated with data reuse. E.g. Carvalheiro (2024) clearly includes in the metadata description the licence applied to their dataset (Creative Commons Attribution 4.0 International) and a link to its definition (https://creativecommons.org/licenses/by/4.0/).	as metadata standard, but no usage rights, licence or restrictions are documented.	
Q11. Was a specific set of standardised terms, like those from the Relation Ontology, used to describe the types of interactions, or other existing domain-specific resources like the Plant-Pollinator Interactions Vocabulary? 🔄	A specific set of standardised terms was used to provide a common terminology across datasets, ensuring consistency and interoperability. E.g. Wolowski et al. (2024) used Relation Ontology to document the interaction types ("pollinated by", "has flowers visited by"), and the PPI vocabulary for documenting the sexual system of recorded plant species.	The dataset relies on ad-hoc or non-standardised terminology, leading to inconsistencies and ambiguity in the data interpretation. E.g. Prior to the FAIRification of the dataset provided by Alves et al. (2024), the data contained ad-hoc and non-standardised column labels. Additionally, the column labels were written in Portuguese, limiting the interpretation of the dataset to Portuguese speakers.	I1 Metadata, I1 Data, I2 Metadata, I2 Data, I3 Metadata, I3 Data.
🔍 Findable 🔓 Accessible 🔄 Interoperable ♻️ Reusable			

'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. Dataset assessments

In the discovery phase of the Agricultural Biodiversity Case Study (reported in WorldFAIR Deliverable 10.1), we presented an overview of data practices in studies of pollinators and their interactions with agricultural crops and other plant species within or outside agroecosystems. We found a total of 8,768 unique datasets in a highly fragmented data landscape and performed an analysis of some of the aspects of the FAIR principles: licences, metadata completeness, persistent identifiers and data file formats. In this section, we map the datasets' characteristics to some of the rubric questions.

Q1. Is the data under review intended to be reused? ♻️

No. The metadata connected to the datasets include administrative information (e.g., email address, author names) and bibliographic information. However, taxonomic, geospatial and temporal coverage of the data is either missing or incomplete. In addition, documentation of the interaction types (e.g., visits flower of, pollinates), and habitat information (e.g., rain forest, savannah) may be included in the (human-readable) abstract, but not available in machine-readable formats in the related metadata fields. The incomplete metadata we observed hints at opportunities to improve ways to extract machine-readable fields from abstracts and/or underlying data instead of relying on the authors to manually enter these metadata fields. Note that in the 8,768 datasets we analysed in D10.1, we found that less than 5% of the datasets use EML. This implies that 5% or less of the surveyed datasets included information like taxonomic coverage in their metadata.

Q2. Is the data under review digitised? 🔓

Yes. We only surveyed digital data. All 8,768 datasets were digital.

Q3. Does the dataset under review use an already existing (meta)data standard (e.g. DwC, EML)? 🔍🔄♻️

As mentioned earlier, less than 5% of the 8,768 datasets used (meta)data standards like DwC and/or EML.

Q4. Does the metadata include at least the following information: dataset's title, authors, licence, sampling methods and efforts, geographic, temporal and taxonomic coverage? 🔍♻️

None of the datasets include *all* the metadata information listed. Fields describing dataset title and authors are present, but metadata fields on geographic, temporal and taxonomic coverage are missing or incomplete.

Q5. Is the data under review published in a data repository or data journal like GitHub, Zenodo, Dryad, Figshare, or Scientific Data, REBIPP? 🔍🔓♻️

Yes. All the datasets we found are in one of these repositories. As listed in table 1 of Deliverable 10.1[5], Figshare accounts for the majority (i.e. 8,563 of 8,768 datasets).

Q7. How can the data under review be cited in a research paper? 🔍♻️

We found that 75% of all datasets have a DOI attached to them. However, many of them are referring to the published article and not to 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.

Q10. Does the description of the dataset under review clearly outline how others may (or may not) reuse the data? ♻️

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 what cannot be done with the data.

4. Automated FAIR-enabled data reviews

4.1. Harnessing the power of machine-actionable metadata 🐝

Peer review is a crucial part of scholarly communication: scientific journals and their editorial boards should only publish submitted articles after review. Scientific article style guidelines and formatting requirements aim to facilitate the review and publication process. For instance, by limiting the total number of words allowed, authors produce works fit for review and publication. Similarly, the authors are required to publish their work using correct spelling and grammar. These requirements are integral to the scientific publication process, and seem like a consensus - who would want to review or publish a paper full of typos?

However, when applied to *scientific data,* the expectations of a review process are often unclear or even ignored. For instance, are the reviewers supposed to check supplementary data for inconsistencies? If so, do the editors provide adequate guidelines on what should be reviewed? Is it allowed to publish tabular data in a proprietary document format like DOCX (Microsoft Word documents) or PDF (Portable Document Format) files? What are the parameters for evaluating the quality of a dataset? It is worth noting that many scientific journals do not require a particular structure or format for the supplementary sections or appendices, which is usually where the row data is presented.

In this section, we show a way to review plant-pollinator datasets. These spell-checks (or data reviews) aim to facilitate data review at all stages of the publication process: just like an author

[5] Trekels, M., Pignatari Drucker, D., Salim, J. A., Ollerton, J., Poelen, J., Miranda Soares, F., Rünzel, M., Kasina, M., Groom, Q., & Devoto, M. (2023). WorldFAIR Project (D10.1) Agriculture-related pollinator data standards use cases report (Version 2). Zenodo. https://doi.org/10.5281/zenodo.8356529

'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.

would use a spell-check on their methods section in their text processing program, we show examples of automated processes that help catch and highlight inconsistencies in *scientific data* automatically.

The data spell-check (or review) process is outlined as follows:

Step 1. A researcher creates and registers their plant-pollinator dataset according to GloBI guidelines[6] (described in Appendix IV).

Step 2. An automated GloBI review "bot" generates a human-readable review report (see Appendix V for documentation).

Step 3. The review report is inspected by the researcher (or their peers).

Step 4. A new version of the dataset is created if desired/needed, triggering a new review (S2).

By design, the outline of the data review process above aligns with the FAIR principles: by completing a review process, the researcher has shown that their data can be **f**ound, **a**ccessed, **i**ntegrated, and **r**eused by a (naive) review bot. So, we can consider this data spell-check process to be a FAIR assessment. Similar to the (self-)assessment provided by the Rubric, our automated review process produces answers to questions like: Is the type of species interaction (e.g., pollination, flower visitation) specified? Or, does the scientific name of a plant appear in distinct taxonomies (e.g. Catalogue of Life - CoL[7] and others)? In the case of reviews included in this report, the Nomer Corpus of Taxonomic Resources (Poelen, 2022) was used. This corpus includes versioned copies of taxonomic resources including Catalogue of Life, GBIF Backbone Taxonomy, Integrated Taxonomic Information System (ITIS), National Center for Biotechnology Information (NCBI) Taxonomy, Index Fungorum, DiscoverLife, World of Flora Online and more.

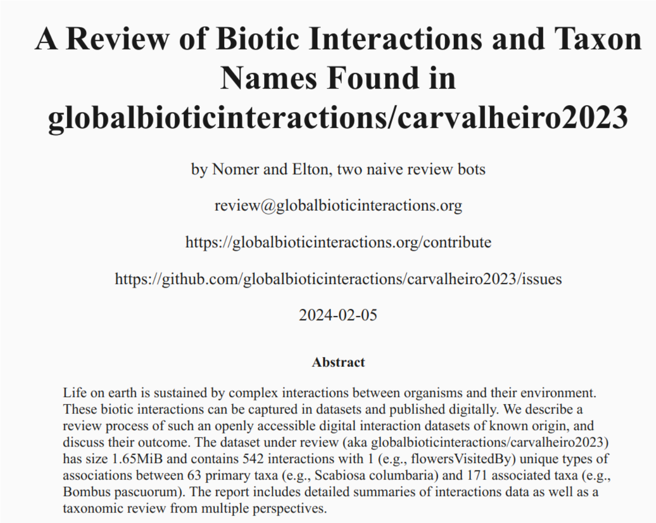
In the subsection below, we highlight two data reviews of WorldFAIR Work Package 10's Pilots Study: Carvalheiro (2024), and Kasina et al. (2024), as produced by Nomer and Elton, two naive GloBI review bots, on 5 February 2024.

To make the data review report readable for researchers, GloBI's bots produce a document resembling a data publication: a document with a title, authors, publication date, abstract, introduction, and so forth.

[6] This process is described in the GloBI webpage https://www.globalbioticinteractions.org/contribute and was explored in Deliverable 10.2. See also the cookbook “Guidelines and Recommendations for Publishing Agricultural-related pollinator data”.

