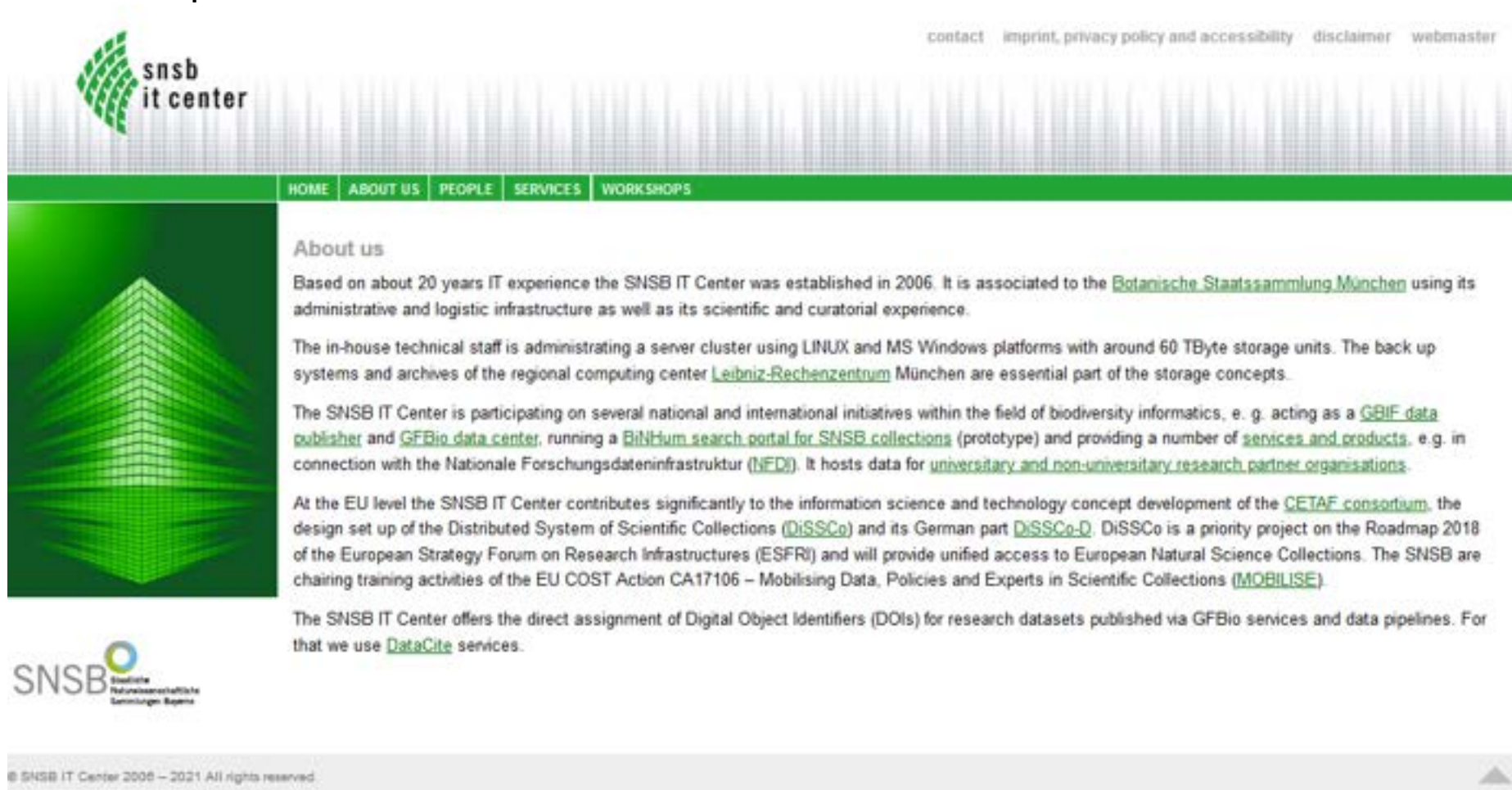


The SNSB data pipelines for publishing occurrence data via GBIF are appropriate for human osteological collections

Dagmar Triebel, Tanja Weibulat, Stefan Seifert, Markus Weiss



- SNSB has a biodiversity informatics department and a number of recognized data repositories



