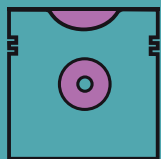
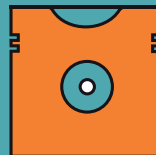




Biomedical  
vocabularies



Tape library



Treatments  
Figures



Database  
(MongoDB)



Database  
(MongoDB)



Distributed tape library



Treatments  
Figures



Mapped  
entities



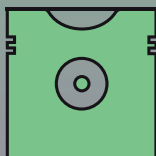
BioC format



Biomedical  
vocabularies



BioC format  
{;}



Database  
(MongoDB)



Mapped  
entities



Tape library



Treatments  
Figures

Scientific Treatments and  
Supplementary Indices  
Revolutionize  
the Reuse of Research Data



Mapped  
entities



## Scientific Treatments and Supplementary Indices Revolutionize the Reuse of Research Data

### Enrichment of published literature at scale

The **automated** generation of **scientific treatments** through analysis of literature has been pioneered by Plazi and SIBiLS in the life sciences and by Europa Institute Basel in the social sciences and humanities. This radically improves accessibility of scientific facts – even material published in copyrighted literature – enabling the creation of new open FAIR digital annotation object resources.

Plazi's **TreatmentBank** and Biodiversity Literature Repository **BLR** are now one of the principle suppliers of taxonomic treatments from existing publications to the Global Biodiversity Information Facility **GBIF**. BLR is a key community of Zenodo, the global catch-all repository for research data. **Plazi** and **SIBiLS** infrastructures form important components of the BiCIKL Horizon2020 project – The Biodiversity Community Integrated Knowledge Library.

### Distributed Research Data Tape Library

**Institutional Guarantee (IG)** solutions being developed by the DataFutures' **hasdai** partnership with CERN break through cost barriers which prevented operation of sustainable data resources. IG sweeps away the prohibitive costs of retaining local IT departments and proprietary alternatives, which have limited retrieval facilities and difficult-to-forecast egress costs.

The Distributed Research Data Tape Library provides a new community-governed and operated infrastructure based on international standards. The tape library provides automated server components, distributed among hosting institutions, plus consortium-operated remote management. Use of **industry-standard LTO technology** enables secure copies of corpora to be stored at multiple locations for decades at low cost, and automatically copied to new media as part of the LTO roadmap.

### SIBiLS Supplementary Data Index

Key information in medical literature – particularly that relating to genomic variants is often not reported in publications' full-text, but only in supplementary material such as imagery and tables.

**Annotating** and **searching** of this **supplementary data** enables significant improvement in the volume of documents retrieved for a variant – **reducing by about 70%** the number of variants for which no match is found in the scientific literature. Supplementary data now represents a paramount source of information for curating variants of unknown significance.

SIBiLS

**Swiss Institute of Bioinformatics Literature Services (SIBiLS)** provides indices of biological literature, allowing fully customizable search for semantically-enriched content, based on bio-medical entities from a growing set of standardized and legacy vocabularies. Its services are used extensively for curation of genes and gene products, by delivering customized literature triage engines to curation teams.

<https://candy.hesge.ch/SIBiLS/>



**Plazi** is a global leader in liberation of data in biodiversity publications as machine-readable open FAIR digital objects. Plazi is the single largest provider of data sets to Zenodo and to the Global Biodiversity Information Facility (GBIF), with over 800K taxonomic treatments and 480K figures from approximately 56 thousand articles liberated into the public domain.

<https://plazi.org/>

datafutures

**DataFutures GmbH** is a not-for-profit company which leads development of annotation and long-term preservation infrastructure in the InvenioRDM consortium. The hasdai partnership of European and U.S. institutions, which operates a network of repositories and tape libraries, is managed by DataFutures and governed under a memorandum with CERN.

<https://www.data-futures.org/>



Biomedical  
vocabularies



Treatments  
Figures



mapped  
entities

