

Bioschemas for Biodiversity

Introducing a White Paper for its Implementation and Adoption in NFDI4Biodiversity and beyond



Lars Möller, Marcus Ernst, David Fichtmüller, Uwe Schindler, Claas-Thido Pfaff, Stefan Seifert, Cornelia Fürstenau, Lorenz Reimer, Julius Witte, Christian Ebeling

Bioschemas

What is Schema.org?

Schema.org is a set of terms to describe web content with structured and machine readable metadata using JSON-LD. This allows search engines to better understand the content of a website, e.g. the opening hours of a store. It was originally developed by a group of major search engine providers and focuses on a wide and generic set of terms.

What is Bioschemas?

Bioschemas is a community driven set of extensions to schema.org specifically designed for applications in the life-sciences. It extends existing terms from schema.org, e.g. Dataset, and defines new ones e.g. Taxon, BioSample or Gene.



What are the advantages of Bioschemas?

Using Bioschemas allows institutions to share metadata about their datasets as well as individual data records in a standardized and widely accepted way. Search engines will recognize the core elements from schema.org, e.g. datasets will show up in the Google Dataset Search.

Bioschemas in NFDI4Biodiversity

Which role will Bioschemas play in NFDI4Biodiversity?

Using Bioschemas as a common meta-data format within NFDI4Biodiversity offers many advantages as it increases the findability, accessibility, and reusability of our biodiversity data sets even beyond the consortium. It captures the essential information from diverse data types in enough detail to fully support the faceted search of the GFBio Search Portal. This also lays the groundwork for future data exchange with other consortia, such as DataPLANT and FAIRagro.

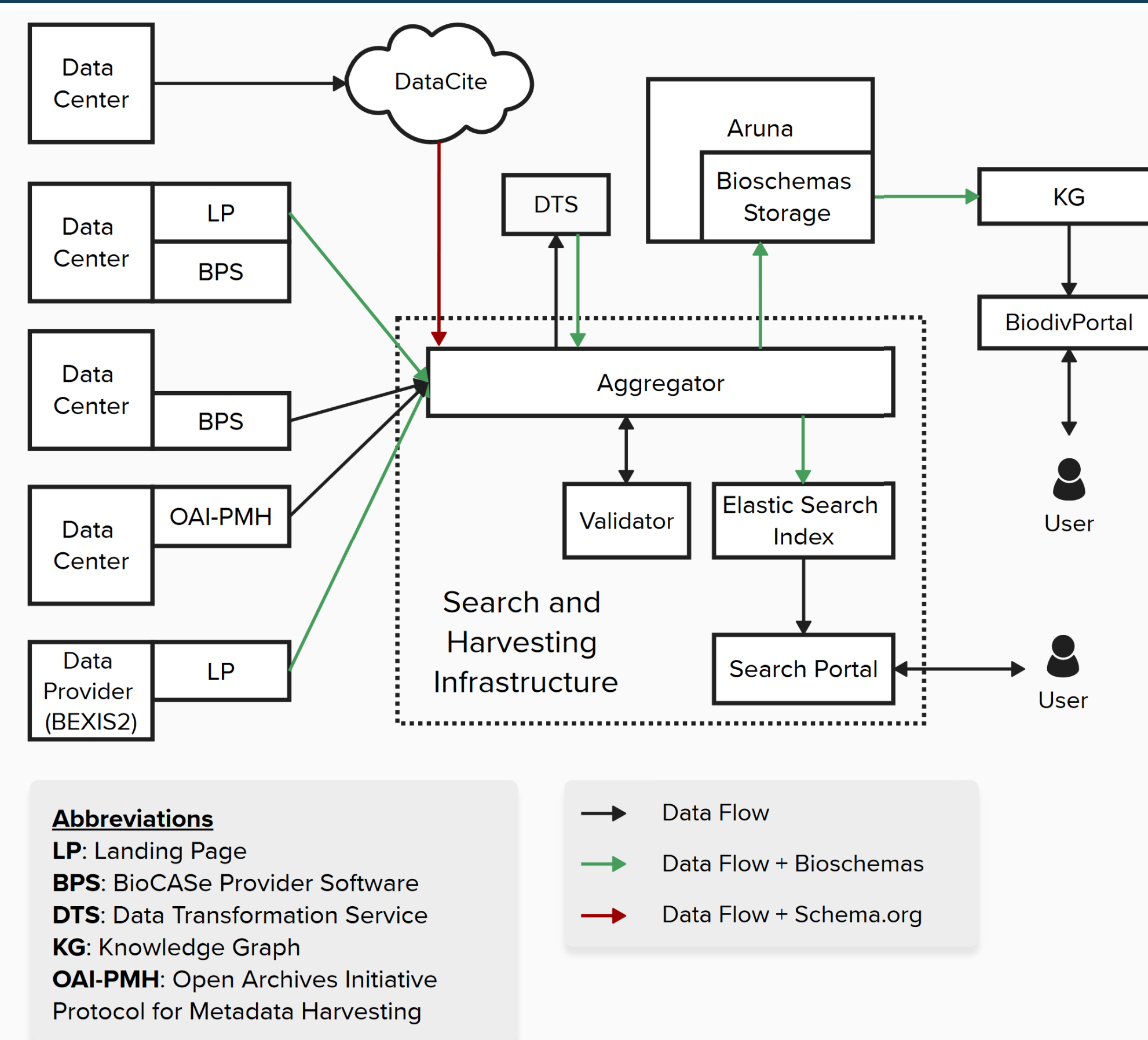


Fig 1: Schematic view of NFDI4Biodiversity Infrastructure Components that could be relevant for Bioschemas metadata generation and processing.

How will the Bioschemas markup be produced?

Bioschemas markup can be generated in several ways: Data Centers and Providers may embed it directly into their dataset landing pages or conveniently convert their existing data using the Data Transformation Service (DTS). This empowers providers to fully leverage the benefits of Bioschemas. Alternatively, if neither of these options are realizable, the Aggregator will automatically handle the conversion of data, ensuring all data are standardized and discoverable.

Bioschemas Implementation Interest Group and the White Paper

What is the BslIG?

The **Bioschemas Implementation Interest Group (BslIG)** is a group of people from NFDI4Biodiversity that explore the further use of schema.org and Bioschemas within the project. We meet about once a month in a video call and one of the outcomes so far is the White Paper presented here. Further information can be found at <https://kb.gfbio.org/display/NFDI/Bioschemas+Implementation+IG>.

What is the White Paper about?

The White Paper introduces schema.org and Bioschemas, as well as the benefits for their adoption. It describes the NFDI4Biodiversity infrastructure components that will play a role in creating and processing the markup and outlines the intended changes and the current state of adoption at different partners. It shows how the *Five Major Data Types for Biological Data* can be expressed in Bioschemas including a detailed mapping for specimen data in ABCD.

Where can I find the White Paper?

The white paper has recently been submitted, and the pre-print can be found under:

Advancing FAIR Biodiversity Data: Bioschemas Implementation in NFDI4 Biodiversity. Lars Möller, Marcus Ernst, David Fichtmüller et al. 2025. F1000. <https://tiny.cc/bioschemas-white-paper>



Joint Plenary
2025 NFDI4Earth meets
NFDI 4
BIODIVERSITY

www.nfdi4earth.de | www.nfdi4biodiversity.org

NFDI4Biodiversity and NFDI4Earth are funded by the German Research Foundation (DFG). NFDI4Biodiversity DFG project number 442032008, NFDI4Earth DFG project number 460036893