The screenshot shows the SNSB IT Center website. At the top left is the SNSB IT Center logo. To the right are links for 'contact', 'imprint, privacy policy and accessibility', 'disclaimer', and 'webmaster'. Below this is a green navigation bar with links for 'HOME', 'ABOUT US', 'PEOPLE', 'SERVICES', and 'WORKSHOPS'. The main content area features a large green graphic on the left and text on the right. The text describes the center's history, technical infrastructure, and participation in various international initiatives. At the bottom left is the SNSB logo, and at the bottom right is a copyright notice: '© SNSB IT Center 2006 – 2021 All rights reserved'.

snsb
it center

contact imprint, privacy policy and accessibility disclaimer webmaster

HOME ABOUT US PEOPLE SERVICES WORKSHOPS

About us


Based on about 20 years IT experience the SNSB IT Center was established in 2006. It is associated to the [Botanische Staatssammlung München](#) using its administrative and logistic infrastructure as well as its scientific and curatorial experience.

The in-house technical staff is administrating a server cluster using LINUX and MS Windows platforms with around 60 TByte storage units. The back up systems and archives of the regional computing center [Leibniz-Rechenzentrum München](#) are essential part of the storage concepts.

The SNSB IT Center is participating on several national and international initiatives within the field of biodiversity informatics, e. g. acting as a [GBIF data publisher](#) and [GFBio data center](#), running a [BINHum search portal for SNSB collections](#) (prototype) and providing a number of [services and products](#), e.g. in connection with the Nationale Forschungsdateninfrastruktur ([NFDI](#)). It hosts data for [university and non-university research partner organisations](#).

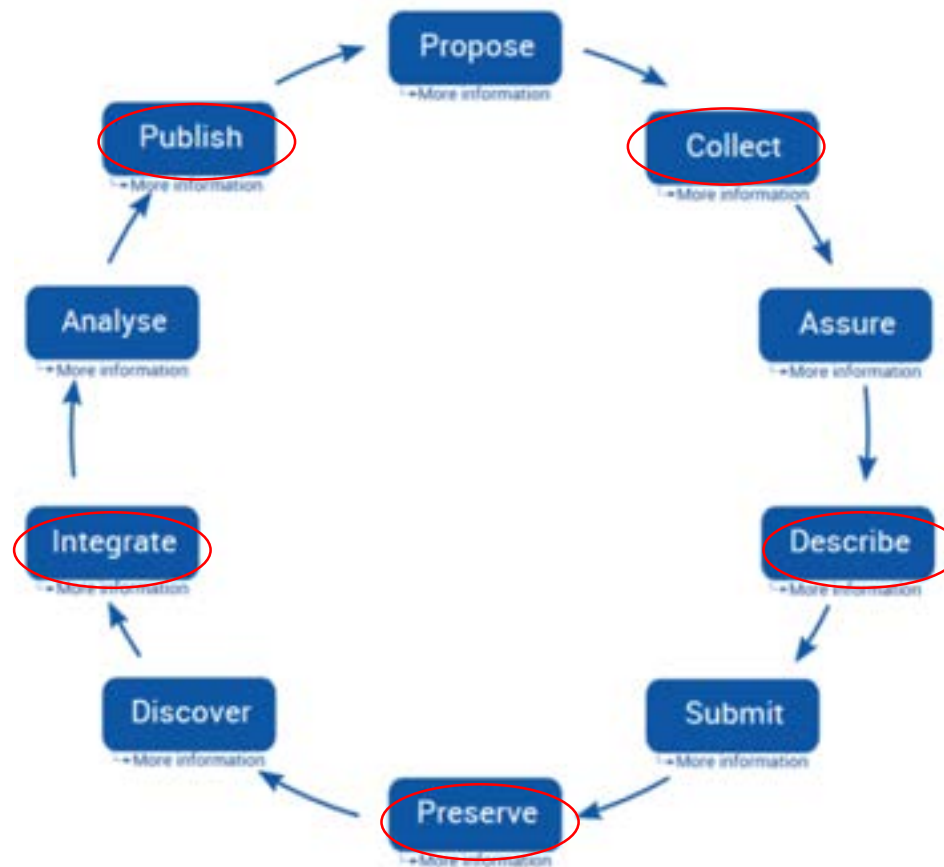
At the EU level the SNSB IT Center contributes significantly to the information science and technology concept development of the [CETAF consortium](#), the design set up of the Distributed System of Scientific Collections ([DiSSCo](#)) and its German part [DiSSCo-D](#). DiSSCo is a priority project on the Roadmap 2018 of the European Strategy Forum on Research Infrastructures (ESFRI) and will provide unified access to European Natural Science Collections. The SNSB are chairing training activities of the EU COST Action CA17106 – Mobilising Data, Policies and Experts in Scientific Collections ([MOBILISE](#)).

