

Background

Taxonomy has a long tradition of describing earth’s biodiversity. For the last 20 years or so, taxonomic revisions have become available in PDF format, which is regarded as a good means of digital dissemination. However, a PDF document is nothing more than a text document that can be transferred easily among researchers.

In today’s world, traditional taxonomic techniques need to be met with novel tools to make data dissemination a reality, make species hypotheses more robust, and open the field up to rigorous scientific testing.

Cybertaxonomic tools provide such methods and are here summarized in the context of revisionary taxonomy. While many of the tools have been around for some time now, very few practicing taxonomists embrace and utilize these tools in their publications.

Cybertaxonomic tools provide ways to:

- re-use, re-purpose, and test previously gathered data easily
- make work of future taxonomists easier
- move taxonomy into the 21st century

The example discussed here is taken from Dikow (2012) – a revision of two mydas-fly genera from Namibia and South Africa.

Open access & unique identifiers

In order for species and specimen data to be linked (Fig. 1), they need to be openly accessible in data depositories and utilize unique identifiers (for example, GUID = Globally Unique Identifier, DOI = Digital Object Identifier).

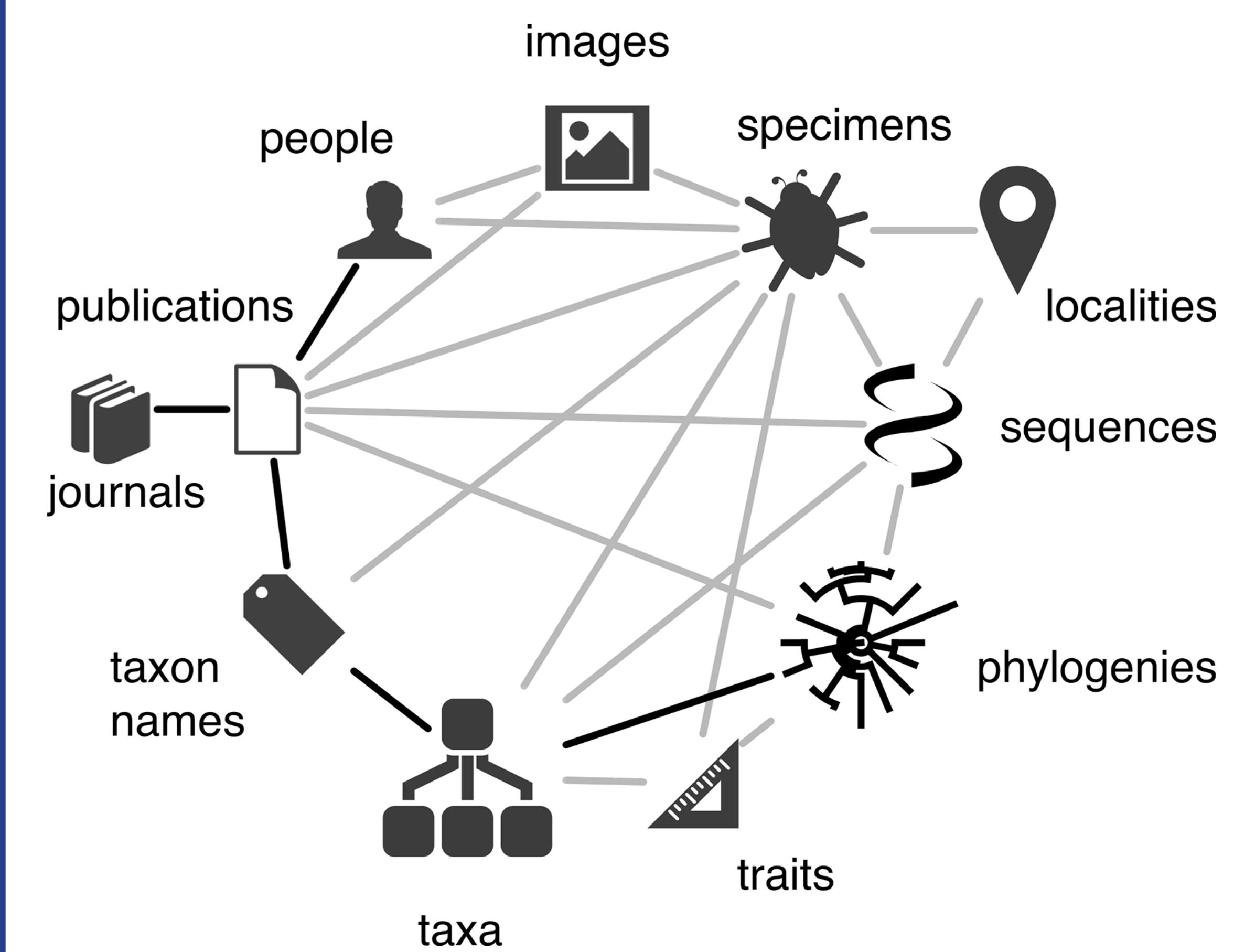


Fig. 1. Biodiversity knowledge graph proposed by Page 2016 (10.3897/rio.2.e8767)

Data uploaded to open-access depositories can easily be re-used and referenced through unique identifiers. Data aggregators such as the Encyclopedia of Life (EoL) can harvest them for species pages and researchers can build data networks to always provide the latest information on species such as a taxonomic catalog (Dikow & Agosti 2015).

Conclusion

Taxonomy and taxonomic revisions need to provide the primary data for biodiversity research. Therefore, a “Data Resources” section should be included to summarize and easily make accessible all data depositories to which data have been uploaded. Users of biodiversity data such as ecologists or conservation scientists can then utilize basic occurrence data, for example, to answer questions across taxa a single taxonomist cannot cover.

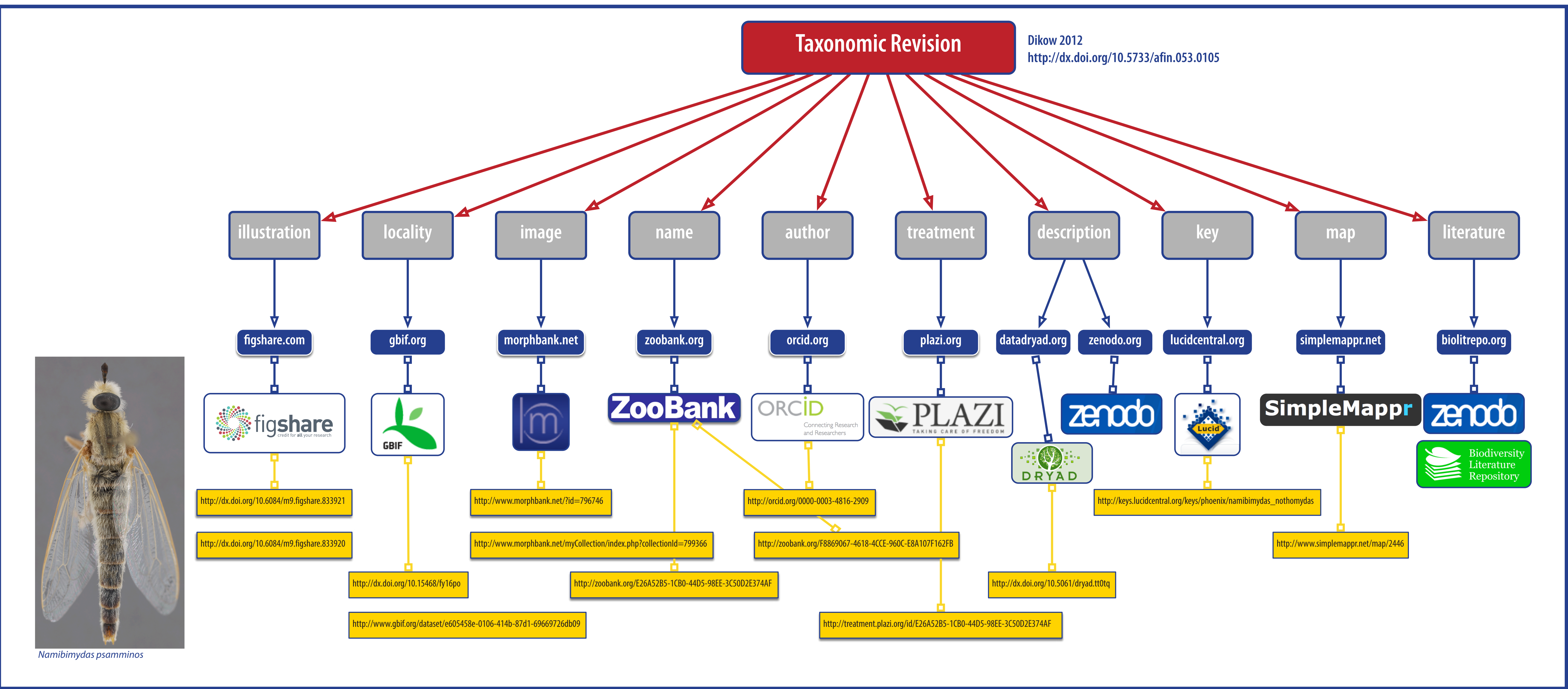
A taxonomist’s legacy can be shaped by making data ready for re-use and re-purposing by others in the future within and outside of taxonomy. Ultimately species hypotheses are more easily testable by providing data in machine-readable form and observations can be replicated.