[7] https://www.catalogueoflife.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.



A Review of Biotic Interactions and Taxon Names Found in globalbioticinteractions/carvalheiro2023

by Nomer and Elton, two naive review bots

review@globalbioticinteractions.org

https://globalbioticinteractions.org/contribute

https://github.com/globalbioticinteractions/carvalheiro2023/issues

2024-02-05

Abstract

Life on earth is sustained by complex interactions between organisms and their environment. These biotic interactions can be captured in datasets and published digitally. We describe a review process of such an openly accessible digital interaction datasets of known origin, and discuss their outcome. The dataset under review (aka globalbioticinteractions/carvalheiro2023) has size 1.65MiB and contains 542 interactions with 1 (e.g., flowersVisitedBy) unique types of associations between 63 primary taxa (e.g., Scabiosa columbaria) and 171 associated taxa (e.g., Bombus pascuorum). The report includes detailed summaries of interactions data as well as a taxonomic review from multiple perspectives.

Figure 2. First page of a data review generated by GloBI bots "Nomer" and "Elton" on 5 February 2024 (Nomer & Elton, 2024d).

In the table below, some examples from the review report are highlighted in the context of an associated review question and FAIR principle as applied to various examples.

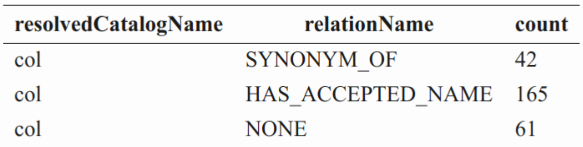
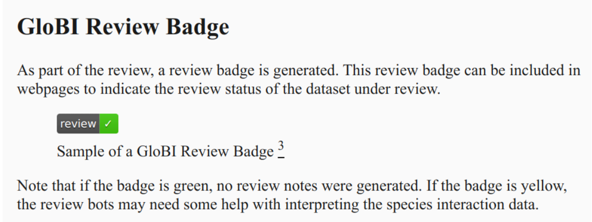
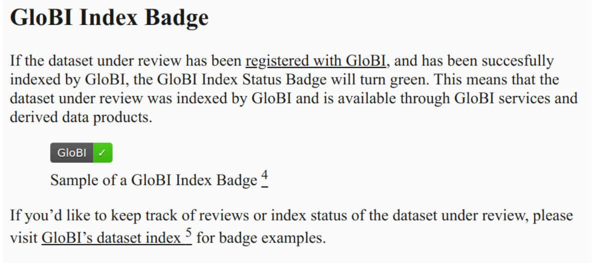
4.1.1 Data Review Example: Carvalheiro (2024)

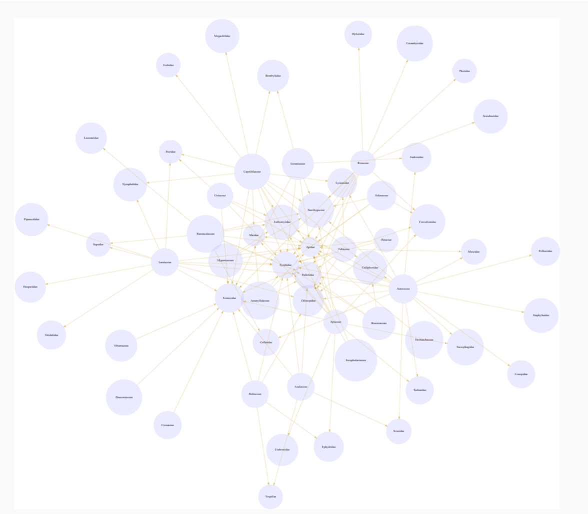
Carvalheiro (2024) decided to make their metadata, data, and review publicly available. This is why we can show examples from their data review as generated on 5 February 2024. See Appendix II for the full review report and references to associated data review products.

Table 2. Data Review of Carvalheiro (2024) by Nomer & Elton (2024d)

🔍 Findable 🔓 Accessible 🔄 Interoperable ♻️ Reusable

Review Question	Answer	Example
R1 Can the data under review be found and accessed? 🔓🔍	The review abstract summarises the plant-pollinator dataset - so yes!	"[...] contains 542 interactions with 1 (e.g., flowersVisitedBy) unique types of associations between 63 primary taxa (e.g., *Scabiosa columbaria*) and 171 associated taxa (e.g., *Bombus pascuorum*) [...]"

Review Question	Answer	Example
R2 Are the taxonomic names recognised by distinct taxonomies like Catalogue of Life and others? 🔄	Yes, and no: according to the (naive) bots and CoL, 165 names are accepted, 42 are synonyms. And for some reason, 61 names were not recognised. Also see the Taxonomic Name Alignment section.	 resolvedCatalogName · relationName · count col · SYNONYM_OF · 42 col · HAS_ACCEPTED_NAME · 165 col · NONE · 61
R3 Can the data be reused? ♻️	The review report shows that "naive" bots can produce different kinds of data summaries. These summaries are examples of reuse that support the claim that the dataset under review can, in fact, be reused. And, if bots can reuse the data, humans can do it too.	Appendix II and a "green" review badge suggest that the data were successfully reused.  **GloBI Review Badge** As part of the review, a review badge is generated. This review badge can be included in webpages to indicate the review status of the dataset under review. review ✓ Sample of a GloBI Review Badge [3] Note that if the badge is green, no review notes were generated. If the badge is yellow, the review bots may need some help with interpreting the species interaction data.
R4 Is the data findable through GloBI? 🔍	Yes! A green GloBI badge is shown in the review report. This indicates that at the time of generation of the report, the (meta)data under review was indexed, searchable, and included in GloBI-interpreted data products.	Compare the section in Appendix II reporting on the index status of the dataset under review.  **GloBI Index Badge** If the dataset under review has been registered with GloBI, and has been succesfully indexed by GloBI, the GloBI Index Status Badge will turn green. This means that the dataset under review was indexed by GloBI and is available through GloBI services and derived data products. GloBI ✓ Sample of a GloBI Index Badge [4] If you'd like to keep track of reviews or index status of the dataset under review, please visit GloBI's dataset index [5] for badge examples.

Review Question	Answer	Example
R5 Can a network diagram be generated from the data under review? 🔄♻️	Yes! The review report contains a network diagram connecting the reported interactions on the family (or user specified) level. Note that this includes only the names accepted by the CoL, showing the interoperability of the taxonomic name in the dataset under review.	 Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

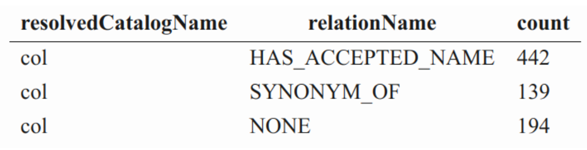
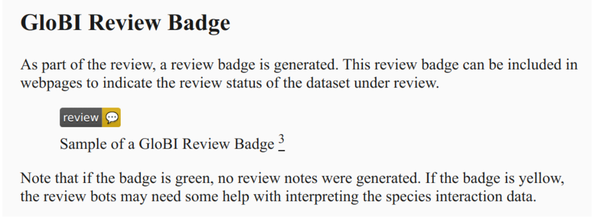
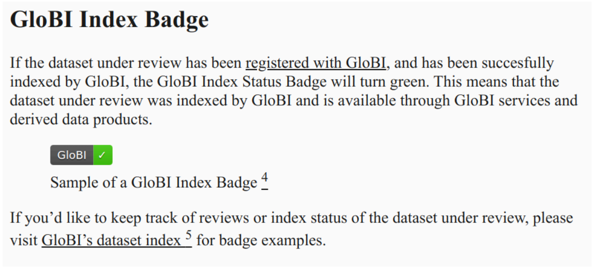
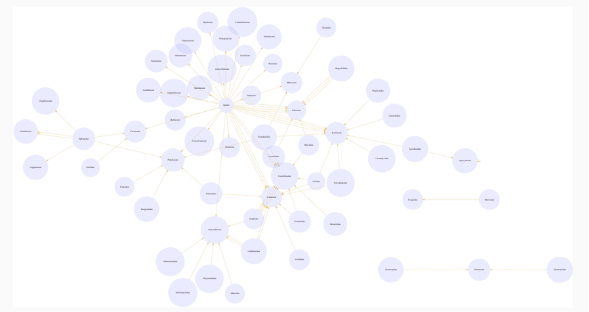
4.1.2. Data Review Example: Plant-pollinator data from Kasina et al. (2024) of Kenya Agricultural and Livestock Research Organization

As part of the Kenya Agricultural and Livestock Research Organization (KALRO), Kasina *et al*. (2024) openly shared "A review of the status of web-based African Plant-Pollinator Interaction data," associated metadata, and the review reports.

