



**BIODT**  
biodiversitydigitaltwin

# School & Hackaton

March 3rd – March 7th  
Officine Cantelmo, Lecce IT

## DATA PAPERS

#CiteTheDOI

Dmitry Schigel – GBIF Secretariate

(supported by Dag Endresen, GBIF Norway, University of Oslo)



Funded by  
the European Union



# READ THE CLASSICS FROM THE INVENTORS

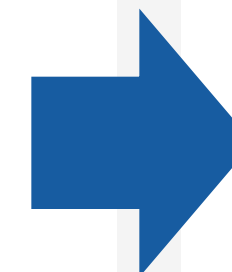
“...a scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than to report a research investigation... contains facts about data, not hypotheses and arguments in support of those hypotheses based upon data, as found in a conventional research article”

- to provide a **citable journal publication** that brings scholarly credit to data creators
- to describe the data in a **structured human-readable** form
- to bring the existence of the data to the **attention** of the scholarly community

Penev L, Mietchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Ó Tuama É, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017)

**Strategies and guidelines for scholarly publishing of biodiversity data.**

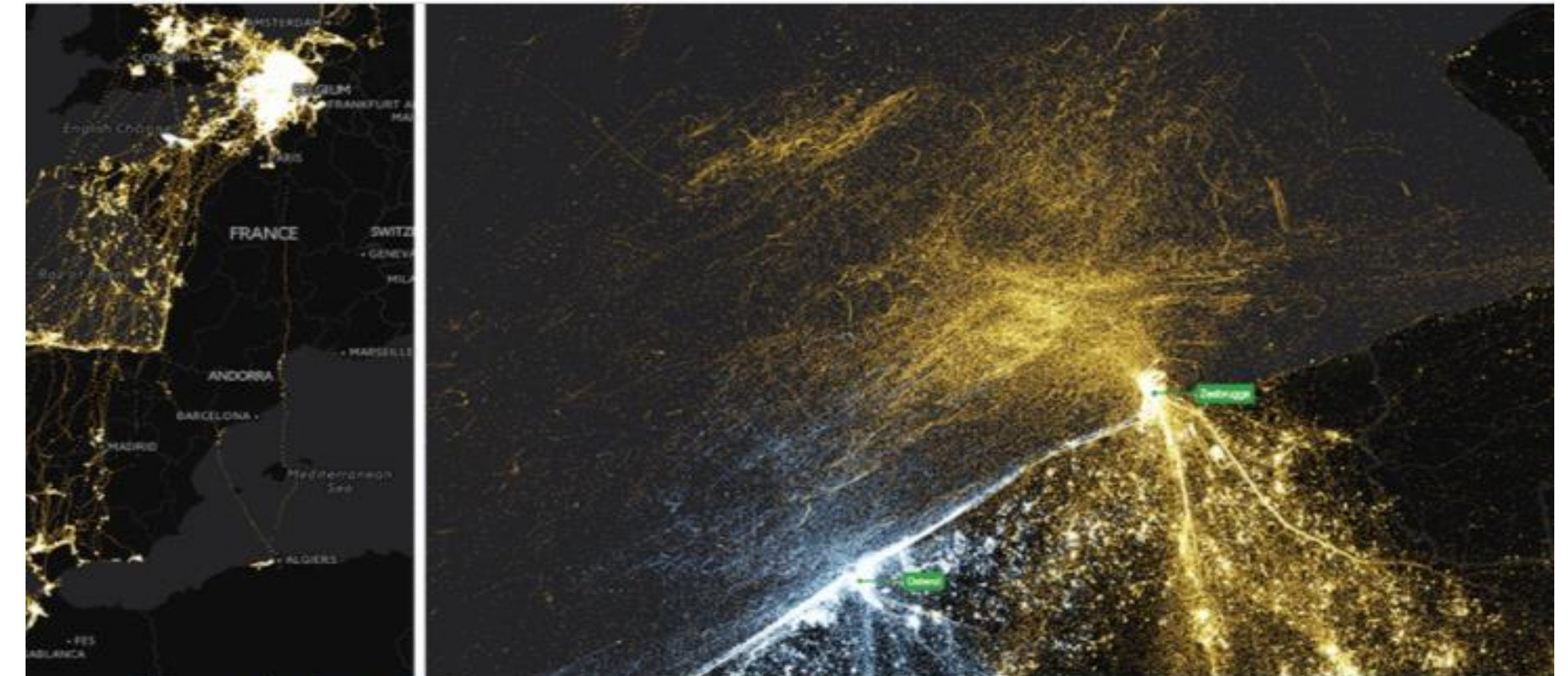
Research Ideas and Outcomes 3:e12431. <https://doi.org/10.3897/rio.3.e12431>