Acknowledgements & References

The author thanks Donat Agosti, Lyubomir Penev, and Rod Page for discussions over the past few years.

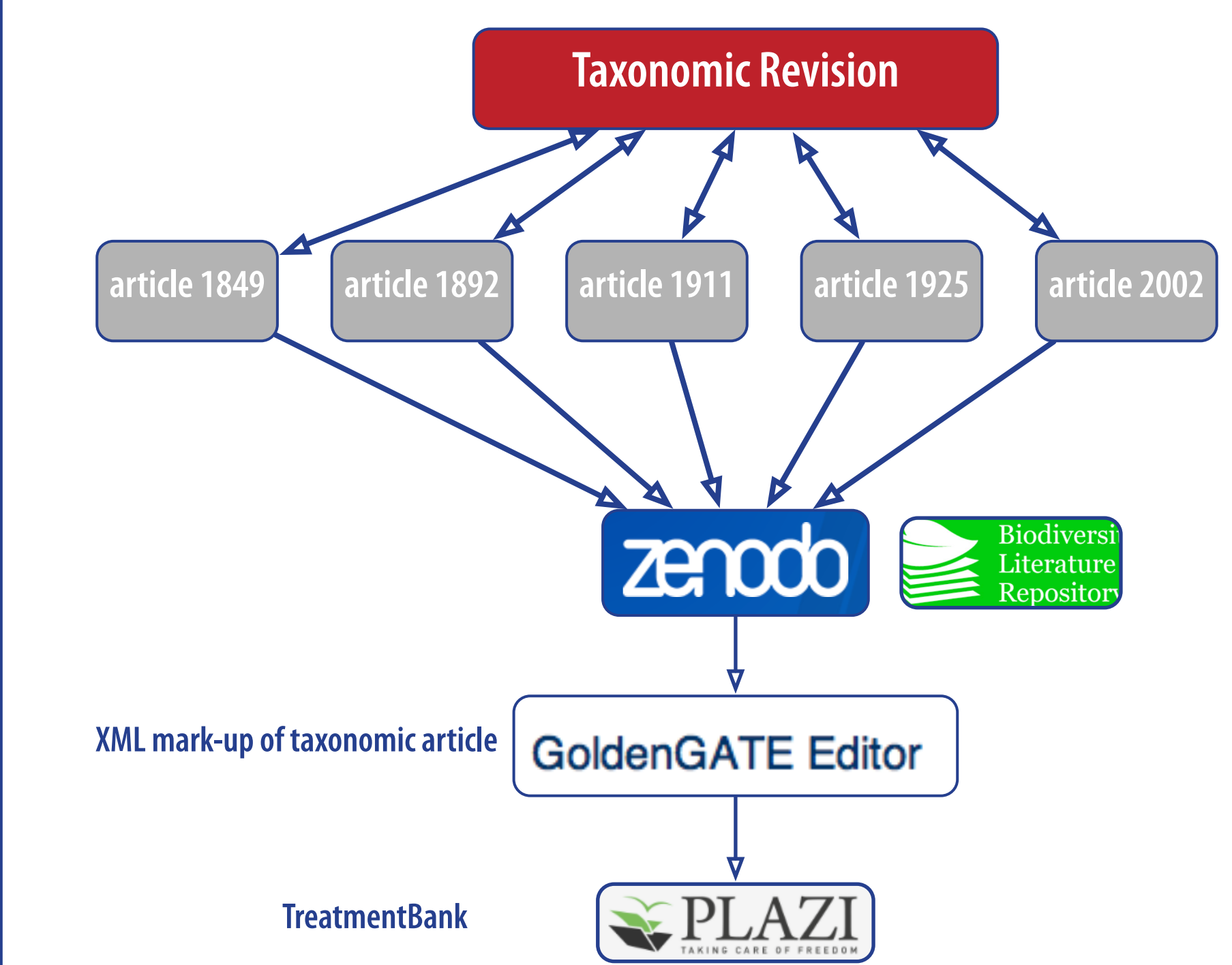
- Catapano 2010. <https://www.ncbi.nlm.nih.gov/books/NBK47081>
Dikow 2012. doi:10.5733/afin.053.0105
Dikow & Agosti 2015. doi: 10.3897/BDJ.3.e5707
Page 2013. doi:10.7717/peerj.190
Page 2016. doi:10.3897/rio.2.e8767
Robertson *et al.* 2014. doi:10.1371/journal.pone.0102623

