Summaries of each infrastructure, their tools and services for the Biodiversity Knowledge Hub

---

Please use this document to start to shorten the summaries of each infrastructure, their tools and services which we will publish to the BKH to help visitors understand what services and resources the various Infrastructures can supply.

Here is the “front-end” where you can see the infrastructure descriptions, if your infrastructure is not there please ensure your description fits within the limits indicated: http://193.204.79.46/bridge/fair-data-place/explore-infrastructures-organisations

The descriptions below **should be shortened to 250 characters at the maximum.** Other two sections should inform the potential user on **what kind of MAIN data and services the RI offers to the USERs** (scientists and others). Each of these Data and Services sections should not exceed 200 characters each.

The part on “What the RI will provide in BiCIKL” should be erased, as the BKH will operate far beyond the expiration of the BICIKL.

The re-worded descriptions should match the following structure (as exemplified of OpenBiodiv and ARPHA-XML RIs)::

OpenBiodiv

Name of the infrastructure: OpenBiodiv

Supporting organisation: Pensoft Publishers

Brief description: A biodiversity knowledge graph encompassing Linked Open Data extracted from literature by Pensoft and Plazi and converted to RDF following the OpenBiodiv-O ontology. [MAXIMUM 250 characters!]

Data types: RDF triples on specimens, sequences, taxon names, taxonomic treatments, persons, articles, bibliographies, and others. [MAXIMUM 200 characters!]

Services: XML-to-RDF conversion, SPARQL endpoint, search engine, user-oriented applications for literature exploration. [MAXIMUM 200 characters!]

Website: https://openbiodiv.net
API: SPARQL endpoint: http://graph.openbiodiv.net/

ARPHA-XML

Name of the infrastructure: ARPHA-XML

Supporting organisation: Pensoft Publishers

Brief description: Full-featured publishing platform supporting the full manuscript life cycle: authoring, peer review, publication and dissemination, with the ARPHA Writing Tool (AWT) at the core of it. The host of the Biodiversity Data Journal, RIO Journal, and others. [MAXIMUM 250 characters!]

Data types: Published information on specimens, sequences, taxon names, taxonomic treatments, bibliographies, and others. [MAXIMUM 200 characters!]

Services: ARPHA-XML provides services to researchers and journals, including several unique services for biodiversity: import, exchange and semantic enrichment of data in an entirely XML-based environment. [MAXIMUM 200 characters!]

Website: https://arphahub.com
API: https://arphahub.com/about/OAI-PMH

Descriptions of Research Infrastructures

EMBL-EBI

Name of the infrastructure: European Nucleotide Archive (ENA)

Supporting organisation: European Molecular Biology Laboratories - European Bioinformatics Institute (EMBL-EBI)

Brief description: The European Nucleotide Archive (ENA) is an open, supported platform for the management, sharing, integration, archiving, and dissemination of sequence data. This authoritative nucleotide sequence repository comprises the European node of the INSDC. [MAXIMUM 250 characters!]

Data types: projects, samples, read data, sequences, assemblies and other analyses, taxon names [MAXIMUM 200 characters!]

Services: ENA Browser, Metadata Search, Taxonomy services, Webin Submission Portal, Helpdesk services, support documentation, RESTful APIs [MAXIMUM 200 characters!]

Website: https://www.ebi.ac.uk/ena
API: Discovery API: https://www.ebi.ac.uk/ena/portal/api; Retrieval API: https://www.ebi.ac.uk/ena/browser/api/; ENA taxonomy: https://www.ebi.ac.uk/ena/taxonomy/rest/#/; Source Attribute Helper API: https://www.ebi.ac.uk/ena/sah/api/; Cross References API: https://www.ebi.ac.uk/ena/xref/rest/ ;

Europe PMC

Name of the infrastructure: Europe PMC

Supporting organisation: European Molecular Biology Laboratories - European Bioinformatics Institute (EMBL-EBI)

Brief description: Europe PMC is a global, free, life sciences literature repository that provides comprehensive access to life sciences literature (research articles and preprints) from trusted sources. It is partnered with PubMed and PubMed Central. [MAXIMUM 250 characters!]

Data types: literature [MAXIMUM 200 characters!]