The SNSB IT Center offers the direct assignment of Digital Object Identifiers (DOIs) for research datasets published via GFBio services and data pipelines. For that we use [DataCite](#) services.

SNSB 
Staatliche
Naturwissenschaftliche
Sammlungen Bayerns

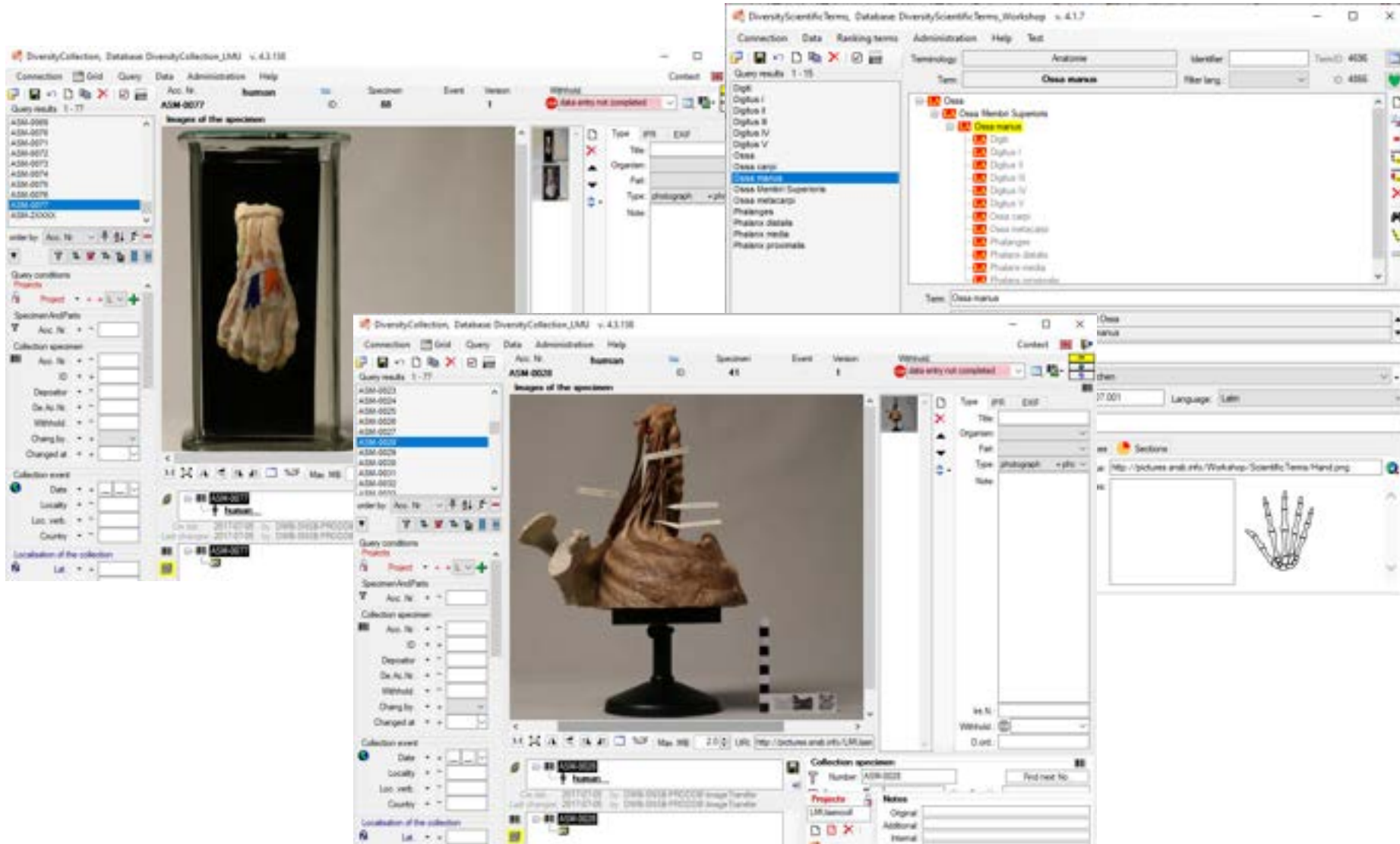
© SNSB IT Center 2006 – 2021 All rights reserved

- The core services for collection data are following the Data Life Cycle and are on the way to be certified.

