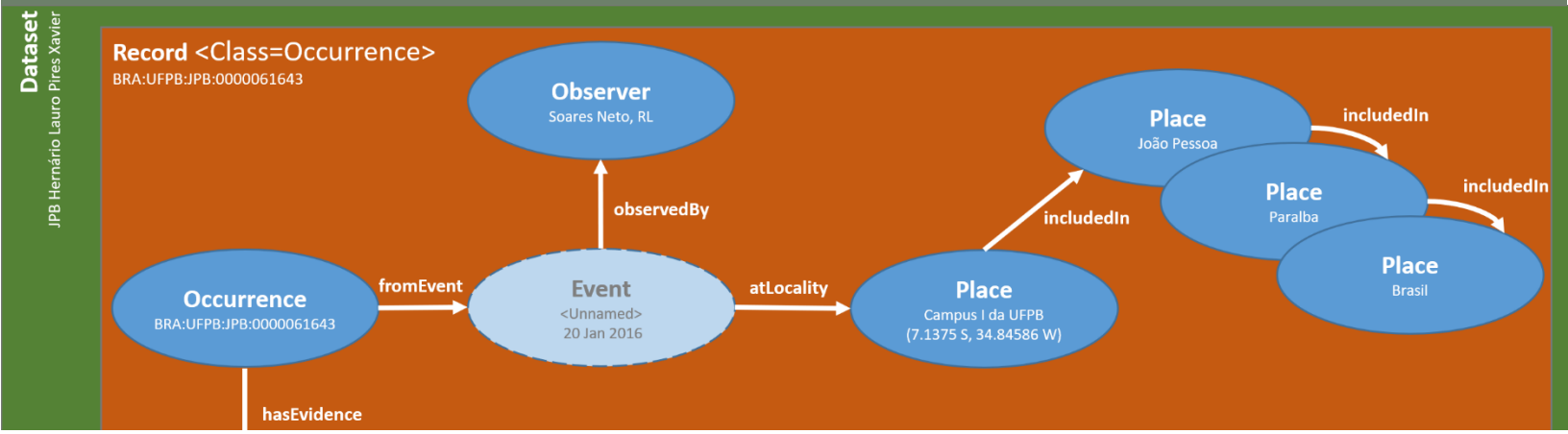
Flexible Packaging of Biodiversity Data



Flexible Packaging of Biodiversity Data

Revise and rationalize existing biodiversity information data standards and adopt a simple but flexible common packaging format for sharing rich data sets on all aspects of biodiversity.

Rationale

International biodiversity informatics initiatives, under the coordination of TDWG (Biodiversity Information Standards), have developed effective standards and models for interchange of data on natural history specimens, field surveys, multimedia objects, species checklists, traits and linkages to genomic data products, among others. Unfortunately, despite efforts within TDWG, these various initiatives have not yet combined to enable data publishers easily to share the entire content of many biodiversity data sets as a single explorable data export or graph. A high-level revision of these standards is required, including class definitions for the common categories of concepts and entities which recur throughout these data sets (taxon concepts, scientific names, specimens, sampling events, etc.) and a modular approach for data holders to publish additional properties related to their observations and measurements. In addition, the community needs to replace the commonly used Darwin Core Archive packaging format with one that can accommodate richer and more complex structures (graphs) of data connections within a dataset.

Status: **RESEARCH** Value: **VERY HIGH** Readiness: **VERY HIGH**

Estimated costs: *Research and Development* - **€150,000** *Operationalisation* - **€300,000** *Ongoing annual* - **€50,000**

Elements to accommodate

- TDWG data standards, including historical work on a TDWG ontology
- Darwin Core Archive format
- Modern data packaging formats, including Frictionless Data and W3C CSV on the Web
- Software and tools currently sharing data using existing data standards

Remaining challenges

- Ensure future-proofing or backwards compatibility for existing published data sets
- Agree high-level domain model for classes of concepts and entities references in biodiversity information
- Resolve overlaps between existing standards and vocabularies
- Establish governance structures for future development of these standards

Flexible Packaging of Biodiversity Data

GBIO Component	Significance of this investment
Published Materials	Opportunity to present rich data graphs representing content of published documents Generalised model for easier adoption by publishers, editors and authors as supplementary materials
Collections and Specimens	Opportunity to publish more complex information structures and relationships to represent the complexity of collections data
Field Surveys and Observations	Greater flexibility to represent a more complete subset of data elements captured during field work, with more sophisticated options for complex data relationships Continued simplicity for most uses and ease of use e.g. in citizen science contexts
Open Access and Reuse Culture	More consistent and flexible standards will simplify and accelerate adoption of consistent practice for open access and resuse
Data Standards	Fundamental modernization of existing standards for interoperable exchange of biodiversity data from diverse sources Common model for packaging data for discovery, reuse and archival
Persistent Storage and Archives	Standardized packaging of data from diverse sources will simplify management, use and storage of data for all stakeholders Opportunity to develop intelligent repositories for biodiversity data sets
Policy Incentives	Greater coordination with CBD Parties will assist with adoption of common approaches to ensure mobilization of necessary data

Supporting stakeholders

>>> Institutional logos here for stakeholders with particular interests in promoting delivery of this component <<<