Services: Europe PMC has built tools and services on top of the literature indexed:

- ORCID claiming tool
- External links service
- Grant finder
- SciLite annotations tool
- Article status monitor [MAXIMUM 200 characters!]

Website: https://europepmc.org/
API: Articles RESTful API: https://europepmc.org/RestfulWebService; Annotations API: https://europepmc.org/AnnotationsApi; Grants RESTful (Grist) API: https://europepmc.org/GristAPI

GBIF

Name of the infrastructure: Global Biodiversity Information Facility

Supporting organisation: Global Biodiversity Information Facility

Brief description: GBIF is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth. [MAXIMUM 250 characters!]

Data types: occurrences, specimens, observations, eDNA, sequences, taxon names, taxonomies, collection catalogue, images, literature, training materials. [MAXIMUM 200 characters!]

Services: portal, Cloud-based downloads, search engine, RESTful JSON based API, taxonomic comparisons, collection catalogue. [MAXIMUM 200 characters!]

Website: https://www.gbif.org/
API: RESTful JSON based: https://api.gbif.org/v1/

LifeWatch ERIC

Description of the infrastructure: [Maximum 1000 characters]

Name of the infrastructure: e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research

Supporting organisation: e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research

Brief description: LifeWatch ERIC provides access to a multitude of datasets, webservices and other research resources, allowing the accelerated capture of data, their analysis and support for knowledge-based decision-making for biodiversity & ecosystem management. [MAXIMUM 250 characters!]

Data types: occurrences, observations, eDNA, sequences, taxon names, taxonomies, images, training materials. [MAXIMUM 200 characters!]

Services: portal, Cloud-based downloads, search engine, RESTful JSON based API, web services on all levels of biodiversity , organised in thematic services, workflows and virtual research environments (VREs). [MAXIMUM 200 characters!]

Website: https://www.lifewatch.eu

API: RESTful JASON APIs available at: https://trainingcatalogue.lifewatch.eu/site/api

CETAF

Consortium of European Taxonomic Facilities, Brussels, Belgium

Brief Description: CETAF is a network of 71 research institutions holding collections that represent more than 80% of the world's bio- and geodiversity. It serves as a knowledge and expertise exchange platform for researchers from a wide variety of scientific disciplines.
Data types: natural science collections, taxonomic expertise, research facilities, institutional.
Services: Distributed European School of Taxonomy (DEST) platform; CETAF Registry of Collections and Expertise
Website: https://cetaf.org/

DiSSCo

Name of the infrastructure: Distributed System of Scientific Collections (DiSSCo)

Supporting organisation: Naturalis Biodiversity Center

Brief description:

A World-class RI for integrated access to natural science collections in Europe that unifies all European natural science assets under common curation and access policies and practices.

Data types: FAIR Digital Objects serialised as JSON and JSON-LD for Digital Specimens and related objects such as Media objects, Annotations, Facilities and Agents.

Services: digital specimens, collection management systems integration, PID infrastructure, unified curation and annotation system, digitisation services integration, unified loans and visits system, DO APIs.

Website: https://dissco.eu, https://dissco.tech/,
API: RESTful JSON based: https://sandbox.dissco.tech/api/swagger-ui/index.html#/

Biodiversity Literature Repository (BLR)

Name of the infrastructure: Biodiversity Literature Repository
Supporting organisation: Plazi http://plazi.org/
Brief description: BHL provides findable, accessible, interoperable and reusable data liberated from scientific publications, with an emphasis on taxonomic treatments and figures with rich metadata.
Data types: Articles, treatments, figures accessible in JSON, HTML, PDF, PNG
Services: search and retrieval service
Website: http://plazi.org/blr/
API: https://developers.zenodo.org/

TreatmentBank

Name of the infrastructure: TreatmentBank https://treatmentbank.org/
Supporting organisation: Plazi http://plazi.org/
Brief description: TreatmentBank is a service liberating data from scientific publications to make the data imprisoned in publications widely accessible, citable and enriched with links to cited specimen, taxonomic and genomic data.
Data types: Articles, treatments, figures accessible in JSON, HTML, PDF, PNG
Services: data conversion, FAIRization, linking, dissemination, search and retrieval service of data liberated from scientific publications
Website: https://treatmentbank.org/
API: https://tb.plazi.org/GgServer/srsStats (treatment data);
https://tb.plazi.org/GgServer/dioStats (article data)

OpenBiodiv

Name of the infrastructure: OpenBiodiv

Supporting organisation: Pensoft Publishers

Brief description: A biodiversity knowledge graph encompassing Linked Open Data extracted from literature by Pensoft and Plazi and converted to RDF following the OpenBiodiv-O ontology.