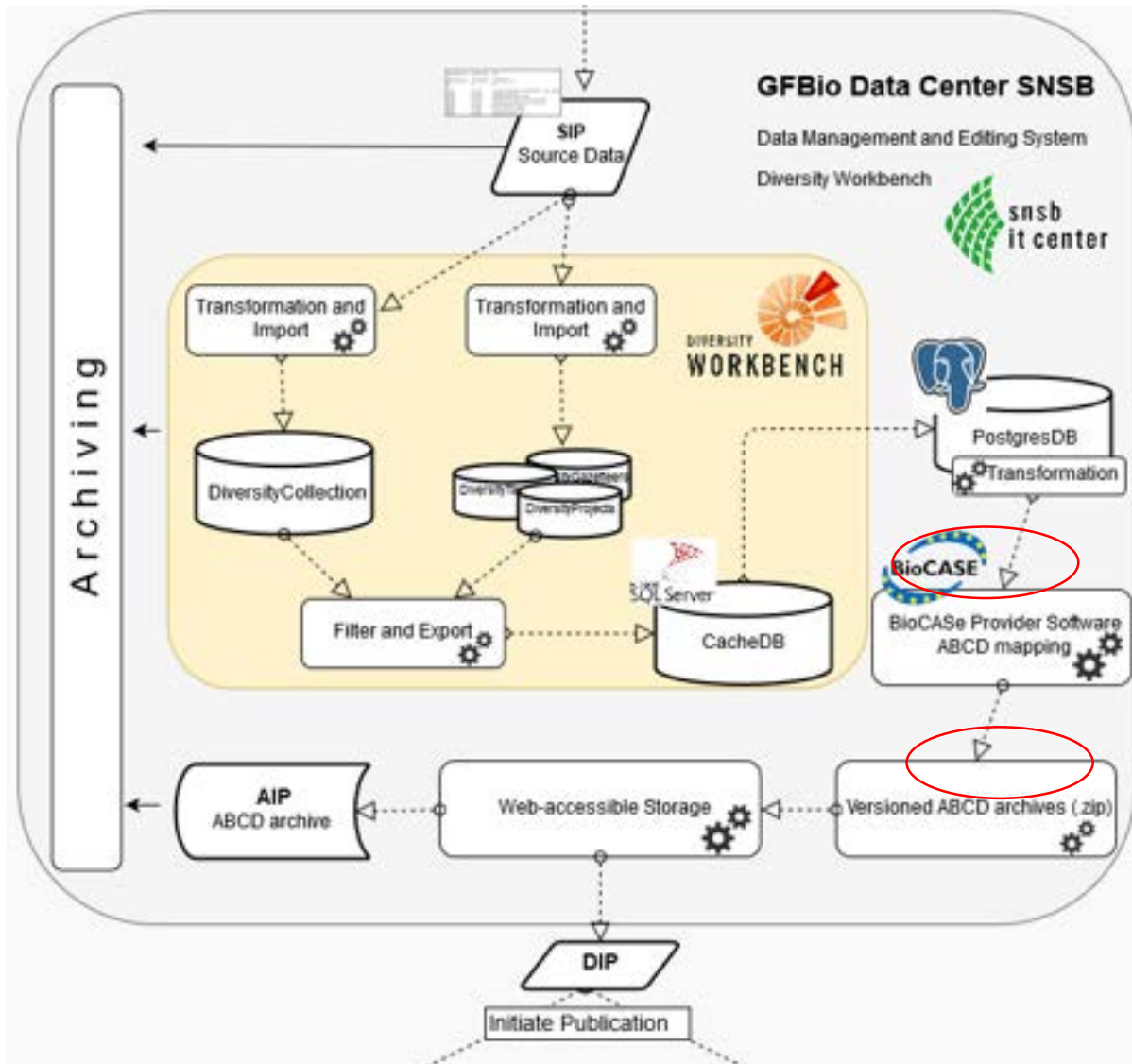


- Data management of all repositories is done in own installations of DWB databases (www.diversityworkbench.net)
- Administration of 200 DWB accounts
- For collection and occurrence data: DiversityCollection






SNSB Data Pipelines



Access to Biological Collection Data (ABCD)

The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data).