Table 3. Data Review Example of Kasina et al. (2024) by Nomer & Elton (2024f)

🔍 Findable 🔓 Accessible 🔄 Interoperable ♻️ Reusable

Review Question	Answer	Example
R1 Can the data under review be found and accessed? 🔓🔍	The review abstract summarises the plant-pollinator dataset - so yes!	"[...] contains 1,023 interactions with 8 (e.g., pollinates) unique types of associations between 512 primary taxa (e.g., Apis mellifera) and 331 associated taxa (e.g., Persea americana) [...]"

Review Question	Answer	Example
R2 Are the taxonomic names recognised by the Catalogue of Life (CoL)? 🔄	Yes, and no: according to the (naive) bots and CoL, 442 names are accepted, 139 are synonyms. And for some reason, 194 names were not recognised. Also see the Taxonomic Name Alignment section.	 resolvedCatalogName | relationName | count col | HAS_ACCEPTED_NAME | 442 col | SYNONYM_OF | 139 col | NONE | 194
R3 Can the data be reused? ♻️	The review report is evidence that a third party can reuse the data.	Appendix II and a "yellow" review badge suggests that the data was reviewed and some review notes were generated.  **GloBI Review Badge** As part of the review, a review badge is generated. This review badge can be included in webpages to indicate the review status of the dataset under review. review 💬 Sample of a GloBI Review Badge [3] Note that if the badge is green, no review notes were generated. If the badge is yellow, the review bots may need some help with interpreting the species interaction data.
R4 Is the data findable through GloBI? 🔍	Yes! According to the review document, the dataset was indexed by GloBI search.	Section in Appendix II reporting on the index status of the dataset under review.  **GloBI Index Badge** If the dataset under review has been registered with GloBI, and has been succesfully indexed by GloBI, the GloBI Index Status Badge will turn green. This means that the dataset under review was indexed by GloBI and is available through GloBI services and derived data products. GloBI ✓ Sample of a GloBI Index Badge [4] If you'd like to keep track of reviews or index status of the dataset under review, please visit GloBI's dataset index [5] for badge examples.
R5 Can a network diagram be generated from the data under review? 🔄♻️	Yes! The review report contains a network diagram connecting the reported interaction on the family (or user specified) level. Note that this includes names recognized by the CoL, showing the interoperability of the taxonomic name in the	 Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

Review Question	Answer	Example
	dataset under review.	

4.1.3. Taxonomic name alignment

In our examples shown in Tables 2 and 3, we highlighted name alignment review results for pilots Carvalheiro *et al.* (2024) and Kasina *et al.* (2024). These results included the number of accepted, synonyms and unrecognised names as determined by a defined taxonomic resource, with detailed tabular views available when needed.

Taxonomic names are essential to understand the kind of organisms that interact, and careful review of these names help catch typos (e.g., *Apis mellifera* vs *Apis melifera*) and alternative names that refer to a single species (e.g., *Apis mellifera* and *Apis mellifera* Linnaeus), that may be easy for bots to spot, but hard to detect by their human counterparts. For example, in the review report for Kasina *et al.* (2024), *Apis mellifera* appears under at least four different names in the dataset (see Table 4 of the corresponding review). This topic was also explored in Deliverable 10.1 (Trekels *et al.* 2023).

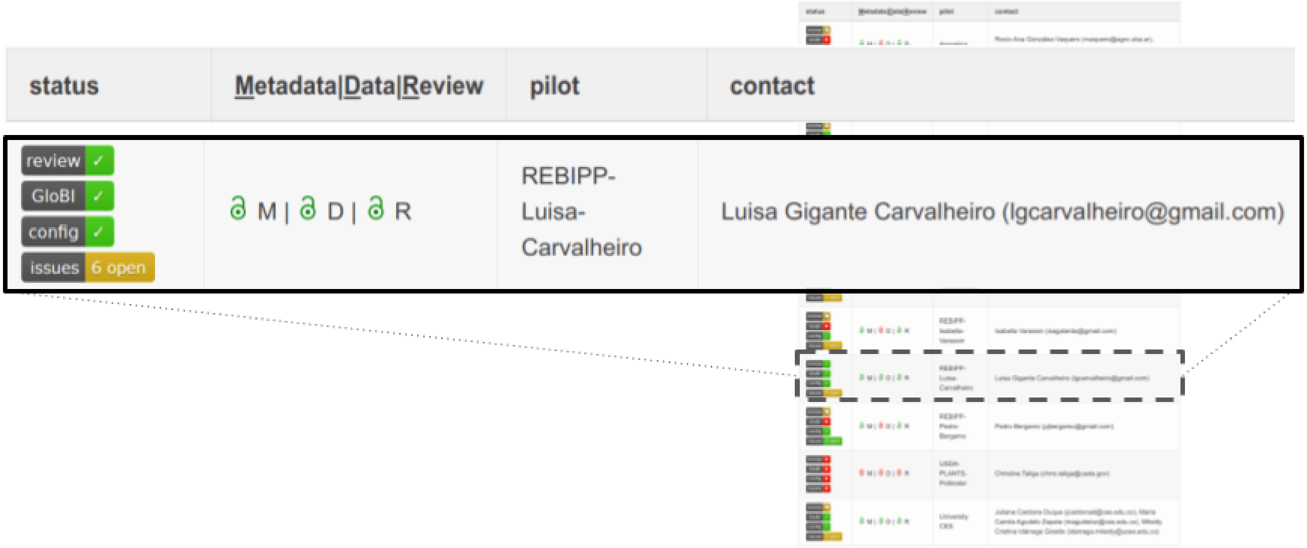
By comparing against different taxonomic resources side-by-side[8], biases and differences in their alignments become apparent. For example, the number of unrecognised names is different across the taxonomic catalogues and may highlight potential taxonomic gaps in the sources. Valid names of rare/endemic flora and pollinators may come up as unrecognised due to (implicit) geographical biases in a given published taxonomic catalogue. For example, when reviewing a plant dataset from Brazil, reviewing the names using a regional checklist may benefit the quality of the review in that more valid names are recognised correctly, increasing the true positive quality of the taxonomic "spellcheck". This is why including both regional and transnational taxonomic name references is needed to get a comprehensive view on the interoperability of the names in the dataset under review. Also, including taxon information from revisionary papers could help to strengthen the taxonomic trees (e.g. see for cyclanths pollinators or floral visitors two contrasting cases: https://www.gbif.org/pt/species/124558104 - taxon already recognized when included in datasets; instead of https://www.gbif.org/occurrence/1932562261 - taxon with issues in “Taxon match higherrank”). Biologists know strengths and weaknesses of published taxonomic resources, but their preferences cannot typically be accommodated in available biodiversity informatics infrastructures. Adding the ability to perform a taxonomic alignment from various perspectives is a key feature that needs to be supported by existing biodiversity informatics infrastructure, and we show, through our automated reviews, that this can be done.

[8] GloBI review reports include alignments with the Catalogue of Life, Integrated Taxonomic Information System (ITIS), NCBI Taxonomy, GBIF Backbone Taxonomy, and more.

‘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.

4.1.4. Review of restricted data

Due to institutional policy, research embargo, or commercial reasons, among others, some plant-pollinator datasets are not shared openly. Because the FAIR principles allow for access restrictions – see sub-principle A1.2 (GO FAIR, 2024) – some of these restricted datasets may still be considered FAIR-compliant. In Figure 3, the variety of openness in the WorldFAIR pilots is shown. Note, however, that "red" does not imply that a dataset is “bad” or “un-FAIR”. For instance, Wolowski (2024) opted to share their metadata and review but not their raw data. This means that GloBI cannot index Wolowski´s data in its open search index. However, as you can see in Appendix II (Nomer & Elton, 2024c), the publicly-available review report of a sample of their data suggests that their data are structured such that it can be reused for those who have access to it. For other pilots, the review tools can be used internally among people with access rights. So, while evidence to support the dataset FAIRness cannot be shared publicly, they may benefit from the assessment we outlined here. For instance, the USDA Plant Pollinator Pilot opted to keep their (meta)data and review private prior to an associated scholarly publication. However, through personal communication, we were able to exchange an automatically-generated data review report to facilitate internal review.



status	Metadata|Data|Review	pilot	contact
review ✓ GloBI ✓ config ✓ issues 6 open	M | D | R	REBIPP-Luisa-Carvalheiro	Luisa Gigante Carvalheiro (lgcarvalheiro@gmail.com)

Figure 3. Screenshots of the GloBI WorldFAIR status page (https://globalbioticinteractions.org/worldfair, as accessed on 9 February 2024) show the various aspects of our pilots. In the "status" column, clickable badges show the public availability of the review report review ✓*, searchability in GloBI* GloBI ✓*, the metadata registration* config ✓*, and the number of active discussions* issues 6 open*. The next column shows the openness choices made by the pilots. In the example of the Carvalheiro pilot, the metadata* M *, data* D *, and review* R *were made openly accessible. Finally, a short description of the pilot and their contacts are included.*

‘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.

Another example of restricted data comes from HiveTracks, which utilises their mobile app to enable beekeepers to collect beekeeper-reported hive and apiary observations across a wide range of locations. This, in turn, provides direct observation data on some interactions (pollinator stressors and the presence of pollinators / floral resources) and provides evidence for inference on other interactions. For example, pollination events can be inferred through the documented flora combined with the honey produced within the same area (i.e., as per the honey harvest record). These data points have also been mapped to the REBIPP template to demonstrate how HiveTracks' data adhere to the Darwin Core standard, as described in Drucker et al. (2024). Specifically, HiveTracks' data show interaction patterns starting with inferred pollination events and extending into other interaction patterns, such as *Varroa ssp.* mite / pollinator interactions - in this case, a different type of interaction. Given that beekeepers' hive locations are considered trade secrets, which is specified in HiveTracks' privacy policy, the HiveTracks Plant Pollinator Pilot opted to keep their (meta)data and review private, and opted to use sample data for their mapping process.

4.1.5. Nuts and bolts of the GloBI review process

The review reports in this document were generated using open-source software such as GloBI's Elton[9] (for parsing interaction data) and Nomer[10] (for name alignment), as well as commonly-used tools like pandoc[11] (for generating documents from structured markdown), and various Linux command-line tools (e.g., sed, cat, awk, grep). Also, the review workflow using these tools is openly available through a bash script named "*check-dataset.sh*[12]". The openness of the review workflow and its associated tools allows for executing the GloBI review process on private servers, but also on infrastructures such as GitHub Actions workflows. This not only enables individuals to review their own private data using publicly-available tools but also allows for fast automated review cycles through the GitHub Action-driven review process. Analogous to how current text processors can perform spell-checks on the fly or suggest grammatical improvements, real-time data reviews can assist researchers in creating FAIR data, irrespective of whether they are open or not.

4.1.6. Pointing at datasets with (aspirationally) persistent identifiers

Scholarly citation is a cornerstone of scholarly communication, and the FAIR principles place "persistent identifiers" front and centre as a preferred method to point to data. In fact, the first FAIR principle, "F1: (Meta) data are assigned globally unique and persistent identifiers", makes the concept of identifiers an essential building block of "FAIR" data. However, studies have shown

[9] https://github.com/globalbioticinteractions/elton - helps to access, review and index existing species interaction datasets.

[10] https://github.com/globalbioticinteractions/nomer - maps identifiers and names to other identifiers and names.

[11] https://pandoc.org - a universal document converter.

[12] "check-dataset.sh" is part of Daniel Mietchen, Jorrit Poelen, & Katja Seltmann. (2024). globalbioticinteractions/globinizer: 0.4.0 (0.4.0). Zenodo. https://doi.org/10.5281/zenodo.10647565

'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.

(Philipson, 2018; Elliott et al. 2020; Elliott et al. 2023) that commonly used identifiers like DOIs and URLs, as popular as they may be, are unreliable and *unverifiable* references to digital data.

But not all is lost: these crucial links between identifiers and their data can be strengthened by constantly examining, monitoring, recording, and reviewing the relations between an identifier (e.g., a DOI, ARK, LSIDs) and their associated digital data. And, we may want to consider taking advantage of commonly-used cryptographic techniques by adding digital fingerprints (or cryptographic hashes, checksums) into our scientific citations (Elliott et al. 2023) to *verifiably* identify immutable digital dataset associations. When using these digital fingerprints in combination with identifiers such as DOIs, we can benefit from existing internet infrastructure (e.g., dynamically redirecting to a human-readable web page) while making sure that the digital data are ready for a future beyond the internet.

In short, we suggest indicating caution in using the term "persistent identifiers" by placing "(aspirationally)" in front of it, as well as imagining a future beyond the internet by asking the question: How will you, or future generations, find that cited plant-pollinator dataset 50 years from now?

4.1.7. Towards a virtuous review cycle

To help alleviate the burden of manually reviewing data as part of reviewing scientific publication, we propose to deploy domain-specific, automated data review processes to help researchers better understand how they can make their data easier to review and reuse. Recognising that publishing reusable, integrated data remains mostly a manual process, we recommend plant-pollinator datasets in specific, and species interaction datasets in general, to register with one or more infrastructures (e.g., GloBI, GBIF) to benefit from the (domain-specific) data review services they offer. Also, we envision that (data) publishers continue to collaborate, or even build/maintain, similar infrastructures to assess, and hopefully increase, the quality (and FAIRness) of published scientific data.

5. Recommendations

Recommendation 1 - Use and leverage existing biodiversity data infrastructures.

a) Type (choose as many as apply):
   - ○ Policy
   - ○ Organisational

b) the stakeholder(s) at which the recommendation is aimed: Researchers, Research Performing Institutions, Data Producers, Journal Editors, and Publishers

Domain-specific data infrastructures such as GBIF, GloBI, and REBIPP provide several services and result in a higher adherence to the FAIR principles, enabling data reusability, particularly

'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.

plant-pollinator interactions data. We recommend plant-pollinator datasets specifically, and species interaction datasets in general, to register with one or more infrastructures (e.g., GloBI, GBIF) to benefit from the (domain-specific) data review services they offer. Also, we envision that (data) publishers continue to collaborate, or even build/maintain, similar infrastructures to *continuously* assess, and hopefully increase, the quality (or FAIRness) of published scientific data. We consider a FAIR assessment essentially time-dependent: a dataset may become less FAIR due to the degradation of digital resources. Or, stated more optimistically, a dataset may increase in FAIRness as their curators continue to exercise their ability to make their digital knowledge easier to reuse, and receive more feedback from the community of users.

Recommendation 2 - Catalyse existing efforts to help promote a diverse community of users from different backgrounds

a) Type (choose as many as apply):
   ○ Policy
   ○ Organisational
b) the stakeholder(s) at which the recommendation is aimed: Global Biodiversity Information Facility (GBIF)

We recommend increasing the leading role that GBIF plays in biodiversity data, information and knowledge worldwide by leveraging existing efforts and helping to promote a diverse community of users from different backgrounds. This means including a large community of infrastructures and tools tailored by the users and allowing for community contributions to the development of FAIR data tools and services, allied to maintaining its own contribution to data publishing. New ideas and initiatives are being developed continuously and GBIF could play a central role as an integration point of this ecosystem of tools and services to accommodate local needs. For instance, as demonstrated in this report: currently, operations like taxonomic alignment and taxonomic name parsing cannot be customised when searching for data in the GBIF infrastructure - users are expected to adopt a single taxonomic perspective even though this set perspective (e.g., Catalogue of Life/GBIF Backbone) is known to cause incomplete and biased results. We would like to encourage GBIF to expand its services to facilitate not only data sharing but also to foster the development of reusable and open biodiversity data tools to better make use of our growing global biodiversity informatics community and its cross-disciplinary collaborators. Examples of such tools include, but are not limited to: (i) high-performance, offline-enabled, taxonomic name alignment tools supporting many taxonomic perspectives; (ii) version tracking of original data, not just interpreted data; (iii) re-packaging and publishing assemblies of original datasets for reuse; and (iv) tracking annotations on records in existing versioned datasets. By embracing this collaborative approach, essential services like a real-time biodiversity data review (as we proposed in D10.3) can emerge organically in a diverse group of researchers and software/data engineers. Our pilot studies have shown that we need localised tools and local talents to work with local data to promote FAIRness globally. And, to help facilitate this, we need global infrastructures like GBIF to go beyond the extraction of local datasets, and help to develop the skills and tools needed to process these

'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.

data into a locally curated, globally-connected, and increasingly FAIR corpus of global digital biodiversity knowledge.

6. Conclusions

This report presents results from the pilot studies of the Agricultural Biodiversity Case Study (WP10), complementing previous efforts with Agricultural Biodiversity FAIR data assessment rubrics. The FAIR assessment tools described in this report, tailored to the plant-pollinator interactions domain, assist researchers and institutions in evaluating adherence to the FAIR principles. We believe that these tools help encourage data publishing and reuse in the plant-pollinator interactions domain, moving away from diverse approaches and isolated initiatives toward widely available plant-pollination interactions FAIR data for scientists and decision-makers. This shift enables the development of integrative studies that enhance our understanding of species biology, behaviour, ecology, phenology, and evolution.