Data types: RDF triples on specimens, sequences, taxon names, taxonomic treatments, persons, articles, bibliographies, and others.

Services: XML-to-RDF conversion, SPARQL endpoint, search engine, user-oriented applications for literature exploration.

Website: https://openbiodiv.net
API: SPARQL endpoint: http://graph.openbiodiv.net/

ARPHA-XML

Name of the infrastructure: ARPHA-XML

Supporting organisation: Pensoft Publishers

Brief description: Full-featured publishing platform supporting the full manuscript life cycle: authoring, peer review, publication and dissemination, with the ARPHA Writing Tool (AWT) at the core of it. The host of the Biodiversity Data Journal, RIO Journal, and others.

Data types: Published information on specimens, sequences, taxon names, taxonomic treatments, bibliographies, and others.

Services: ARPHA-XML provides services to researchers and journals, including several unique services for biodiversity: import, exchange and semantic enrichment of data in an entirely XML-based environment.

Website: https://arphahub.com
API: https://arphahub.com/about/OAI-PMH

LifeWatch ERIC. Biodiversity and Ecosystem Virtual Research Environments (BER_VREs)

LifeWatch ERIC, Seville, Spain
https://www.lifewatch.eu/web/guest/catalogue-of-virtual-labs

User defined platform of VREs uniquely using blockchain technology

LW ERIC is the e-Science European Research Infrastructure for **B**iodiversity and **E**cosystem **R**esearch. It provides access to a multitude of datasets, e-Services and tools enabling the construction and operation of Virtual Research Environments (VREs) through innovative technologies, which permit the accelerated capture of data, their analysis and knowledge-based decision-making support for biodiversity and ecosystem management. It can be split into 2 groups of services: **BER_e-Infra** & **BER_VREs**.

BER_VREs consists of 3 cutting-edge technologies:

1. LifeBlock, a Blockchain technology for transparency and immutability, guaranteeing FAIR-compliant data.
2. Tesseract, the technical composability layer to integrate web services, enabling the development of VREs for users to combine and arrange services and software into multiple workflows.

3. The Artificial Intelligence virtual laboratory (vLab), an online service allowing users to locate others working on similar subjects and the hubs that link them.

Role in BiCIKL: [Maximum 250 characters]

Within BiCIKL, LifeWatch ERIC will be particularly active in gathering and analysing users' technical requirements, as well as implementing the BKH. It will also contribute to ensuring the FAIR principles, supporting virtual access to **BER_VREs**.

LifeWatch ERIC. e-infrastructure services (BER_e-Infra)

LifeWatch ERIC, Seville, Spain
https://lifewatch.eu

Access to data, services and VREs on biodiversity and ecosystem research (BER)

Description of the infrastructure: [Maximum 1000 characters]

LW ERIC is the e-Science European Research Infrastructure for **B**iodiversity and **E**cosystem **R**esearch. It provides access to a multitude of datasets, e-Services and tools enabling the construction and operation of Virtual Research Environments (VREs) through innovative technologies, which permit the accelerated capture of data, their analysis and knowledge-based decision-making support for biodiversity and ecosystem management. It can be split into 2 groups of services: **BER_e-Infra** & **BER_VREs**.

BER_e-Infra provides web services, applications, software and data, guaranteeing FAIR-compliant data and FAIR(/SHARP)-compliant services, and their integration into workflows and VREs on: taxonomy & classification, phylogenetic algorithms, biogeography, ecology, metabarcoding, genomics and data services.