Poster at figshare doi:10.6084/m9.figshare.5346259



Plazi

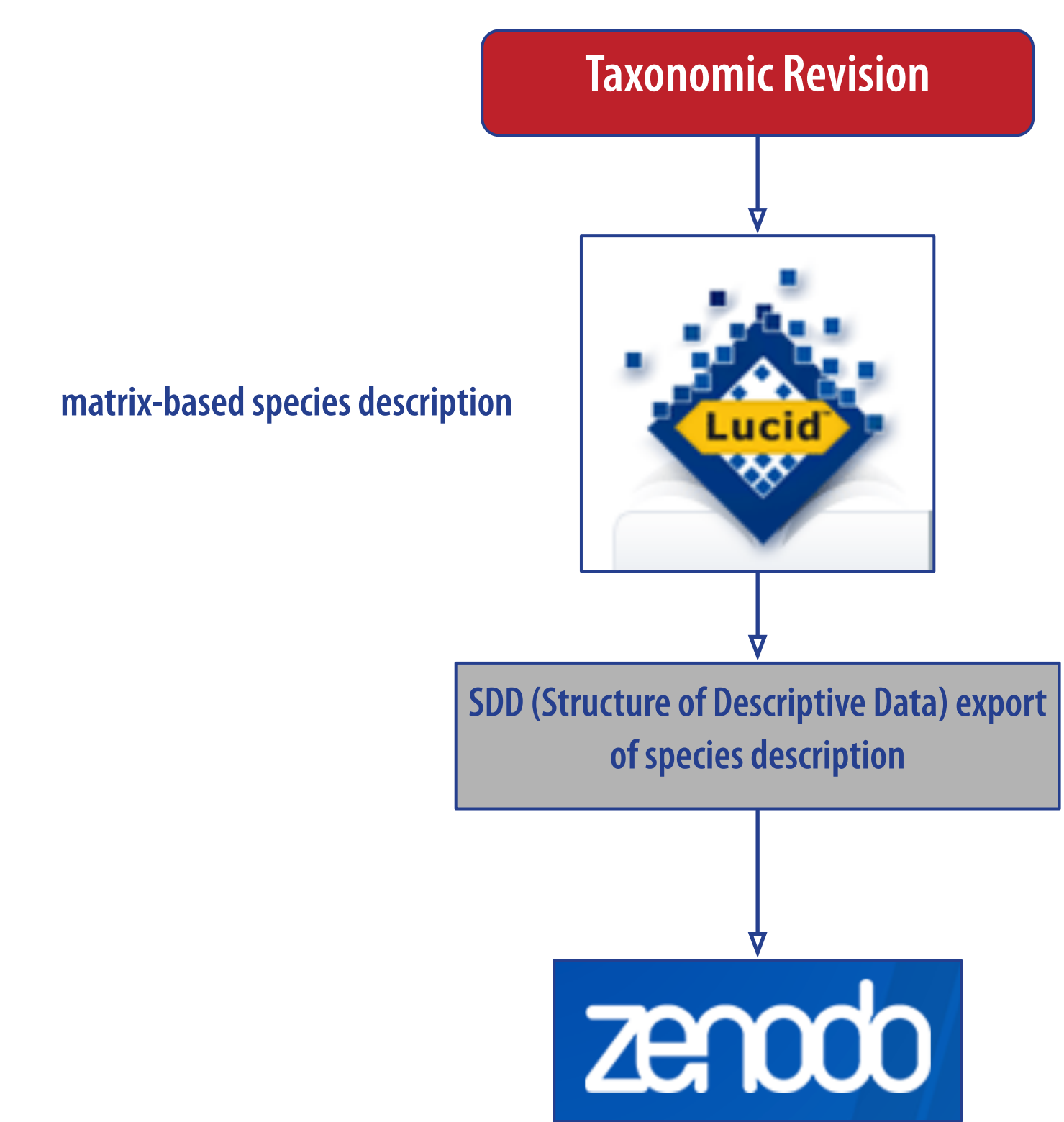
Plazi is an association supporting and promoting the development and service of persistent and openly accessible digital taxonomic literature and its contents. The main emphasis is to provide human- and machine-readable access to taxonomic treatments in TreatmentBank (<http://plazi.org/resources/treatmentbank/>) in TaxonX XML-format and data therein as well as to make them easily citable and retrievable. A treatment is a part of an article that is explicitly provided by an author to define his understanding of a taxonomic name usage at the time of publication (Catapano 2010). Treatments can either be provided for new species or for re-descriptions and include references to previous name usages and treatments.



Structure of Descriptive Data

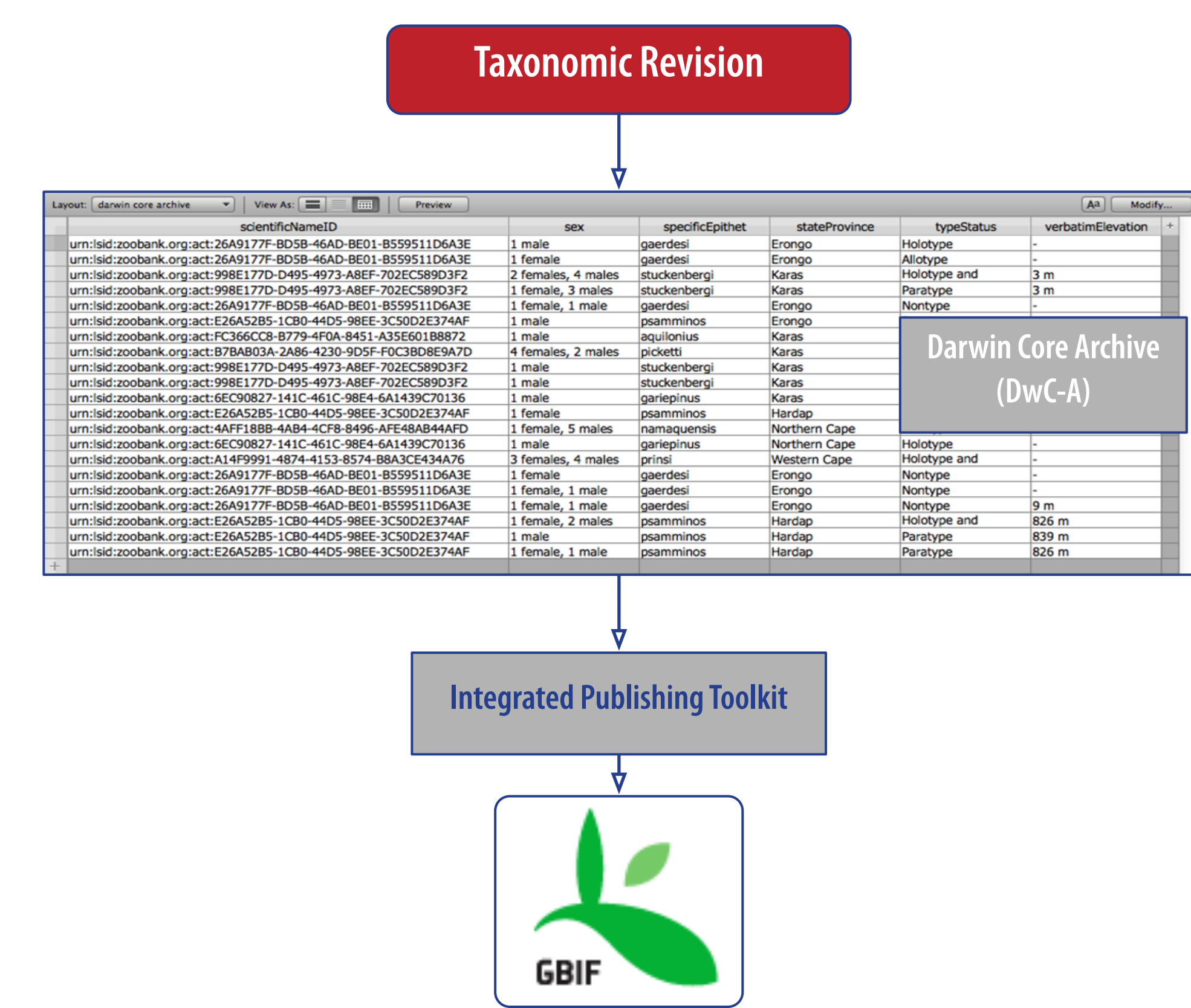
Developing species descriptions in a matrix editor such as Lucid Builder or Delta has several advantages. From such a matrix, matrix-based interactive identification keys, natural-language species descriptions, and XML versions can be exported.

Structure of Descriptive Data (SDD) is an XML-standard that allows for species descriptions be machine-readable. Providing your new species descriptions or re-descriptions of known species in SDD-format, allows future users to utilize exactly the same features for a new species description. It will make taxonomic treatments more comparable and ideally promote the usage of identical terminology and feature sets within a given taxon.



GBIF

The Global Biodiversity Information Facility (GBIF) is a data aggregator that gathers specimen occurrence data from numerous natural history collections and herbaria around the world. It provides free and open access to biodiversity data on all types of life on Earth. Data can also be directly submitted by a researcher to GBIF through the use of an Integrated Publishing Toolkit (IPT, Robertson *et al.* 2014). A third way to upload data to GBIF is through articles in which occurrence / observation records are either marked-up *a priori* (e.g., Pensoft’s Biodiversity Data Journal, <https://bdj.pensoft.net/>) or *a posteriorly* (Plazi treatments) and made accessible via a Darwin-Core Archive (DwC-A) to GBIF.



Literature

Imagine you have every original description and every subsequent re-description of a species at your fingertips.

The Biodiversity Heritage Library (BHL, <http://www.biodiversitylibrary.org>) is a digital archive of natural history literature and works collaborative to make biodiversity literature openly available as part of a global biodiversity community. It digitizes any natural history literature that is out of copyright and published prior to 1923. In copyright books and journals can also be digitized by BHL after an agreement with a publisher has been signed. The community can propose titles to be digitized by BHL free of charge on the BHL scanning request form.

BioStor (Page 2013) works with BHL content and itemizes the digitized journal volume into the respective articles, which are the

standard unit of citation. It also makes the itemized article available as a single page PDF download and therefore provides a great resource to obtain open-access publications digitized by the BHL.

The Biodiversity Literature Repository (BLR) is part of CERN’s digital Zenodo archive for scientific data. The BLR is focused on biodiversity literature, specifically articles and illustrations. The items stored are at article- or subarticle-level (e.g., individual treatments) for which a Digital Object Identifier (DOI) is provided, which allows for a citation of the item in a standardized way. Furthermore, it has the potential to assign DOIs to all legacy literature and with that make these articles first-class citizens. Upload of articles is open and free to anybody. Generally, any article published before 2000 without a DOI can be made accessible.