ABCD version
2.06 

Find us on

Permanent IRI

<http://www.tdwg.org/standards/115>

Abstract

The Access to Biological Collections Data (ABCD) comprehensive standard for the access to and specimens and observations (a.k.a. primary bio Schema attempts to be comprehensive and hij supporting data from a wide variety of databar several existing data standards. Parallel structu both) atomised data and free-text can be acco and 2.06 are currently in use with the GBIF (Glc Information Facility) and BioCAsE (Biological C for Europe) networks. Apart from the GBIF and potential for the application of ABCD extends t in-house legacy data access (e.g. datasets from shall not be converted and integrated into an i be kept separately, though easily accessible). B between terms, ABCD is a step towards an ont collections.

Creator

Access to Biological Collections Data task grou
Information Standards (TDWG)

Bibliographic citation

Access to Biological Collections Data task group (2007) Access to Biological Collection Data (ABCD), Version 2.06, Biodiversity Information Standards (TDWG) <http://www.tdwg.org/standards/115>

technical specification 2005 standard 2005

Standards / Access to Biological Collection Data (ABCD)

Header section

Title

Access to Biological Collection Data (ABCD) Schema

Date created

2005-09-16

Status

Current (2005) standard

Category

Technical specification

ABC D 3.0

... is here!

The ABCD 3.0 project was successfully completed on 2019-01-31. This site gives an overview about the outcomes.

What's new?

XML Schema

We developed a new [XML Schema](#). The changes we made were focused on element reuse, implementation of requests from the community and harmonization with the new ABCD 3.0 Ontology. A detailed documentation of all changes as well as an XPath Mapping can be found [here](#).

ABCD Ontology

The ABCD Standard is now described as an XML Schema and an Ontology. This allows the access of the standard through semantic queries, encourages element reuse and serves as basis for future software and services in the area of semantic web. The XML Schema is linked to the Ontology via SAWSDL Annotations. An Introduction to the features of the ontology is available [here](#).

- DWB data in BioCASE Provider Software:

» standard conversion, creation of XML archives

BioCASE Provider Software 3.7.3

Start

Welcome to the BioCASE provider software entrance page. This is BPS version 3.7.3.

Your BioCASE installation is up to date.

Documentation



Check the PyWrapper Wiki to find tutorials on installation, configuration, mapping, debugging, and other useful tips.

Config Tool

Configure new datasources, general options, the querytool, statistics, etc.

Utilities

Several other utilities useful when managing your data provider software.

Query Tool

Query this datasource using a generic software that works with any database.

Report a Bug

If you find a bug, please send us a short message.

DataSources

Each data source in a BioCASE service represent a database map (Darwin Core or ABCD). Click on its name to get more information

- ANGSwebdbescoll
- BCPtest
- BFLportal01coll
- BFLportal04coll
- BFLtestcoll
- BIODAlichencoll
- Blettacoll
- BSHerycoll
- BSHfungicoll
- BSHgrassescoll
- BSHlichencoll
- BSHlichfungicoll
- BSHmicrofescoll
- BSHmyxomascoll
- BSHreubert
- BSHschinfcoll
- BSHschmittler
- BSHvplantcoll
- BSHweinstockcoll
- BSHweidcoll
- BSPGcoll
- BSPGpiscescoll
- DiversityIndexing_GRIFFdoppelhaue
- GFBio201900216SNSB
- GFBio201900230SNSB
- GFBio201900204SNSB
- GFBio201900208SNSB
- GFBio202000316SNSB
- GFBio202000340SNSB
- GFBio202000341SNSB
- GFBio202000344SNSB
- GUNcoll
- GOETkavocoll

- GOETvplantcoll
- HALcoll
- HYHIScoll
- IBFfungicoll
- IBFgallicoll
- IBFlichencoll
- IBForthoptercoll
- IBFplantcoll
- JHEfossilcoll
- JHEpiscoll
- JHEfungicoll
- LEmyxcoll
- Hbcoll
- Herphyllcoll
- MSBvplantcoll
- REGvplantcoll
- SAPMmammalcoll
- SAPMpiscescoll
- SMR0langicoll
- SMR0capidcoll
- SMR0capiderstutcoll
- SMNS-E-aransecoll
- SMNS-2-herpcoll
- TUBvplantcoll
- ZSMarthrovaccoll
- ZSMhaystudiescoll
- ZSMhbfcoll
- ZSMpiscescoll

PyWrapper Manual Query Form

Home | