Notes on Methodology

CURATION AND STANDARDISATION OF PLANT–POLLINATOR INTERACTION DATA UNDER FAIR PRINCIPLES: EXPERIENCE FROM PAMPEAN AGROECOSYSTEMS

Rocío Ana González-Vaquero[1,3,*], Debora Pignatari Drucker[2], Mariano Devoto[1,3]

[1] *Universidad de Buenos Aires, Facultad de Agronomía, Cátedra de Botánica General, Av. San Martín 4453, C1417DSE. Ciudad Autónoma de Buenos Aires, Argentina.*

[2] *Embrapa Agricultura Digital, Brazilian Agricultural Research Corporation. Av. André Tosello, nº 209, Campus da Unicamp, Barão Geraldo, 13083-886 - Campinas - SP, Brazil.*

[3] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.*

Journal of Pollination Ecology,
41(1), 2025, pp 1-9

DOI: 10.26786/1920-7603(2026)873

Received 26 May 2025,
accepted 15 December 2025

*Corresponding author:
rvaquero@agro.uba.ar

Abstract—Plant–pollinator interactions are essential to the functioning of natural and agricultural ecosystems. Although the volume of available data on these interactions is steadily increasing, their utility for integrative studies is limited by heterogeneity in data formats and poor documentation. The WorldFAIR project aimed to improve data interoperability and promote the adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) principles. In this study, we evaluated the effectiveness of two standardisation approaches (manual and semi-automated) applied to five datasets with varying levels of cleanliness, complexity, and structure. Our results show that for small datasets (fewer than 637 records), the manual method is more suitable, while for larger datasets, the semi-automated approach—based on tools such as OpenRefine—is more efficient after an initial learning curve. Additionally, we compiled a list of common issues encountered during the standardisation process and suggested possible solutions. This study aims to guide those interested in applying FAIR principles to ecological interaction data, supporting the planning and decision-making process when choosing the most suitable approach.

Keywords—Pollination, bioinformatics, Darwin Core, REBIPP, WorldFAIR project, OpenRefine

INTRODUCTION

In recent years, the exponential growth in data generation and availability has transformed the way biological systems are studied, from ecological interactions to global biodiversity patterns (Kitchin 2014). However, this increased data availability also presents significant challenges related to heterogeneity, accessibility, and reusability. In biological research, the diversity of data formats and the lack of common standards hinder the integration of essential information needed to address complex issues such as biodiversity loss or climate change. To overcome these challenges, global initiatives have promoted principles such as FAIR (Findable, Accessible, Interoperable, Reusable), aimed at maximising the utility of scientific data (Wilkinson et al. 2016).

Plant–pollinator interactions play a key role in the functioning of both natural and agricultural ecosystems, being fundamental to food production and biodiversity conservation alike. An estimated 75% of major food crops rely at least partially on animal pollination, highlighting the importance of these interactions for food security and global economic well-being (Klein et al. 2007; Trekels et al. 2023). The ongoing loss of pollinators could have significant impacts on food production and ecosystem health, threatening not only biodiversity but also human nutrition (Potts et al. 2016). This concern has led to growing interest in pollination ecology, as evidenced by the substantial and sustained increase in scientific publications on the topic over the past 25 years (Ollerton 2021).

Despite the increasing volume of available data on plant–pollinator interactions, much of it remains scattered, presented in diverse formats, and poorly documented. Most studies do not define the interaction per se (i.e., when the specimen physically contacts the sexual parts of the flower), nor the time of day the interaction was recorded, nor if the insect was just observed or captured and then deposited at an entomological collection. This heterogeneity limits the potential for integrative studies at regional and global scales, hampering the advancement of knowledge about the biology and ecology of these interactions (Drucker et al. 2024a; Ollerton et al. 2025). To address this issue, the aforementioned FAIR principles offer a comprehensive framework for improving data management and reusability across scientific disciplines (Wilkinson et al. 2016). Nevertheless, the uptake of these principles within plant–pollinator research has been uneven, with obstacles such as inconsistent metadata quality and insufficient data access (i.e., data not openly available for reuse) still prevailing (Drucker et al. 2024a).