The screenshot shows the Rio journal website. The header includes the Rio logo (an elephant) and navigation links: Home, Articles, About, About Pensoft, Books, and Journals. Below the header, there's a 'Guidelines' tab and a 'Reviewed v1' badge. The article title is 'Data Published in Data Papers'. The abstract defines a data paper as a scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than to report a research investigation. It contains facts about data, not hypotheses and arguments in support of those hypotheses based upon data, as found in a conventional research article. Its purposes are three-fold: to provide a citable journal publication that brings scholarly credit to data creators, to describe the data in a structured human-readable form, and to bring the existence of the data to the attention of the scholarly community. The text further explains that the description should include several important elements (usually called metadata, or "description of data") that document, for example, how the dataset was collected, which taxa it covers, the spatial and temporal ranges and regional coverage of the data records, provenance information concerning who collected and who owns the data, details of which software (including version information) was used to create the data, or could be used to view the data, and so on. It also states that most Pensoft journals welcome submission and publication of data papers, that can be indexed and cited like any other research article, thus bringing registration of priority, a permanent publication record, recognition, and academic credit to the data creators. In other words, the data paper is a mechanism to acknowledge efforts in authoring 'fit-for-use' and enriched metadata describing a data resource. The general objective of data papers in biodiversity science is to describe all types of biodiversity data resources, including environmental data resources. Finally, it notes that an important feature of data papers is that they should always be linked to the published datasets they describe, and that link (a URL, ideally resolving a DOI) should be published within the paper itself.



### Getting scholarly recognition for your datasets



Map excerpt from [data paper](#) by Stienen et al. in Zookeys describing a bird tracking dataset.

A recorded webinar introducing data papers and providing helpful guidelines and tips for writing them is available from [here](#).

A data paper is a peer reviewed document describing a dataset, published in a peer reviewed journal. It takes effort to prepare, curate and describe data. Data papers provide recognition for this effort by means of a scholarly article.



Search videos, people



GBIF  
Global Biodiversity  
Information Facility

- Takes time to write a data paper and (usually) costs money to publish it
- Improves reusability of your data by providing a rich source of metadata to the user
- Currently datasets are not being properly cited (see published study), especially when only published to GBIF and not accompanied by a data paper. The data paper provides a mechanism for proper recognition/ citation of you as the data publisher