Role in BiCIKL: [Maximum 250 characters]

Within BiCIKL, LifeWatch ERIC will be particularly active in gathering and analysing users' technical requirements, as well as implementing the BKH. It will also contribute to ensuring the FAIR principles, supporting virtual access to **BER_e-Infra.**

SIB

SIBiLS
SIB, Geneva, Switzerland

Name of the infrastructure: SIBiLS (Swiss Institute of Bioinformatics Literature Services Services and HES-SO/HEG Genève)

Supporting organisation: Swiss Institute of Bioinformatics Text Mining group

Brief description: SIBiLS provide holistic access to the biological and biodiversity literature (e.g. PMC, MEDLINE, TreatmentBank): all articles are semantically enriched with reference terminologies and the resource is freely accessible via REST APIs. [MAXIMUM 250 characters!]

Data types: biomedical articles and material citations with annotations in JSON/JATS/BioC, queries in Lucene syntax and search results in Elasticsearch 's JSON. [MAXIMUM 200 characters!]

Services: literature retrieval engine. [MAXIMUM 200 characters!]

Website: https://candy.hesge.ch/SIBiLS
API: https://candy.text-analytics.ch/SIBiLS/MEDLINE/search.jsp

MBG

Name of the infrastructure: BR Herbarium

Supporting organisation: Meise Botanic Garden, Meise, Belgium

Brief description: A large botanic garden with all the collections and facilities you might expect from that. Large living collections both outdoor and in greenhouses. A large herbarium of plants, algae and fungi most of which has been digitised, most with an image.

[MAXIMUM 250 characters!]

Data types: Both living and preserved collections of plants, algae and fungi [MAXIMUM 200 characters!]

Services:

- Specimen portal (https://www.botanicalcollections.be/#/en/home)
- Living collections (https://www.botanicalcollections.be/#/en/search/living-collection)

Website: https://www.plantentuinmeise.be/en/
API: The data are accessible through the dataset published to GBIF https://www.gbif.org/dataset/b740eaa0-0679-41dc-acb7-990d562dfa37

BGBM-FUB

Name of the infrastructure: Herbarium B

Supporting organisation: Botanic Garden Berlin, Freie Universität Berlin

Brief description: The Botanic Garden Berlin is a centre of biodiversity research facilitating access to biodiversity data including data generated by the BGBM and data from cooperating scientists and infrastructures. The BGBM is committed to FAIR principles and provides services for 20 partner institutions.

Data types: Both living and preserved collections of plants, algae and fungi. Mutimedia.

Services:

- Specimen portal via JACQ (https://www.jacq.org/)
- Herbonauten (https://www.herbonauten.de/

Website: https://www.bo.berlin/

API: (via JACQ)
https://services.jacq.org/jacq-services/rest/JACQscinames/description.html

Zenodo

Name of the infrastructure: Zenodo

Supporting organisation: CERN

Brief description: Zenodo is a general-purpose repository for the long tail of research, developed and hosted at CERN. It eliminates barriers and enables researchers across all disciplines to share, preserve, cite, and discover all kinds of digital research artefacts in line with the FAIR principles.

Data types: No restrictions on data types. Currently, hosting mainly taxonomic treatments, and other related content coming from literature extraction (through Plazi).

Services: user-oriented (UIs) and machine-oriented (APIs) application for data deposition.

Website: https://zenodo.org
API: https://zenodo.org/api (docs: http://developers.zenodo.org)

PlutoF

Name of the infrastructure: PlutoF
Supporting organisation: University of Tartu
Brief description: PlutoF provides online services to manage biological data fully integrated and ready for analyses. Users can manage datasets through a full data life cycle - from uploading to publishing and archiving in FAIR format.

Data types: taxon names, taxon hypotheses, specimens, DNA sequences, samples, traits, observations, locality data, references.
Services: data management, data publishing in GBIF, INSDC, Datacite DOI, data export and import, collection management, molecular laboratory.

Website: https://plutof.ut.ee

API: https://plutof.docs.apiary.io/

Catalogue of Life (COL)

Name of the infrastructure: Catalogue of Life & ChecklistBank

Supporting organisation: Species 2000

Brief description: COL aims to deliver a consistent and up-to-date listing of all the world's known species; the COL Checklist. ChecklistBank, developed by GBIF and COL, is a publishing platform and data infrastructure for taxonomic and nomenclatural datasets.

