

# Plant-pollinator interactions data standardization

Community Review



# Summary

- Why do we need a data standard?
- What is a data standard?
- The Darwin Core example
- What are Darwin Core Extensions?
- A plant-pollinator interaction DwC Extension
- What are the next steps?
- How GitHub can help us?

# Why do we need a data standard?

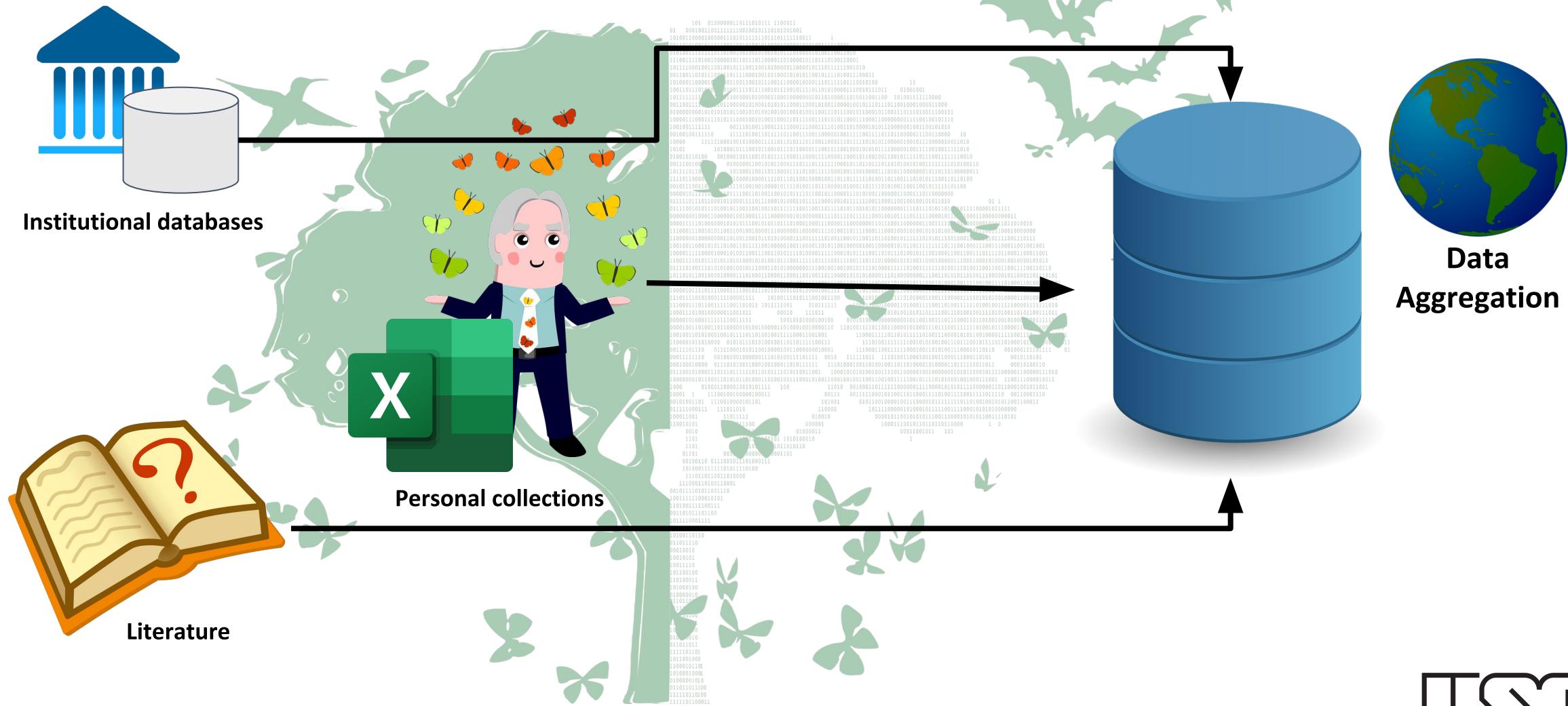


## Institutional databases



## Personal collections

# Why do we need a data standard? (cont.)



# What is a data standard?



- "Standards make it easier to **create, share, and integrate data** by making sure that there is a **clear understanding of how the data are represented** and that the data you receive are in a **form that you expected.**" ([USGS](#)).
- "The goal of a data standard is to enable the **sharing or exchange** of information between **multiple parties** in a way that guarantees that the interacting parties share the **same understanding of what is represented** within that information" (<http://www.knowledge-integrity.com/columns/dmr200401.htm>)
- "The use of **common terminology** and **common data element definitions** enables the **integration of databases**, and promotes more efficient and effective use of data by users of commonly defined data from **disparate sources.**" ([Geoscience Australia](#))

# The Darwin Core Standard

- Darwin Core is a standard for sharing data about **biodiversity** – the **occurrence** of life on earth and its **associations** with the environment.
- The last version defines 171 terms.

Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	
MeasurementOrFact	measurements, facts, characteristics, assertions, references	Generic Darwin Core (relational)

# What are DwC extensions?

- Darwin Core aims to cover the **common ground in biodiversity**, it **inevitably lacks** terms that are of interest to more **specialized** groups.
- A DwC Extension consists of **additional terms** describing a complementary, related domain, or guidance on the use of Darwin Core within a **specific sub-domain of biodiversity**.

- Registered DwC Extensions:

<https://tools.gbif.org/dwca-validator/extensions.do>

# A plant-pollinator interaction DwC extension

- DwC does not have specific terms to describe biotic interactions
- Then, we started the definition of a set of terms related to plant-pollinator interactions (almost 5 years ago):
  - We started from a list of **147 descriptors**.
  - After 4 workshops and many discussions we have refined the vocabulary to **51 terms**.
- TDWG Task Group of Biological Interactions Data:  
<https://www.tdwg.org/community/interaction/>
- <https://github.com/tdwg/interaction>
- <https://biocomp-usp.github.io/rebipp-data-standard/terms/>
- <https://github.com/BioComp-USP/rebipp-data-standard>

# What are the next steps?

- **Community Review: (we are here!)**
  - Review terms definitions and their relationships.
  - Gathering examples of datasets and terms usage.
- **Formalization: (and here)**
  - Finalize the data model to share plant-pollinator interaction data (depends on terms definition).
- **Publishing:**
  - Make it public and encourage the adoption to a broader community.

# How GitHub can help us?

- GitHub provides hosting for software development and version control using Git.
- But today its usage goes beyond software development, and it has been successfully used as a platform for tracking data standards evolution.
- **What is interesting to us?**
  - **Issues Tracking System:** handle discussions for each term separately.
  - **Version control:** history of data standard versions of each term.
  - **It is free!**

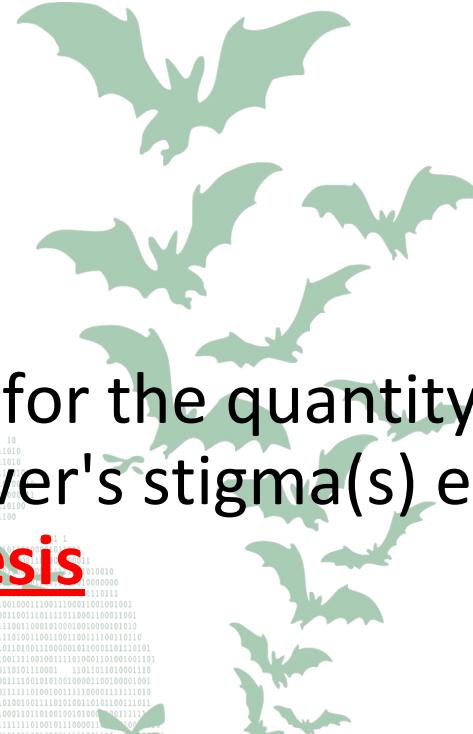
# Just an example

**Term:** co-specificPollenGrainsQuantity

**Definition:** The number or enumeration value for the quantity of coespecific pollen grains deposited on the flower's stigma(s) exposed to multiple visitors at the end of flower anthesis

**Problems:**

- multiple visitors implies **multiple Interactions**. The term is linked to one or more Interactions or none at all (if the Interaction is not being recorded, just the coespecific pollen grains in an exposed flower).
- at the end of flower anthesis implies that the measurement have be taken in a specific time of flower longevity. It is some kind expressing the protocol used, and not a general term for pollen grains quantity. Should we change it to be more general? How?



# What do we expect from you?

- Engage on GitHub discussions
  - express your point of view on the Terms, their definitions and applications
  - Help to achieve a consensus (the best practical option, not necessarily the perfect one)
  - Eliminate ambiguity on terms definitions
- Provide examples of datasets
- Help to find and adopt controlled vocabularies
- Be part of the international community that defines biodiversity standards
  - recognition, and have your voice heard

# Hands-on GitHub!

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:  
THERE ARE  
14 COMPETING  
STANDARDS.

14?! RIDICULOUS!  
WE NEED TO DEVELOP  
ONE UNIVERSAL STANDARD  
THAT COVERS EVERYONE'S  
USE CASES.

YEAH!



SOON:

SITUATION:  
THERE ARE  
15 COMPETING  
STANDARDS.