<https://academic.oup.com/database/article/doi/10.1093/database/bay033/4965857>



## 3 years ago | More



GBIF BUSINESS

+ Follow

More from GBIF

☒ Autoplay next video



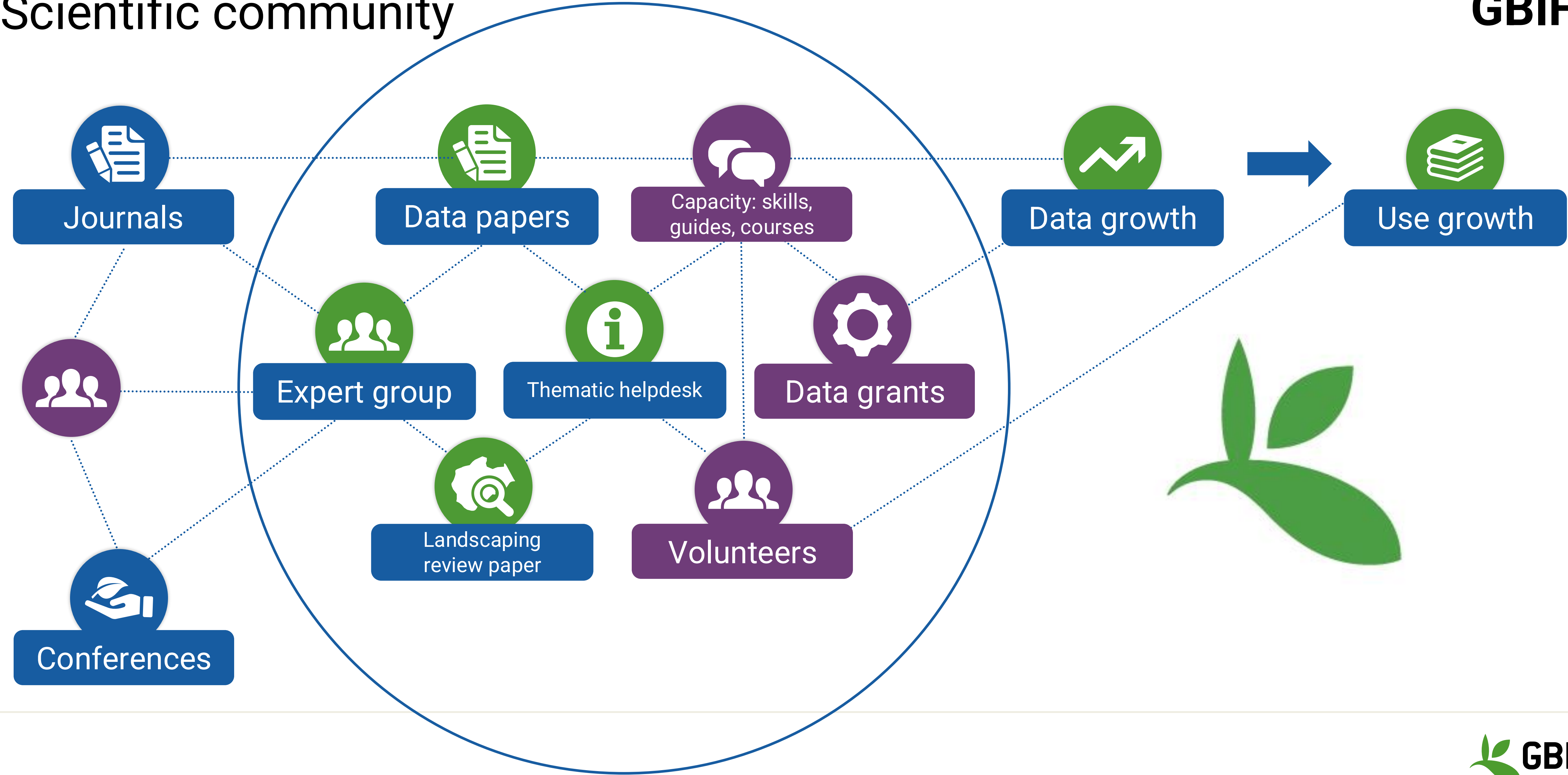
Webinar: In

GBIF

# DRIVING THEMATIC ROADS

Scientific community

GBIF

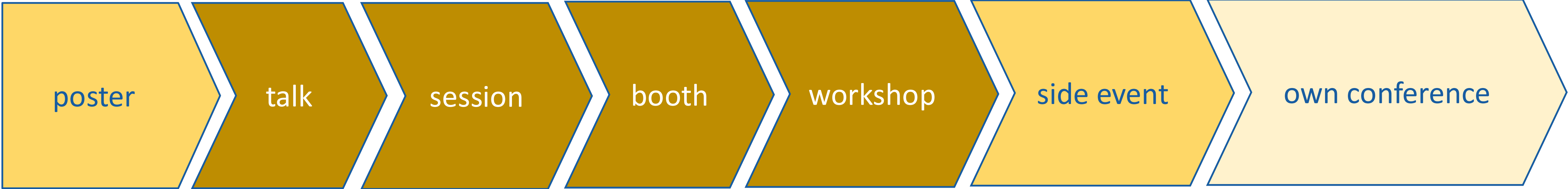


# TOOL: CONFERENCES



Conference

awareness  
publisher, user  
recruitment  
promoting calls  
projects, partnerships  
brewing



-> -> -> effort, efficiency, cost -> -> ->

# TOOL: DATA PAPERS (SUPPORTED FROM GBIF CALLS)



Data papers

5 years  
7 special issues  
2,217,088 new records in GBIF



Data growth

|         |                               |                           |           |                   |
|---------|-------------------------------|---------------------------|-----------|-------------------|
| 2020-21 | Biota of Russia               | Biodiversity Data Journal | 59 papers | 1,171,270 records |
| 2022    | Northern Eurasia              | Biodiversity Data Journal | 11 papers | 86,833 records    |
| 2022    | Vectors of human diseases I   | GigaByte                  | 11 papers | 555,805 records   |
| 2023    | Freshwater biodiversity data  | Journal of Limnology      | 12 papers | 134,973 records   |
| 2023    | Soil biodiversity             | Biodiversity Data Journal | 3 papers  | 96,585 records    |
| 2023    | Vectors of human diseases II  | GigaByte                  | 8 papers  | 171,622 records   |
| 2025    | Vectors of human diseases III | GigaByte (in progress)    | 0 papers  | 0 records         |