In conjunction with Deliverables 10.1 and 10.2, this work significantly contributes to promoting plant-pollinator interaction data interoperability and availability for reuse, which is the ultimate goal of the Agricultural Biodiversity Case Study. With examples from our pilot studies representing initiatives in Europe, South America, Africa, North America and elsewhere, our tools facilitate FAIR assessments and highlight best practices developed throughout the WorldFAIR project. This approach helps to understand the FAIR principles in a domain-specific manner. We are confident that this effort can assist similar initiatives in embracing interoperability standards within this domain, aligning with the FAIR principles. Through the adoption of standards such Ecological Metadata Language, Darwin Core, Plant-Pollinator Interactions Vocabulary and Relation Ontology, we aim to enhance the understanding of how plant-pollinator interactions contribute to sustaining life on Earth while ensuring that data is easily discoverable, accessible, and reusable for further research and analysis.

‘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. Appendix I. FAIR Implementation Profile (FIP) for plant-pollinator interactions data

A FAIR Implementation Profile (FIP) outlines technology choices to uphold FAIR Guiding Principles, decided collectively by a community. The FIP Wizard[13] captures FIPs via a questionnaire answered by a Community Data Steward. Published as FAIR and Open data, FIPs serve as a reference for FAIR data stewardship. This encourages reuse, saving time and promoting convergence on FAIR implementation. FIPs are periodically revised to reflect community needs and technological advancements, making FAIRification more structured and efficient (Schultes et al. 2020). Here we present the WorldFAIR WP10 Plant-Pollinator FIP01, with additional comments and examples that are not included in the original WP10 FIP published by Drucker (2022).

FAIR Implementation Community: WorldFAIR WP10 Plant-Pollinator Community.

Community Data Steward: Debora Pignatari Drucker (ORCID: 0000-0003-4177-1322).

Start date for the validity of the FIP: 2022-08-08.

End date for the validity of the FIP: 2024-06-30.

7.1. Declarations for Findability

Declaration F1 Metadata: Globally unique, persistent, resolvable identifier service for metadata records

A wide array of persistent identifier types were utilised in the pilots, including but not limited to: hashes, DOIs, URIs, etc. Examples:

10.1016/j.biocon.2005.12.009 (DOI from USDA)

mailto:Chris.Taliga@usda.gov (email URL USDA)

http://purl.obolibrary.org/obo/RO_0002455 (purl from Universidad CES)

https://sciendo.com/article/10.2478/jas-2013-0004 (url from KALRO)

10.2478/jas-2013-0004 (doi from KALRO)

UCES:CBUCES:F122 (darwin core triple from Universidad CES) etc.

[13] https://fip-wizard.ds-wizard.org/wizard/

‘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.

Declaration F1 Data: globally unique, persistent, resolvable identifier service do you use for datasets

Digital object type: Persistent Identifier

Type	DOI
Provider	Zenodo
How	“Zenodo will automatically register a Digital Object Identifier (DOI) for a record once you publish it. The DOI is a globally unique persistent identifier which ensures that the record can be uniquely cited which is important for reproducibility and attribution of credit. Zenodo register DOIs with DataCite.” (Zenodo, 2024).
Examples	https://doi.org/10.5281/zenodo.8176978

Digital object type: Persistent Identifier

Type	URI
Provider	GBIF
How	Each observation record is assigned a unique GBIF identifier.
Examples	https://www.gbif.org/occurrence/4507695028

Digital object type: Persistent Identifier

Type	Hash
Provider	GloBI
How	
Examples	hash://sha256/dec6efdd95fd64d5c38480e0db0dfa329c94e8e0fc0736f0769cafb470fd13ce\

‘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.

Declaration F2: Metadata schemas for findability

The following schemas have been used to annotate plant-pollinator interactions **data sets** in the pilots, and are recommended by this WorldFAIR Deliverable 10.3.

Digital object type: Metadata schema

Name	Ecological Metadata Language (EML)
Namespace	https://eml.ecoinformatics.org/eml-2.2.0
Description	"The Ecological Metadata Language (EML) metadata standard was originally developed for the earth, environmental and ecological sciences. It is based on prior work done by the Ecological Society of America and associated efforts. It has been developed to document any research data, and as such can be used outside of these original subject areas. EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset." (FAIRsharing Team, 2015).

Digital object type: Metadata schema

Name	Plant-Pollinator Interactions Vocabulary (PPI)
Namespace	https://ppi.rebipp.org.br/terms/
Description	PPI is a "vocabulary of terms and a data model for sharing plant–pollinator interactions data based on the Darwin Core standard. The vocabulary introduces 48 new terms targeting several aspects of plant–pollinator interactions and can be used to capture information from different approaches and scales. Additionally, we provide solutions for data serialisation using RDF, XML, and DwC-Archives and recommendations of existing controlled vocabularies for some of the terms. Our contribution supports open access to standardized data on plant–pollinator interactions." (Salim et al. 2022).

Digital object type: Metadata schema

'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.

Name	Darwin Core (DwC)
Namespace	http://rs.tdwg.org/dwc/terms/
Description	"Darwin Core is a standard for sharing data about biodiversity – the occurrence of life on earth and its associations with the environment." (Wieczorek et al., 2012).

Regarding metadata records, we initially employed the EML schema due to its compatibility with GloBI. However, in line with the principles of cross-domain interoperability (CDIF Working Group et al., 2023), we proposed a mapping between EML and more generic standards like Schema.org and DCAT. This mapping is available on GitHub[14] (Drucker et al., 2024). Considering this mapping, alternative schemas are recommended in addition to EML. It is important to note that despite these alternatives, a metadata record in EML remains necessary for interoperability with platforms such as GloBI and GBIF.

Digital object type: Metadata schema/ontology

Name	Schema.org
Namespace	https://schema.org/

Digital object type: Metadata schema

Name	Data Catalog Vocabulary (DCAT)
Namespace	http://www.w3.org/ns/dcat#

Digital object type: Ontology

Name	Friend of a Friend (FOAF) Ontology
Namespace	http://xmlns.com/foaf/0.1/

Digital object type: Metadata schema

[14] URL: https://github.com/globalbioticinteractions/carvalheiro2023/issues/1#issuecomment-1855661190 (retrieved on 02/09/2024).

'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.

Name	Dublin Core
Namespace	http://purl.org/dc/terms/

Declaration F3: Schemas used to link the persistent identifiers of the data to the metadata description

Digital object type: Ontology

Name	DataCite Ontology
Namespace	http://purl.org/spar/datacite
Description	The DataCite Ontology (DataCite) is an ontology that enables the metadata properties of the DataCite Metadata Schema Specification (i.e., a list of metadata properties for the accurate and consistent identification of a resource for citation and retrieval purposes) to be described in RDF.

Declaration F4 Metadata: Services to publish metadata records

GloBI and GBIF function as metadata registries, indexing metadata records to enhance data discoverability.

Digital object type: Metadata registry

Resource	GBIF search engine
URL	https://www.gbif.org/
Description	GBIF search engine provides free and open access to biodiversity data.

Digital object type: Metadata registry

Resource	GloBI search engine
URL	https://www.globalbioticinteractions.org/
Description	Global Biotic Interactions (GloBI) provides open access to finding species

'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.

	interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing open datasets using open source software.

Declaration F4 Datasets: Services used to publish datasets

Digital object type: Metadata registry

Resource	GBIF search engine
URL	https://www.gbif.org/
Description	GBIF search engine provides free and open access to biodiversity data.

Digital object type: Metadata registry

Resource	GloBI search engine
URL	https://www.globalbioticinteractions.org/
Description	Global Biotic Interactions (GloBI) provides open access to finding species interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing open datasets using open source software.

7.2. Declarations for Accessibility

Declaration A1.1 Metadata: Standardised communication protocol used for metadata records

Digital object type: Data communication protocol

Name	Hypertext Transfer Protocol Secure (HTTPS)
Description	Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The protocol is therefore also referred to as HTTP over TLS, or HTTP over SSL.

'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.

Digital object type: Data communication protocol

Name	Representational state transfer (REST)
Description	REST defines a set of constraints for how the architecture of an Internet-scale distributed hypermedia system, such as the Web, should behave.

Declaration A1.1 Datasets: Standardised communication protocol for datasets

The same as for metadata records.

Declaration A1.2 Metadata: Authentication and authorization services used for metadata records access

Digital object type: Authorization protocol

Name	Open Authorization (OAuth)
Description	OAuth 2.0 is the industry-standard protocol for authorization. OAuth 2.0 focuses on client developer simplicity while providing specific authorization flows for web applications, desktop applications, mobile phones, and living room devices.

Concept: Authorization protocol

Name	Open Data
Description	Practice of sharing data publicly and reusably

Declaration A1.2 Datasets: Authentication and authorization services used for datasets

The same as for metadata records.

‘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.

Declaration A2: Metadata preservation policies adopted

No implementation choice has been made by this community.

7.3. Declarations for Interoperability

Declaration I1 Metadata: Knowledge representation languages (allowing machine interoperation) used for metadata records

Digital object type: **Knowledge representation language**

Name	Darwin Core Archive (DwC-A)
Description	"DwC-A is a biodiversity informatics data standard that makes use of the Darwin Core terms to produce a single, self contained dataset for sharing species-level (taxonomic), species-occurrence data, and sampling-event data. An archive is a set of text files, in standard comma- or tab-delimited format, with a simple descriptor file (called meta.xml) to inform others how the files are organised." (GBIF 2021).

Digital object type: **Knowledge representation language**

Name	JavaScript Object Notation (JSON)
Description	"JavaScript Object Notation (JSON) is a lightweight, text-based, language-independent data interchange format. It was derived from the ECMAScript Programming Language Standard. JSON defines a small set of formatting rules for the portable representation of structured data. This RFC specification aims to remove inconsistencies with other specifications of JSON, repair specification errors, and offer experience-based interoperability guidance." (Bray 2017).

The following knowledge representation languages have not been implemented by this project, but are recommended.

Digital object type: **Knowledge representation language**

Name	eXtensible Markup Language Schema (XMLS)
Description	"XMLS defines and describes a class of XML documents by using schema components to constrain and document the meaning, usage and relationships of their constituent parts: datatypes, elements and their content and attributes and their values." (W3C 2004).

'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.

Digital object type: Knowledge representation language

Name	Resource Description Framework Schema (RDFS)
Description	“RDF Schema (RDFS) is the RDF vocabulary description language. RDFS defines classes and properties that may be used to describe classes, properties and other resources.” (W3C, 2014).

Declaration I1 Datasets: Knowledge representation languages (allowing machine interoperation) used for data sets

The same as for the metadata records.

Declaration I2 Metadata: Structured vocabularies used to annotate metadata records

See Declaration F2.

Declaration I2 Datasets: Structured vocabularies used to encode datasets

The same as in Declaration F2, with the addition of Relations Ontology to specify the types of plant-pollinator interactions.

Digital object type: Ontology

Name	Relations Ontology (RO)
Namespace	http://purl.obolibrary.org/obo/ro.owl
Description	“RO is a collection of relations intended primarily for standardisation across ontologies in the OBO Foundry and wider OBO library. It incorporates ROCore upper-level relations such as part of as well as biology-specific relationship types such as ‘develops from’.” (Relations Ontology, 2024).

Declaration I3 Metadata: Semantic model used for metadata records

Darwin Core, Darwin Core Archive, and Relations Ontology.

‘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.

Declaration I3 Datasets: Semantic model used for datasets

DwC-A | Darwin Core Archive.

7.4. Declarations for Reusability

Declaration R1.1 Metadata: Licence used for your metadata records

Digital object type: Licence

Name	CC0 1.0 | CC0 1.0 Universal Public Domain Dedication
Description	You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.
URL	https://creativecommons.org/publicdomain/zero/1.0/deed.en

Digital object type: Licence

Name	CC BY 4.0 | Attribution 4.0 International
Description	Using this licence you are free to share and adapt the resource but you must give appropriate credit.
URL	https://creativecommons.org/licenses/by/4.0/deed.en

Digital object type: Licence

Name	CC BY-NC 4.0 | Attribution-NonCommercial 4.0 International
Description	This licence allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.
URL	https://creativecommons.org/licenses/by-nc/4.0/deed.en

Declaration R1.1 Datasets: Licence used for datasets

The same as in Declaration R1.1 Metadata.

‘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.

Declaration R1.2 Metadata: Metadata schema used for describing the provenance metadata records

Digital object type: Ontology

Name	PROV-O | W3C PROV Ontology
Namespace	http://www.w3.org/ns/prov#
Description	The PROV Ontology (PROV-O) expresses the PROV Data Model using the OWL2 Web Ontology Language (OWL2). It is intended for the Linked Data and Semantic Web community. It provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. It can also be specialised to create new classes and properties to model provenance information for different applications and domains. PROV-O is one serialisation of PROV-DM, the other two being PROV-N and PROV-XML. PROV-DM and PROV-O define how to represent provenance on the World Wide Web, and as such additional documentation has been included in this record for PROV-AQ (Access and Query), a note which describes how standard web protocols may be used to locate, retrieve and query provenance records. PROV-DC provides a mapping from Dublin Core to PROV-O, and is listed in this record. For the purpose of this specification, provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing. In particular, the provenance of information is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources, and how to give credit to its originators when reusing it. In an open and inclusive environment such as the Web, where users find information that is often contradictory or questionable, provenance can help those users to make trust judgements. (W3C 2013).

Declaration R1.2 Datasets: Metadata schema used for describing provenance of datasets

The same as in Declaration R1.2 Metadata.

Declaration R1.3: Your community uses this FAIR Implementation Profile to link to domain-relevant community standards

Yes.

'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. Appendix II. Full Review Reports

This appendix includes references to the review reports of participating pilot studies. Please note that the accessibility of the reviews varies: some pilots publish their reviews, metadata, and data openly, whereas others opted to keep some, or all, restricted.

Nomer, & Elton. (2024a). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Agudelo-Zapata MC, Álvarez Correa C, Bedoya Duque C, Cardona-Duque J, Idárraga M, Marentes-Herrera E & Molina JA. 2023. Dimensiones de la biodiversidad del Refugio de Vida Silvestre Alto de San Miguel. Universidad CES-Secretaría de Medio Ambiente de Medellín, 2023 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647879

Nomer, & Elton. (2024b). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Alves, Denise A.; Bento, José M. S.; Carvalheiro, Luísa G. 2023. Contribution of insect pollinators to orange production and quality. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647830

Nomer, & Elton. (2024c). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Bergamo P. 2010 Dados provenientes de estudos realizados no Brasil sobre sistema reprodutivo e polinização/polinizadores de cultivares; manejo de polinizadores; doenças e agrotóxicos em polinizadores; paisagem, mudanças climáticas e conservação no contexto da polinização; e artigos de revisão [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647758

Nomer, & Elton. (2024d). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Carvalheiro, LG; Barbosa, E.R.M. & Memmott, J. 2008. Pollinator networks, alien species and the conservation of rare plants: *Trinia glauca* as a case study. Journal of Applied Ecology, 45,1419-1427. DOI: https://doi.org/10.1111/j.1365-2664.2008.01518.x . [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647856

Nomer, & Elton. (2024e). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Christine Taliga, Diana Cox-Foster 2023. USDA NRCS PLANTS Pollinator Interaction Prototype Data [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647942

Nomer, & Elton. (2024f). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Kasina M, Kimani I, Mulwa S and Muliaro W. A review of the status of web-based African Plant-Pollinator Interaction data. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647892

Nomer, & Elton. (2024g). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Marques, Bruno Ferreira; Carvalheiro, Luísa G. 2023. Orange (*Citrus sinensis* L. Osbeck, var. Pera-rio) insect floral visiting data of orchards in Itaberaí, Goiás, Brasil [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647981

Nomer, & Elton. (2024h). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Plant-flower visitor interactions recorded in 49 sites in Argentina (Buenos Aires: Carlos Casares county) by

'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.

Marcos Monasterolo (2013-15) and Antonio López Carretero (2016). [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10648048

Nomer, & Elton. (2024i). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Max Rünzel, Drew Robinson. 2023. HiveTracks WorldFAIR Test Data. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10694980

Nomer, & Elton. (2024j). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Tinoco, C.F. 2023. Floral visitation in restored areas/remnants of natural vegetation in the Xingu region. Zenodo. https://doi.org/10.5281/zenodo.10695013

Nomer, & Elton. (2024k). Review of WorldFAIR Agricultural Plant-Pollinator Data Pilot: Varassin, I.G., de Souza, T.M. 2023. The Ecology of Plant Hummingbird Interactions (EPHI) - Brazil [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10647779

'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.

9. Appendix III. Annotated data sets

Alves, D. A. (2024). Contribution of insect pollinators to orange production and quality [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679727

Bedoya Duque, C., Correa Álvarez, C. M., Cardona Duque, J., Molina, A., Juan Fernando, A., Vélez-Naranjo, M. C., Marentes Herrera, E., Agudelo Zapata, M. C., & Idárraga Giraldo, M. C. (2024). Web interactions between insects and some common plants in the "Refugio de Vida Silvestre Alto de San Miguel" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10678237

Carvalheiro, L. G. (2024). Plant-flower visitor network from Avon Gorge, UK [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679322

Ferreira Marques, B., & Carvalheiro, L. G. (2024). Orange (Citrus sinensis L. Osbeck, var. Pera-rio) insect floral visiting data of orchards in Itaberaí, Goiás, Brasil [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679666

González-Vaquero, R. A., & Devoto, M. (2024). Plant-flower visitor interactions recorded in Argentina (Buenos Aires: Carlos Casares county) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10669877

Graham, C., Varassin, I., & Machado-de-Souza, T. (2024). The Ecology of Plant Hummingbird Interactions (EPHI) - Brazil [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679802

Kasina, M., Kimani, I., Mulwa, S., Wafula Muliaro, J., & Kenyatta, J. (2024). A review of the status of web-based African Plant-Pollinator Interaction data [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679258

Rünzel, M., & Robinson, D. (2024). HiveTracks WorldFAIR Test Data [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10692151

Tinoco, C. (2024). Floral visitation in restored areas/remnants of natural vegetation in the Xingu region [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679395

Wolowski, M., Agostini, K., Freitas, L., Bergamo, P., & Salim, J. A. (2024). Data compiled from published (original or review) studies carried out in Brazil on the reproductive system and pollination/pollinators of crop plants. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10691993

'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.

10. Appendix IV. GloBI contribution guidelines

(As accessed on 2024-02-22 at https://globalbioticinteractions.org/contribute.)

You can contribute to Global Biotic Interactions in many different ways: use it, join a discussion, share data, contribute code or donate.

Use it

By using GloBI, you support it's mission: to help make species interaction datasets more accessible.

Join a Discussion

Ask questions, share ideas and stay informed by reviewing, commenting on, or opening, a github issue.

Share Data

Various methods exist to share existing interaction data through Global Biotic Interactions:

1. Take a picture, upload it to iNaturalist.org and identify the interacting species using observation fields. For example see Scott Loarie. 2013. *Haemorhous mexicanus* eating *Heteromeles arbutifolia*. iNaturalist.org. Accessed at https://inaturalist.org/observations/432688 on 28 Aug 2014. With a few exceptions, most research-grade iNaturalist interaction data is automatically included in GloBI. For a full list of iNaturalist observation fields GloBI imports, see the iNaturalist to GloBI interaction map . For specific instructions, please see Ken's how-to-add-an-inaturalist-interaction document

2. Have a look at the Dataset Management page, create a discoverable GitHub repository (or use our template dataset repository to get started). Your GloBI-compatible data repository will now be automatically included in GloBI and should appear on the status page within a couple of days. Use GitHub<>Zenodo integration to make your data citable (see next step for more info). GloBI supports many existing interaction data formats, including but not limited to IPT RSS feeds (e.g., Symbiota Collections of Arthropods Network (SCAN)) and DarwinCore Archives (e.g., Illinois Natural History Survey Insect Collection). See all registered datasets and the blog post "Models in Fashion" for more examples.

3. Publish your dataset on Zenodo and add your publication to Zenodo's Global Biotic Interaction community. For more information, see an example data publication.

4. Open an issue and provide a (permanent) url to a web-accessible existing interaction dataset along with a data citation. Any structured data format / API will do, and csv/tsv file formats are preferred. Examples include references to openly accessible datapaper (e.g. Raymond et al. 2011, Ferrer-Paris et al. 2014), data hosted in Github (e.g. Hurlbert 2014) or publicly accessible APIs (e.g. iNaturalist). For citations, DOIs are preferred, but any will do as long as they describe the source of

'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 data. If you don’t have one already, services like figshare, dryad and Zenodo allow you to get one.

5. In case you are publishing a (data) paper that contains species interaction data, cite Poelen et al. 2014. The GloBI citation helps us to easily find your paper and make the published data easy to access.

6. In case the data is not (yet) web accessible, please open an issue in which you describe the dataset, and we can have a discussion on how to make the data available through GloBI.

As the automated updates occur on a daily basis, it may take a day or two for updates or corrections to be available through GloBI and related libraries like rglobi. For more information, see GloBI’s Data Integration Process page or Poelen et al. 2014.

Contribute Code

Improve GloBI by contributing to rglobi, elton, this website or data crunching libraries.

Donate

We would appreciate if you would consider donating time, funds, server space and/or data storage to help make GloBI more useful and resilient? Please contact the main author of the 2014 GloBI paper for details.

‘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.

11. Appendix V. GloBI integration and review process

(As accessed on 2024-02-22 at https://globalbioticinteractions.org/process.)

Data Integration Process

To enable the discovery of existing species interaction datasets, Global Biotic Interactions (GloBI) continuously tracks existing datasets and integrates the discovered interaction records. These integrated interaction records form the basis of the GloBI's interpreted interaction data.

The process described in Figure 4 is an evolution of the process described in the original GloBI methods paper (Poelen et al., 2014).

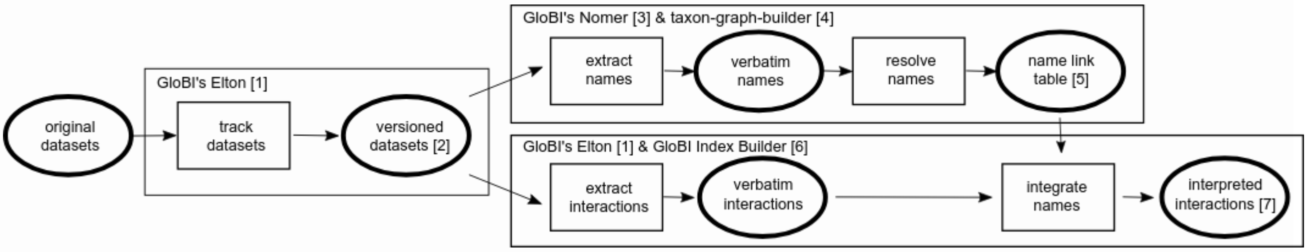


Figure 4. The integration process consists of the following phases:

a. `track` Every other day, Elton ([1]) is used to execute `elton track`. This command queries GitHub and Zenodo for species interactions datasets registration entries (e.g., NHM Interactions Bank, Soleto-Casas & Simões 2020). The information in these entries allow Elton (or any other tool) to locate resources that contain interaction datasets. After discovering dataset locations, all resources related to these datasets are downloaded, versioned and stored. Once in a while, a collection volume of these versioned datasets is added to the Elton Dataset Cache ([2]).

b. `resolve` Regularly, as part of the automated taxon-graph-builder program, Elton command `elton names` is used to extract all names from the versioned datasets. Now, Nomer is used to associate verbatim names to (taxonomic) names known to existing name authorities (e.g., ITIS, WoRMS) using existing services (e.g., https://resolver.globalnames.org, ITIS data products, NCBI api) (e.g., ITIS, WoRMS). The resulting name links are recorded in a name link table. Once updated, the name link table is published as part of the GloBI Taxon Graph ([5]).

c. `integrate` Every other day, GloBI's Index Builder takes the most recent versions of species interaction datasets, extracts the interactions and integrates the resulting records using, among other things, a published version of the name link table ([7]). The resulting integrated, or interpreted, species interaction data forms the basis of the GloBI's interpreted interaction data.

Bias and Errors

As with any analysis and processing workflow, care should be taken to understand the bias and error propagation of data sources and related data transformation processes. The datasets indexed

'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.

by GloBI are biased geospatially, temporally and taxonomically (Hortal et al. 2015 doi:10.1146/annurev-ecolsys-112414-054400, Cains et al. 2017 doi:10.5281/zenodo.814978). Also, mapping of verbatim names from datasets to known name concepts may contain errors due to synonym mismatches, outdated names lists, typos or conflicting name authorities. Finally, bugs may introduce bias and errors in the resulting integrated data product.

To help better understand where bias and errors are introduced, only versioned data and code are used as an input: the datasets, name maps and integration software are versioned so that the integration processes can be reproduced if needed. This way, steps taken to compile an integrated data record can be traced and the sources of bias and errors can be more easily found.

Customization

The modular GloBI integration workflow is designed to facilitate maintenance, troubleshooting, scaling, and stability of the integration process. This means that, in theory, specifically curated name maps and source datasets can be used to compile an integrated data product to answer a specific research question. For instance, when studying bats and the viruses that they host, only relevant input datasets and up-to-date name maps can be curated and constructed. And, a name map can be constructed manually instead of using the Taxon Graph Builder. Alternatively, the verbatim interaction can be extracted from selected datasets using `elton interactions` and other tools can be used to resolve names or otherwise enrich/process the verbatim interactions.

Notes

[1] Elton is a command-line tool to help track, version and access species interaction data. See https://github.com/globalbioticinteractions/elton and doi:10.5281/zenodo.998263.

[2] Versioned datasets, or GloBI's Elton Dataset Cache, contains versioned species interactions datasets and are the result of the `elton track` command. See also doi:10.5281/zenodo.2007418 .

[3] Nomer is a command-line tool to help map identifiers and names to taxonomic names and ontological terms. See https://github.com/globalbioticinteractions/nomer/ and doi:10.5281/zenodo.1145474 .

[4] Taxon Graph Builder is a script using standard linux tools to map (new) names in versioned interaction datasets to known name concepts. The process produces a version of a GloBI Taxon Graph.

[5] Name link table, or GloBI Taxon Graph, is the (published) outcome of the Taxon Graph Builder. The link table associated verbatim names to known taxon name concepts. See published versions at doi:10.5281/zenodo.755513 .

[6] GloBI's Index Builder is an automated process that integrates versioned datasets and a published name map (e.g., GloBI's Taxon Graph) to create integrated species interaction data

'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.

products at https://globalbioticinteractions.org/data . See also https://github.com/globalbioticinteractions/globalbioticinteractions/ .

[7] Interpreted, or integrated, interactions are one of the outcomes of the described GloBI processes. Also see doi:10.5281/zenodo.3950589 and the data page .

'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.

12. Bibliography

Alves, D. A. (2024). Contribution of insect pollinators to orange production and quality [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679727

Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool. Zenodo. https://doi.org/10.5281/zenodo.6361400 r

Bray, T. (Ed.). (2017). The javascript object notation (Json) data interchange format (RFC8259; p. RFC8259). RFC Editor. https://doi.org/10.17487/RFC8259

Carvalheiro, L. G. (2024). Plant-flower visitor network from Avon Gorge, UK [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679322

Carvalheiro, L. G. 2023. Plant-flower visitor network from Avon Gorge, UK. https://github.com/globalbioticinteractions/carvalheiro2023

Carvalheiro, LG; Barbosa, E.R.M. & Memmott, J. 2008. Pollinator networks, alien species and the conservation of rare plants: Trinia glauca as a case study. Journal of Applied Ecology, 45,1419-1427. DOI: https://doi.org/10.1111/j.1365-2664.2008.01518.x .

CDIF Working Group, Richard, S., Gregory, A., Hodson, S., Fils, D., Kanjala, C., Bell, D., Winstanley, P., Edwards, M., Heus, P., Brickley, D., Rizzolo, F., Maxwell, L., Luis, G., Buttigieg, P. L., & Le Franc, Y. (2023). Cross Domain Interoperability Framework (CDIF): Discovery Module (v01 draft for public consultation) (Version 01). Zenodo. https://doi.org/10.5281/zenodo.10252564

Drucker, Debora. (2022). WorldFAIR WP10 Plant-Pollinator FIP01. FIP Wizard. URL: https://fip-wizard.ds-wizard.org/wizard/projects/9542ef0d-66d3-4851-b53a-3eaf5ece921c (retrieved on 02/08/2024).

Drucker, Debora et al. (2024). D10.2 Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations. (Version 01). Zenodo. https://zenodo.org/doi/10.5281/zenodo.10666593

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

Elliott M.J., Poelen J.H., Fortes J.A.B. (2020). Toward Reliable Biodiversity Dataset References. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2020.101132 hash://sha256/136c3c1808bcf463bb04b11622bb2e7b5fba28f5be1fc258c5ea55b3b84f482c

FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). Zenodo. https://doi.org/10.15497/rda00050

FAIRsharing Team. (2015). Fairsharing record for: Ecological metadata language [dataset]. FAIRsharing. https://doi.org/10.25504/FAIRSHARING.R3VTVX

GBIF (2021) Darwin Core Archives – How-to Guide, version 2.2. Copenhagen: GBIF Secretariat. https://ipt.gbif.org/manual/en/ipt/3.0/dwca-guide

Global Biotic Interactions (GloBI). What is GloBI? URL: https://www.globalbioticinteractions.org/about (retrieved on 02/02/2024).

GOFAIR. FAIR Principles. URL: https://www.go-fair.org/fair-principles/ (retrieved on 08/02/2024).

'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.

González-Vaquero, R. A., Devoto, M. (2023). Plant-flower visitor interactions recorded in 49 sites in Argentina (Buenos Aires: Carlos Casares county) by Marcos Monasterolo (2013-15) and Antonio López Carretero (2016). https://github.com/globalbioticinteractions/gonzalez-vaquero2023

Graham, C., Varassin, I., & Machado-de-Souza, T. (2024). The Ecology of Plant Hummingbird Interactions (EPHI) - Brazil [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679802

Jorrit H. Poelen, James D. Simons and Chris J. Mungall. (2014). Global Biotic Interactions: An open infrastructure to share and analyse species-interaction datasets. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Kasina, M. (2024). A review of the status of web-based African Plant-Pollinator Interaction data [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10679258

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)

Martín González, A.M., Dalsgaard, B., Nogués-Bravo, D., Graham, C.H., Schleuning, M., Maruyama, P.K., Abrahamczyk, S., Alarcón, R., Araujo, A.C., Araújo, F.P., de Azevedo, S.M., Jr, Baquero, A.C., Cotton, P.A., Ingversen, T.T., Kohler, G., Lara, C., Las-Casas, F.M.G., Machado, A.O., Machado, C.G., Maglianesi, M.A., McGuire, J.A., Moura, A.C., Oliveira, G.M., Oliveira, P.E., Ornelas, J.F., Rodrigues, L.d.C., Rosero-Lasprilla, L., Rui, A.M., Sazima, M., Timmermann, A., Varassin, I.G., Vizentin-Bugoni, J., Wang, Z., Watts, S., Rahbek, C. and Martinez, N.D. (2015), Macroecology of hummingbird–plant networks. Global Ecology and Biogeography, 24: 1212-1224. https://doi.org/10.1111/geb.12355

Philipson, J. (2018). About a BUOI: Joint Custody of Persistent Universally Unique Identifiers on the Web, or, Making PIDs More FAIR. In: González-Beltrán, A., Osborne, F., Peroni, S., Vahdati, S. (eds) Semantics, Analytics, Visualization . SAVE-SD SAVE-SD 2017 2018. Lecture Notes in Computer Science(), vol 10959. Springer, Cham. https://doi.org/10.1007/978-3-030-01379-0_3

Poelen, J. H. (2022). Nomer Corpus of Taxonomic Resources hash://sha256/f4e2b9806440d0605f60b81feb9782655291aac2d000c74e4e8fdeb937e29b1d (0.6) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7065661

Poelen, Jorrit H., James D. Simons and Chris J. Mungall. (2014). Global Biotic Interactions: An open infrastructure to share and analyse species-interaction datasets. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Relations Ontology (RO). http://purl.obolibrary.org/obo/ro.owl (retrieved on 02/10/2024).

Salim, J. A., Saraiva, A. M., Zermoglio, P. F., Agostini, K., Wolowski, M., Drucker, D. P., Soares, F. M., Bergamo, P. J., Varassin, I. G., Freitas, L., Maués, M. M., Rech, A. R., Veiga, A. K., Acosta, A. L., Araujo, A. C., Nogueira, A., Blochtein, B., Freitas, B. M., Albertini, B. C., … Brito, V. L. G. (2022). Data standardisation of plant–pollinator interactions. GigaScience, 11, giac043.

'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.

https://doi.org/10.1093/gigascience/giac043

Schultes, E., Magagna, B., Hettne, K.M., Pergl, R., Suchánek, M., Kuhn, T. (2020). Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann, G., Ram, S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science(), vol 12584. Springer, Cham. https://doi.org/10.1007/978-3-030-65847-2_13

Thébault, E., & Fontaine, C. (2022). A database of plant-pollinator networks (Version 2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6630184

Trekels, M., Pignatari Drucker, D., Salim, J. A., Ollerton, J., Poelen, J., Miranda Soares, F., Rünzel, M., Kasina, M., Groom, Q., & Devoto, M. (2023). WorldFAIR Project (D10.1) Agriculture-related pollinator data standards use cases report. Zenodo. https://doi.org/10.5281/zenodo.8176978.

W3C. Prov-o: The prov ontology. (2013). https://www.w3.org/TR/prov-o/ (retrieved on 02/10/2024).

W3C. RDF Schema 1.1. (2014). https://www.w3.org/TR/rdf-schema/ (retrieved on 02/10/2024).

W3C. Xml schema part 1: Structures second edition. (2004). https://www.w3.org/TR/xmlschema-1/ (retrieved on 02/10/2024).

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (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.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wolowski, M., Agostini, K., Freitas, L., Bergamo, P., & Salim, J. A. (2024). Data compiled from published (original or review) studies carried out in Brazil on the reproductive system and pollination/pollinators of crop plants. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10691993

Zenodo. About records. URL: https://help.zenodo.org/docs/deposit/about-records/ (retrieved on 02/02/2024).

'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.