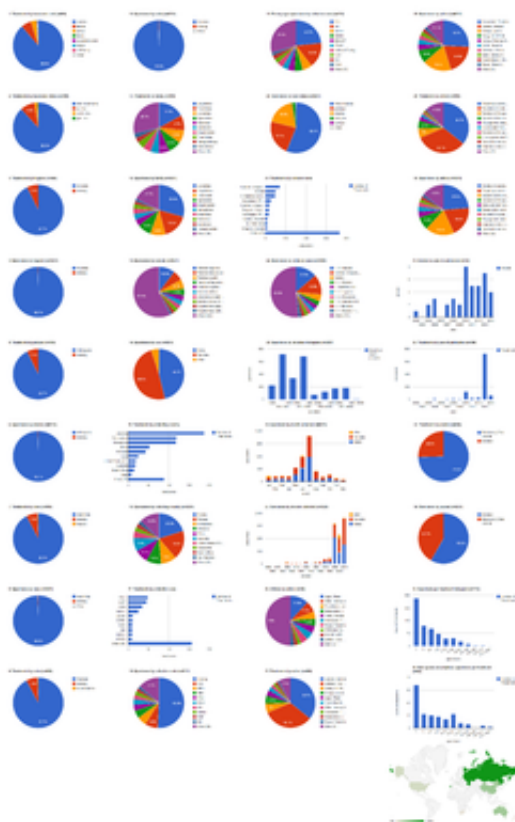


Integrating and visualizing primary biodiversity data from prospective and legacy taxonomic literature

May 15 2015



Dashboard chart summarizing content from 37 open access articles published in *Zootaxa* and five articles published in *Biodiversity Data Journal* containing treatments on spiders. These charts illustrate interoperability of data from XML-based publishing and subsequently marked up legacy literature. Credit: Jeremy A. Miller

XML markup of taxonomic research and specimen data is a valuable tool for structuring the incessantly accumulating biodiversity knowledge. It allows for the opportunity to collectively use the currently fragmented information for more detailed analysis.

A new [research paper](#), published in the *Biodiversity Data Journal*, demonstrates how XML markup using GoldenGATE can address the challenges presented by unstructured legacy data, like those presented in the widely used PDF format. The paper demonstrates how structured primary biodiversity data can be extracted from such legacy sources and aggregated with and jointly queried with data from other Darwin Core-compatible sources, to present a visualization of these data that can communicate key information contained in biodiversity literature.

Specimen data in taxonomic literature are among the highest quality primary [biodiversity data](#). Innovative cybertaxonomic journals such as the *Biodiversity Data Journal* are using workflows that preserve the data's structure and semantic specificity and disseminate electronic content to aggregators and other users that makes these data reusable.

Such structure however is lost in traditional taxonomic publishing and currently, access to that resource is cumbersome, especially for non-specialist data consumers.

The question is: how do you manage this vast distributed repository of knowledge about biodiversity to make it easily available reusable for future research?

To answer this challenge this project queried XML structured articles published in *Biodiversity Data Journal* along with historical taxonomic literature marked up using GoldenGATE, and represents the results as a series of standard charts. XML structured documents are maintained by the Swiss NGO [Plazi](#) and are freely available online.

In such form, data associated with specimens becomes much more valuable as it can reveal key information about a particular species, and even about the scientists who investigate them. Charts indicate at a glance, for example, what time of year and elevation range a species is likely to be found at, useful information if you want to search for it in the field.

Our accumulated biodiversity knowledge includes an estimated 2-3 billion specimens in natural history collections and 500 million pages of printed text. These are the data we need to answer questions that are relevant to our world today, like setting conservation priorities and anticipating the effects of climate change on biodiversity and ecosystem functions that affect the lives of people.

"In short, we have half a billion pages worth of biodiversity knowledge and are just learning how to query it. The real power comes when data from many articles are combined, queried, and reused for new purposes. Potential applications span the scientific, policy, and public spheres. When we all have better access to the information that already exists in the global corpus of [biodiversity](#) literature, this helps us do a better job of exploring what we don't know and wisely applying what we do." explains the lead author Dr Jeremy Miller, Naturalis Biodiversity Center.

More information: Miller J, Agosti D, Penev L, Sautter G, Georgiev T, Catapano T, Patterson D, King D, Pereira S, Vos R, Sierra S (2015) Integrating and visualizing primary data from prospective and legacy taxonomic literature. *Biodiversity Data Journal* 3: e5063. [DOI: 10.3897/BDJ.3.e5063](https://doi.org/10.3897/BDJ.3.e5063)

Provided by Pensoft Publishers

Citation: Integrating and visualizing primary biodiversity data from prospective and legacy taxonomic literature (2015, May 15) retrieved 15 June 2024 from <https://phys.org/news/2015-05-visualizing-primary-biodiversity-prospective-legacy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.