The WorldFAIR project (https://worldfair-project.eu/) aimed to enhance data interoperability and promote the implementation of FAIR principles across a wide range of scientific disciplines and communities worldwide. The "Agricultural Biodiversity" Case Study (https://worldfair-project.eu/agricultural-biodiversity) focuses specifically on plant–pollinator interactions. As part of this effort, best practice guidelines, a standardised vocabulary, and tailored FAIR assessment tools have been developed, along with estimates of the adoption costs of these standards, measured in terms of time spent on relevant tasks (Drucker et al. 2024a). This includes data standardisation based on Darwin Core terms, a widely used biodiversity framework for describing and sharing data about organisms, including their names, geographic locations, and ecological relationships (https://dwc.tdwg.org; Wieczorek et al. 2012). Additionally, it incorporates a specialised vocabulary proposed by the Brazilian Network of Plant–Pollinator Interactions (REBIPP, https://www.rebipp.org.br), a collaborative network of experts in Pollination Biology that aims to encourage the development of scientific and educational activities in the area. These tools enable coherent data structuring and facilitate their integration into global platforms such as GBIF (https://www.gbif.org/new-data-model) and GloBI (https://www.globalbioticinteractions.org; Poelen et al. 2014) (Salim et al. 2022). Pilot studies conducted within WorldFAIR have demonstrated the applicability of these standards, generating structured, machine-readable data suitable for both local and global analysis (Drucker et al. 2024a,b).

Standardising plant–pollinator interaction data can be approached through manual or semi-automated methods, each with its own advantages and limitations. Manual standardisation—typically conducted using spreadsheet tools—allows for fine-grained control but can be time-consuming and prone to human error, especially with large datasets (Drucker et al. 2024a). By contrast, semi-automated approaches, often implemented using software such as OpenRefine (https://openrefine.org), enable quicker and more consistent data transformation, although they require greater technical knowledge and an initial setup. The choice between these approaches depends on several factors, including dataset size and complexity, user expertise, availability of technical resources, and the time allocated for data cleaning and standardisation (Drucker et al. 2024a).

This study evaluates the effectiveness of standardisation methods applied to plant–pollinator interaction datasets, using the tools and standards developed under the WorldFAIR project. We compare manual and semi-automated approaches, highlighting their respective strengths and weaknesses in improving data quality and alignment with the FAIR principles (Drucker et al. 2024a). Our aim is to support researchers in selecting the most appropriate method for their needs and to promote broader standardisation and reuse of ecological interaction data, contributing to a deeper understanding of these processes essential for sustainability and biodiversity conservation.

DATASET DESCRIPTION

Plant–pollinator interaction data are typically organised in spreadsheets, where each row represents a recorded interaction between a plant species and a floral visitor, along with details such as visitation frequency, sampling location, and date. These data are critical for understanding pollination ecology and assessing the roles played by different species in both agricultural and natural ecosystems (Kaiser-Bunbury & Blüthgen 2015; Jordano 2016; Ollerton et al. 2025).

For this study, we used five datasets, each containing between 500 and 2000 rows. Each row represented a specific interaction between a plant individual and a floral visitor species (insect), while the number of individuals visiting the flowers—a common proxy for interaction strength or frequency—was recorded in a column labelled "frequency". These records formed a total of 120 plant–pollinator interaction networks (a representation useful for studying community structure), sampled across different sites within the Pampean region (Argentina), including field margins and crops of soybeans, potatoes, and sunflowers. Field margins are vegetated strips along the boundaries of crop fields, which provide ecosystem services (e.g., pollination, pest control) to crops.

The selected datasets differed markedly in terms of data "cleanliness" (the more typographical mistakes a dataset has, the less "clean" it is), the amount of details provided, and the formatting used, allowing for an evaluation of standardisation methods under diverse conditions. This variation partly reflected the origin of the data: some datasets came from ongoing doctoral theses, while others had already been processed and were associated with published studies. Table 1 summarises the main characteristics of the five datasets included in the study.

Table 1. Summary of processed datasets. The number of plant–pollinator interaction networks corresponds to the number of sampling sites. The number of records refers to the number of observed plant–pollinator interactions (spreadsheet rows). Manual = Dataset processed using the manual method. Semi-aut = Dataset processed using the semi-automated method. FM = field margin. DMS = Degrees, minutes and seconds. DD = Decimal degrees. Information regarding the authors of the names of the species, and the family each species was assigned to, was available for some datasets, recorded in this table under "Authors/families plant/insect species". The number of lines showing data issues (typographical errors, mistakes or whitespaces) is indicated in the table with asterisks, meaning: (*) = less than 10% of lines in the dataset showing data issues, (**) = 10-30% of lines in the dataset showing data issues, and (***) = >30% of lines in the dataset showing data issues. The counties are located in Buenos Aires Province, Argentina.

	Dataset 3 (manual/semi-aut)	Dataset 1 (semi-aut)	Dataset 2 (semi-aut)	Dataset 4 (semi-aut)	Dataset 5 (semi-aut)
# Plant–pollinator interaction networks	20	20	29	35	16
# Records	522	1953	518	1978	781
Publication	Tavella et al. 2022	Monasterolo et al. 2020	Unpublished	Doctoral thesis in progress (Juan Peña)	Doctoral thesis in progress (Ignacio Marchi)
Sampling	FM + soybean	FM	FM + soybean	FM + potato	FM + sunflower
Year	2018	2013-2015	2016	2020-2023	2022-2023
County	Carlos Casares	Carlos Casares	Carlos Casares	Balcarce	Balcarce
Authors plant species	NO	NO	NO	YES	YES
Authors insect species	NO	NO	NO	YES	NO
Families plant species	NO	NO	NO	YES	YES
Families insect species	YES	YES	YES	YES	YES
Empty rows/one part of interaction missing	NO	NO	NO	NO	YES
Coordinates (system)	YES (DMS)	YES (DMS)	NO	YES (DD)	YES (DMS)
Elevation	YES	YES	NO	YES	NO
Data issues	*	*	*	**	***

DATA STANDARDISATION METHODS: MANUAL AND SEMI-AUTOMATED APPROACHES

In this study, we compared two approaches to cleaning and standardising plant–pollinator interaction data: a manual method and a semi-automated method, each involving specific procedures and tools to transform datasets in accordance with the Darwin Core standard and the REBIPP vocabulary (Drucker et al. 2024a). Standardisation entails structuring data in a consistent and accurate manner to ensure its integration into international repositories and their reusability by other researchers. For biological interaction data, this includes the normalisation of scientific names, geographic coordinates, dates, and descriptions of ecological interactions such as the species involved, the interaction type, the role of each organism, and the frequency of interaction. Standards like Darwin Core provide a common framework for describing a wide range of fields, from taxonomic information to location- and time-related metadata (Wieczorek et al. 2012). The REBIPP vocabulary, meanwhile, introduces specialised terms for the precise description of plant–pollinator interactions, enabling the capture of detailed information about floral morphology and pollinator behaviour. These processes not only ensure data quality and consistency, but also foster global interoperability, facilitating integrative and comparative studies.

MANUAL METHOD

The manual method was implemented using Microsoft Excel, which allows for direct data manipulation via filtering, mass search and replace operations. The process followed a structured protocol comprising the following steps:

- **Column mapping:** Each column in the original spreadsheet was compared with the terms from Darwin Core (https://dwc.tdwg.org/terms) and REBIPP (https://ppi.rebipp.org.br/terms), and renamed accordingly.
- **Error detection and correction:** Filters were applied to each column to identify and correct errors such as misspellings, inconsistent records, or missing data. See more details about the correction process in Table 3.
- **Addition of extra information:** New columns were added where necessary to include relevant additional information, ensuring that column names conformed to Darwin Core and REBIPP terms.
- **Standardisation of dates and coordinates:** The Canadensys tool (https://data.canadensys.net/tools) was used to convert dates and coordinates to standard formats.
- **Taxonomic validation:** A list of unique scientific names was generated and validated using the Global Names Resolver (https://resolver.globalnames.org), with the GBIF Backbone Taxonomy selected as thereference taxonomy (GBIF Secretariat 2023). The validity of names, higher taxonomy (family and order), and species authorship were verified and compared with the original spreadsheet. In cases of inconsistencies or unmatched names (match: 0), local catalogues were consulted, or expert taxonomists were contacted. Synonyms were replaced with the corresponding accepted names.

SEMI-AUTOMATED METHOD

The semi-automated method was implemented using the OpenRefine software (https://openrefine.org). OpenRefine allows saving of workflows and applying steps performed on one dataset to another in the future. These steps can be useful to track the changes made to a given dataset, and can be included in a report if needed. In this case, the workflow included the following steps:

- **Column mapping:** As in the manual method, columns were compared with Darwin Core and REBIPP terms and renamed accordingly.
- **Error detection and correction:** OpenRefine enabled the use of facets, filters, and clustering functions to detect and consolidate alternative values referring to the same concept, thereby reducing redundancy and improving dataset consistency.
- **Addition of extra information:** Additional columns were created when needed to include relevant information, ensuring compliance with Darwin Core and REBIPP terminology.

- **Standardisation of dates and coordinates:** Standardised data for dates and coordinates were retrieved automatically from Canadensys. This process was facilitated by connecting OpenRefine to external APIs (application programming interfaces), which allowed for automatic integration of results into the corresponding dataset columns.
- **Taxonomic validation:** Scientific names were validated using a routine (see González-Vaquero et al. 2024) that retrieved data directly from the GBIF Backbone Taxonomy (GBIF Secretariat 2023) via its API. This process not only validated the names but also automatically added higher taxonomy and species authorship in new columns appended to the dataset.

METHOD COMPARISON

To evaluate the manual method, we used dataset 3, which was also processed using the semi-automated method for comparison. Dataset 3 was chosen because its data had been processed for publication, so the spreadsheet was considerably "clean", and it contained a manageable number of records. The experience gained from applying the semi-automated method was documented in a tutorial for OpenRefine (González-Vaquero et al. 2024). This tutorial was later used to process datasets 4 and 5, which originated from doctoral projects and included more complex information and error types. Finally, datasets 1 and 2 (similar in structure and complexity to dataset 3) were also processed following the steps of the tutorial. We were not able to process datasets 1, 2, 4, and 5 using the manual method. Since we are only comparing the two methods on a single dataset (dataset 3), we recognize that there are limitations to the conclusions we can make.

PROCESSING TIME

Processing time can be categorised into fixed and variable costs. **Fixed costs** refer to the time required to become familiar with the Darwin Core standard, the REBIPP plant–pollinator interaction terms, and—in the case of the semi-automated method—learning to use OpenRefine. These costs are incurred only once by the person performing the task and are independent of the dataset. **Variable costs** are those related to the time needed for data cleaning and standardisation, as well as for adding new columns with relevant information. Taxonomic validation and transferring data to the REBIPP template (a step adopted by the Case Study to share the dataset in the GloBI repository) are also considered variable costs, as they depend on the specific dataset. Processing times for each step and each dataset are detailed in Table 2 (modified from Drucker et al. 2024a).

For the processed datasets, we observed that when dealing with a small number of records (in this case, fewer than 637, based on the intersection of both slopes), the manual method is more convenient (Fig. 1A). However, for larger datasets, the time investment required to become familiar with OpenRefine is justified, and the semi-automated method proves to be more efficient.

Datasets 1 and 4 were the quickest to process (0.26 and 0.29 minutes per record, respectively). These spreadsheets were relatively "clean", the user already had some experience with data processing, and the steps from the tutorial (González-Vaquero et al. 2024) were simply followed. Both datasets had a notably high number of records (more than double those of the others), which contributes to optimising variable costs. The variable costs for dataset 5 were the highest, despite having fewer records and fewer names to validate (Table 2). This was due to the spreadsheet being "dirtier", with typographical errors and missing information (such as species authors and elevation), which slowed down processing. As a result, the time function slope was steeper ($m = 1.04$), although still much lower than that of the manual method ($m = 2.93$). The time spent on taxonomic validation was considerably higher than for the other datasets (4.20 minutes per name), since many of the name-related errors were only detected during this step.

Furthermore, the efficiency of the semi-automated method is evident even when processing challenging datasets. Dataset 5, which was the 'dirtiest' in our study, required 1.04 minutes per record. While this was the highest time cost among the datasets processed semi-automatically, it is crucial to note that it was still approximately three times faster than the manual method applied to the relatively clean Dataset 3 (2.93 minutes per record). This demonstrates that

Table 2. Time (in minutes) required for each step of the manual and semi-automated approaches. The number of names to validate excludes morphospecies identified with numbers. The number of records refers to the number of interactions (rows in a spreadsheet) processed under each method. Manual = Dataset processed using the manual method. Semi-aut = Dataset processed using the semi-automated method. NA = not applicable.

	Dataset 3 (manual)	Dataset 3 (semi-aut)	Dataset 1 (semi-aut)	Dataset 2 (semi-aut)	Dataset 4 (semi-aut)
Learn DwC	540	540	540	540	540
Learn REBIPP	270	270	270	270	270
Learn OpenRefine	NA	1380	1380	1380	1380
Fixed costs	810	2190	2190	2190	2190
Cleaning + standardisation	890	190	130	120	180
Validation	340	120	320	80	230
REBIPP template	300	90	60	50	160
Variable costs	1530	400	510	250	570
#Names for validation	67	67	200	83	158
#Records	522	522	1953	518	1978
∑Variable costs/#records	2.93	0.77	0.26	0.48	0.29
∑Minutes validation/name	5.07	1.79	1.6	0.96	1.46

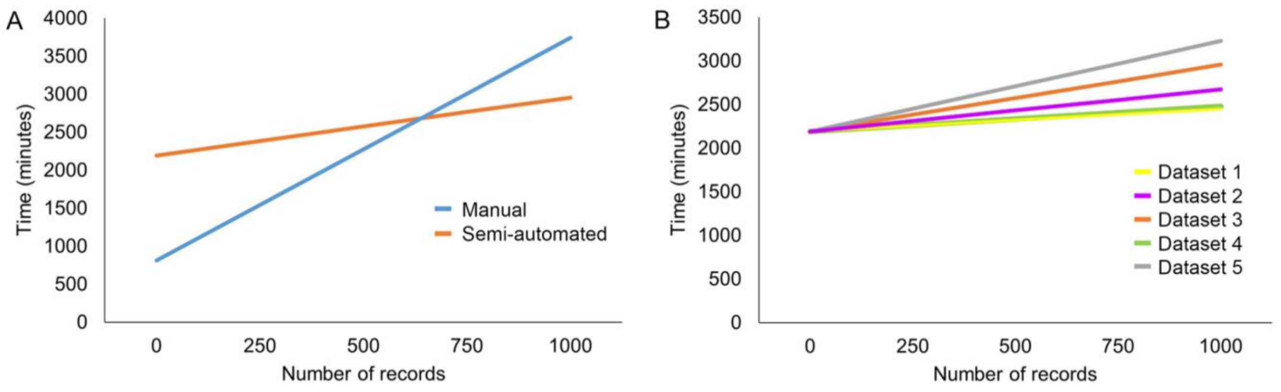

Figure 1. Time invested in processing plant–pollinator interaction datasets using different methodologies. Time functions are derived from the time values presented in Table 2, where the y-intercept represents fixed costs (in minutes), and the slope represents the cost per record. A, Time functions for the manual approach (blue) and semi-automated approach (orange), applied to the same dataset (dataset 3); the intersection of the two functions occurs at 637 records, beyond which the semi-automated method becomes more efficient; B, Time functions for the semi-automated approach applied to five different datasets: dataset 1 (yellow), dataset 2 (violet), dataset 3 (orange), dataset 4 (green), and dataset 5 (grey).

the semi-automated approach maintains a significant efficiency advantage over the manual method, even under suboptimal conditions of high data complexity and error density, once the initial learning curve is overcome.

RECOMMENDATIONS

Table 3 presents examples of the most common issues that may arise when processing a plant-pollinator interaction dataset for "FAIRification", along with suggested solutions based on the experience gained during this study.

Table 3. Common problems in the "FAIRification" process, and suggested solutions.

Problem	Suggested Solution
Rows lack a unique identifier	If spreadsheet rows are not individually identified, it is advisable to add an "identifier"—a name that is preferably unique to each specific interaction, both in the spreadsheet and across all datasets produced by the institution. Although OpenRefine assigns an internal row number, this identifier is not retained in the final output. A typical identifier includes the acronym of the collection or institution, a code referring to the location or the researcher's name, and a serial number with multiple digits. In Darwin Core, this should be placed in the catalogNumber field.
Missing name for one interaction partner	If a row lacks the name of either the plant or the animal, the record should be discarded, since the objective here is to report interactions (not occurrences).
Missing collector name(s)	If the name of the person who recorded each interaction is not available, the name of the field campaign coordinator should be entered in the recordedBy field.
Morphospecies-level identifications	As the scientificName field does not support morphospecies (e.g. codes or numbers), only the lowest confirmed taxonomic rank should be reported (e.g. genus or family). Morphospecies labels can be included in the identificationRemarks field, along with the name of the person responsible for the identification.
Unstandardised dates and/or coordinates	Use the Canadensys tools (https://data.canadensys.net/tools/dates or https://data.canadensys.net/tools/coordinates) or query them via OpenRefine to standardise the values. It is recommended to keep the original data in verbatimLatitude, verbatimLongitude, and verbatimEventDate, and to add the standardised values in decimalLatitude, decimalLongitude, and eventDate, respectively.
Suspicious dates (e.g., 24/02/1900)	Consult the dataset owner and/or review other rows to infer the correct value (e.g., if other entries show 24/02/1990, this may be a typo). If corrected, report the accurate date in eventDate and retain the original in verbatimEventDate.
Ambiguous dates (e.g., 02/06/2020)	If it's unclear whether the day and month have been inverted, and clarification is not possible, leave the date in verbatimEventDate and do not fill in eventDate. In case there is a column for year, fill it in.
Supplementary data stored in a separate file	Migrate the supplementary data using OpenRefine, following the steps outlined in the tutorial (González-Vaquero et al. 2024). Often, coordinates and elevation data are found in separate spreadsheets.
Erroneous coordinates	Check that the site coordinates fall within the sampling area, e.g. using Google Earth. Outliers should be discussed with the dataset owner.
Missing coordinates	If coordinates are missing but other information is available (e.g. locality), consider georeferencing following the guidelines in Chapman & Wieczorek (2020).
Missing elevation data	Elevation values can be retrieved from Google Earth or a dedicated website such as https://www.gpsvisualizer.com/elevation
Scientific name is a synonym	Replace with the accepted name in the scientificName field, and record the decision in identificationRemarks, including the taxonomic source or catalogue used.
Incorrect family assigned to a name	Correct the family name in the family field and document the change in identificationRemarks, citing the taxonomic reference.
Conflicting taxonomic catalogues	Choose one source and record the decision in identificationRemarks, indicating the selected catalogue or database.
Scientific name split across multiple columns	Combine columns (typically genus and specific epithet) to create the scientificName, following the steps in the tutorial (González-Vaquero et al. 2024: page 40).

CONCLUSIONS

The process of "FAIRification" of biological interaction data involves standardising and cleaning spreadsheets using either manual or semi-automated methods. The choice between these approaches is not solely determined by the number of records and the skills of the person responsible. Other critical factors include the inherent complexity and 'dirtiness' of the data, where the semi-automated method's clustering algorithms excel; the need for a reproducible and documented workflow for auditing or reuse; the frequency with which the curation task will be performed; and the requirement for complex data enrichment from external APIs. For projects requiring repeatability, handling complex data errors, or integrating with external biodiversity services, the semi-automated method offers significant long-term advantages that extend beyond raw processing speed. Figure 1A provides a helpful approximation, but it is important to note that both methods were applied only to a single dataset, and the functions may differ significantly for any other given dataset. We recommend developing a detailed and realistic plan, avoiding underestimating the time required for each step. In this study, we provide a list of common characteristics found in interaction datasets (Table 3), which we hope will be useful for such planning.

Our results indicate that the tools developed under the WorldFAIR project—such as the tutorial for applying the semi-automated standardisation method (González-Vaquero et al. 2024)—allow large datasets to be transformed into standardised formats more efficiently than the manual approach. This efficiency is robust: even our most challenging dataset (Dataset 5) was processed three times faster using the semi-automated method than a cleaner dataset (Dataset 3) was using the manual method. In addition, the potential for error introduction is higher with the manual method. Many terms automatically retrieved by a routine in OpenRefine, should be either typed or copy-pasted from the results obtained by the Global Names Resolver, and mistakes can be made in this process. While the time investment per record may initially be higher with the semi-automated method, it decreases substantially after processing the first dataset. Practice in data cleaning and standardisation, as well as familiarity with the species from the biogeographic region of origin, contributes to this reduction.

The semi-automated approach enhances data interoperability, facilitates taxonomic validation, and supports integration with extended data models (see Salim et al. 2022), contributing to more comprehensive research at both regional and global levels. In addition, these standardisation tools promote data reuse, aligning with FAIR principles and supporting open access to critical information for the management and conservation of pollinators (Drucker et al. 2024a, Ollerton et al. 2025). Although this study focused on plant–pollinator interactions, the approaches, tools, and recommendations presented here are readily applicable to other types of ecological interactions.

ACKNOWLEDGEMENTS

We are grateful to Jeff Ollerton, José Augusto Salim, Jorrit Poelen, and Paula Zermoglio for their valuable input in addressing questions that arose during various stages of this study, and to Julia Tavella, Marcos Monasterolo, Tania Palacios, Juan Peña, and Ignacio Marchi for their kind assistance with the datasets. "This work was developed as part of the "'Global cooperation on FAIR data policy and practice' (WorldFAIR), funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393, and we acknowledge the support provided by the project partners and resources".

AUTHOR CONTRIBUTION

Concept and design RAGV & MD, data analysis RAGV, writing RAGV & MD, edits and approval for publication RAGV, DPD & MD.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

DATA AVAILABILITY STATEMENT

The data used to write this article are available at https://doi.org/10.5281/zenodo.10669877.

REFERENCES

Chapman AD, Wieczorek JR (2020) Guía de buenas prácticas de georreferenciación. Secretariado de GBIF, Copenhague. https://doi.org/10.15468/doc-gg7h-s853

Drucker D, Salim JA, Poelen J, Soares FM, Gonzalez-Vaquero RA, Devoto M, Ollerton J, Kasina M, Carvalheiro LG, Bergamo PJ, Alves DA, Varassin I, Tinoco FC, Rünzel M, Robinson D, Cardona-Duque J, Idárraga M, Agudelo-Zapata MC, Marentes Herrera E, Taliga C, Parr C, Cox-Foster D, Hill E, Motta Maués M, Agostini K, Rech AR, Saraiva A (2024) WorldFAIR (D10.3) Agricultural biodiversity FAIR data assessment rubrics (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10719265

Drucker D, Salim JA, Poelen J, Soares FM, Gonzalez-Vaquero RA, Ollerton J, Devoto M, Rünzel M, Robinson D, Kasina M, Taliga C, Parr C, Cox-Foster D, Hill E, Motta Maués M, Saraiva AM, Agostini K, Carvalheiro LG, Bergamo P, Varassin I, Araujo Alves D, Marques B, Tinoco C, Rech AR, Cardona-Duque J, Idárraga M, Agudelo-Zapata MC, Marentes Herrera E, Trekels M (2024) WorldFAIR (D10.2) Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10666593

GBIF Secretariat (2023) GBIF Backbone Taxonomy. Checklist dataset [online] URL: https://doi.org/10.15468/39omei

Gonzalez-Vaquero RA, Drucker D, Salim JA, Soares FM, Zermoglio PF, Devoto M (2024) WorldFAIR (D10.2) Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations: Tutorial. Zenodo. https://doi.org/10.5281/zenodo.10688865

Jordano P (2016) Sampling networks of ecological interactions. Functional Ecology 30(12):1883–1893. https://doi.org/10.1111/1365-2435.12763

Kaiser-Bunbury CN, Blüthgen M (2015) Integrating network ecology with applied conservation: A synthesis and guide to implementation. AoB Plants 7:plv076. https://doi.org/10.1093/aobpla/plv076

Kitchin R (2014) Big Data, new epistemologies and paradigm shift. Big Data & Society 1(1): 1-12. https://doi.org/10.1177/2053951714528481

Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T (2007) Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B: Biological Sciences 274(1608):303–13. https://doi.org/10.1098/rspb.2006.3721

Monasterolo M, Poggio SL, Medan D, Devoto M (2020) Wider road verges sustain higher plant species richness and pollinator abundance in intensively managed agroecosystems. Agriculture, Ecosystems & Environment 302:107084. https://doi.org/10.1016/j.agee.2020.107084

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

Ollerton J, Taliga C, Salim JA, Poelen JH, Drucker DP (2025) Incorporating measures of data quality into plant-pollinator databases. Journal of Pollination Ecology. https://doi.org/10.26786/1920-7603(2025)844

Poelen JH, Simons JD, Mungall CJ (2014) Global Biotic Interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Potts SG, Imperatriz-Fonseca V, Ngo HT, Aizen MA, Biesmeijer JC, Breeze TD, Dicks LV, Garibaldi LA, Hill R, Settele J, Vanbergen AJ (2016) Safeguarding pollinators and their values to human well-being. Nature 2016;540(7632):220–9. https://doi.org/10.1038/nature20588

Salim JA, Saraiva AM, Zermoglio PF, Agostini K, Wolowski M, Drucker D, Soares FM, Bergamo PJ, Varassin IG, Freitas L, Maués MM, Rech AR, Veiga AK, Acosta AL, Araujo AC, Nogueira A, Blochtein B, Freitas BM, Albertini BC, Maia-Silva C, Nunes CEP, Pires CSS, dos Santos CF, Queiroz EP, Cartolano EA, de Oliveira FF, Amorim FW, Fontúrbel FE, da Silva GV, Consolaro H, Alves-dos-Santos I, Isabel C Machado IC, Silva JS, Aleixo KP, Carvalheiro LG, Rocca MA, Pinheiro M, Hrncir M, Streher NS, Ferreira PA, de Albuquerque PMC, Maruyama PK, Borges RC, Giannini TC, Brito VLG (2022) Data standardization of plant–pollinator interactions. GigaScience 11:giac043. https://doi.org/10.1093/gigascience/giac043

Tavella J, Windsor FM, Rother DC, Evans DM, Guimarães PR, Palacios TP, Lois M, Devoto M (2022) Using motifs in ecological networks to identify the role of plants in crop margins for multiple agriculture functions. Agriculture, Ecosystems & Environment 331:107912. https://doi.org/10.1016/j.agee.2022.107912

Trekels M, Pignatari Drucker D, Salim JA, 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 1). Zenodo. https://doi.org/10.5281/zenodo.8176978

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

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, 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 2016;3(1):160018. https://doi.org/10.1038/sdata.2016.18