# Data sharing: the *when* choices

SHORT TERM

LONG TERM

perspective

## RESEARCH PHASE

- file formats
- ownership
- metadata
- storage
- backups

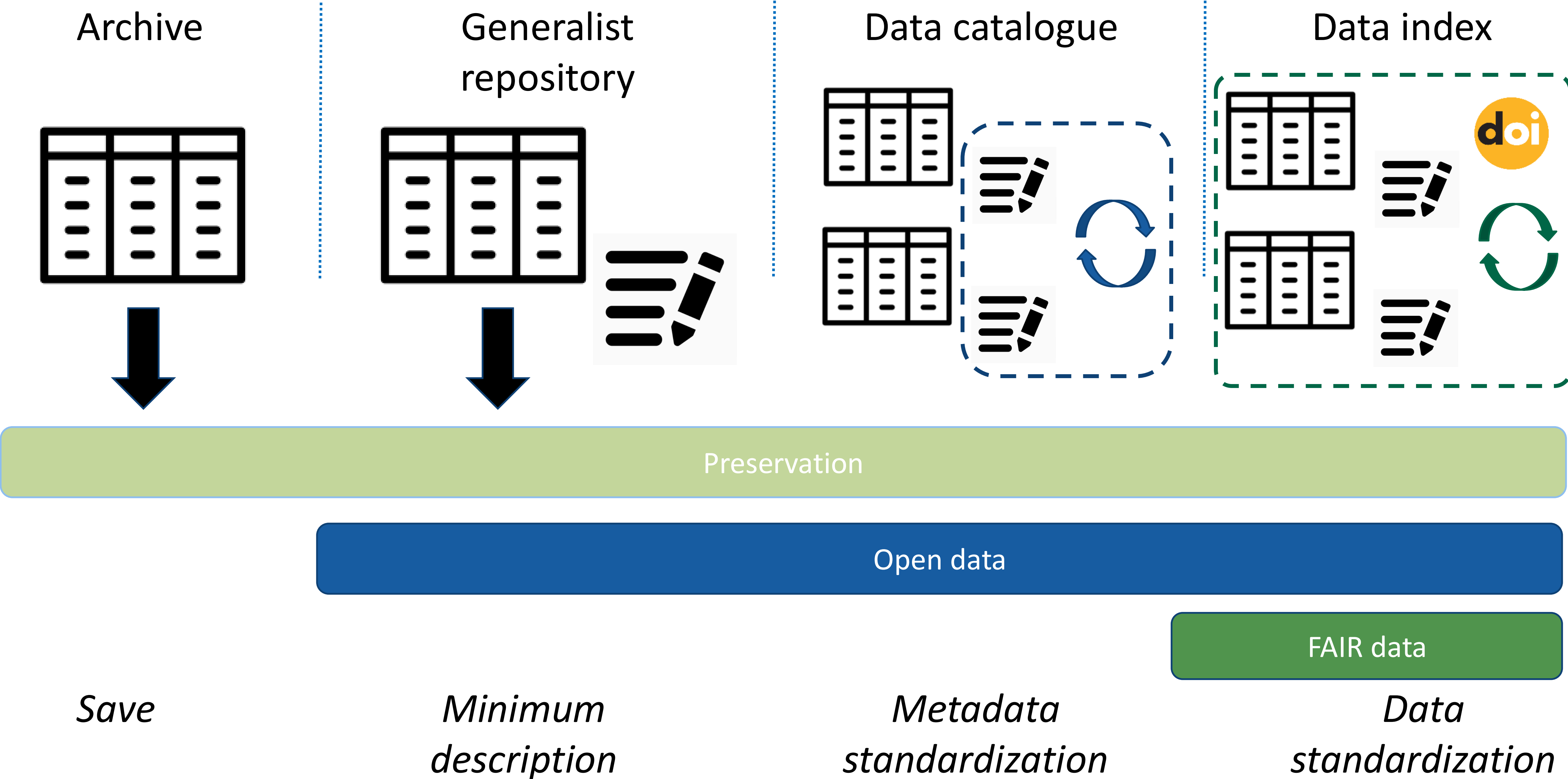
## DISSEMINATION PHASE

- share with whom?
- embargo?
- licensing
- metadata

## PRESERVATION PHASE

- repository?
- long-term manager?

# Data sharing: the **where** choices



Dmitry Schigel  
dschigel@gbif.org

@dschigel