Data types: taxonomic checklists, nomenclature datasources literature references, type specimens, species descriptions/treatments, OTUs

Services: DOIs for COL Checklist releases, taxon name identifiers, transformation into data standards (e.g. DwC-A, ColDP, DOT, Newick); tools to share, discover, analyse, compare and download taxonomic datasets

Websites: Catalogue of Life & ChecklistBank
API: https://api.checklistbank.org

Biodiversity Heritage Library (BHL)

Name of the infrastructure: Biodiversity Heritage Library (BHL)

Supporting organisation: Smithsonian Libraries and Archives

Brief description: BHL provides free access to nearly 300,000 volumes (over 60.5 million pages), from the 15th-21st centuries. Content includes public domain content as well as in-copyright materials made available through permission from rights holders.
Data types: Names, bibliographies, articles, full-text, PDFs
names, taxonomic treatments, persons, articles, bibliographies, and others.

Services: Citable bibliography references to taxonomic information; downloadable PDFs; bibliographic DOIs.
Website: https://www.biodiversitylibrary.org/
API: Available APIs listed on this page

Global Genome Biodiversity Network (GGBN)

Name of the infrastructure: GGBN
Supporting organisation: Botanic Garden Berlin, Freie Universität Berlin

Brief description: GGBN is an international network of institutions that share an interest in long-term preservation of genomic samples representing the diversity of non-human life on Earth. GGBN provides a platform for biodiversity biobanks worldwide to harmonize exchange and use of material in accordance with national and international legislation and conventions.

Data types: tissue/DNA samples, links to voucher specimens

Services: GGBN Data Portal, GGBN Document Library

Website: https://www.ggbn.org/

iBOL

Centre for Biodiversity Genomics in Guelph, Canada
https://ibol.org/

Description of the infrastructure: [Maximum 1000 characters]
The International Barcode of Life (iBOL, https://ibol.org/) Consortium brings together scientists and research organisations in more than 45 countries with a shared interest in developing DNA-based species identification systems. iBOL promotes the capture of short sequences known as DNA barcodes and maintains BOLD (Barcode of Life Data Systems, http://www.boldsystems.org/) as the reference database for these barcodes. Closely related barcodes are clustered and assigned a Barcode Index Number (BIN). Sequences from curated specimens enable many of these BINs to be associated with named species. As new specimens are sequenced, more of the BINs will be identified. The BIN system supports identification of DNA from any source, including specimens of all life stages, gut contents, soil, water and other environmental samples. High-throughput sequencing platforms allow large numbers of species to be diagnosed from a single sample, a process known as metabarcoding.

Role in BiCIKL: [Maximum 250 characters]
iBOL will contribute to BiCIKL through access to the tools and data associated with BOLD, working to enhance the linkages between barcode sequences and the specimens from

which they were derived and to support integration of BOLD BINs into the taxonomic framework developed by Species 2000 and other BiCIKL partners.

Supporting Research Infrastructures

Name of Infrastructure	Hosting institution	Location	Web site	Summary of service offered	Logo
BHL	Smithsonian Libraries and Archives	Washington D.C., USA	https://www.biodiversitylibrary.org/	Free access to global biodiversity literature repository	
GGBN	Botanic Garden and Botanical Museum Berlin	Berlin, Germany	https://www.ggbn.org/	Access to genomic biodiversity samples and data	
iBOL	Centre for Biodiversity Genomics in Guelph	Guelph, ON Canada	https://ibol.org/		
Europe PMC	EMBL-EBI	Hinxton, Cambridgeshire, UK	https://europepmc.org/	Open database of worldwide life sciences literature	

Biodiversity Information Standards (TDWG)

Name of the Infrastructure: Biodiversity Information Standards (TDWG)

Supporting organisation: International Taxonomic Databases Working Group (TDWG)

Brief Description: Biodiversity Information Standards (TDWG) creates, maintains, evolves, and promotes open, community-driven standards to enable sharing and use of biodiversity data for all. Our vision seeks to realise open, free, and interconnected biodiversity knowledge. Much of the data being linked through the BKH use standards and standard data packages developed by the international TDWG community. In addition to our website, you can find us on Slack, on Twitter @tdwg, and on FB.

Data types: data standards for natural science collections, taxonomic data, observations, genomic data, trait data, media data, geospatial data, citizen science, remote monitoring, data quality, api standards, attribution standards

Services: TDWG communities of expertise https://www.tdwg.org/community/; TDWG Standards https://www.tdwg.org/standards/

Website: https://www.tdwg.org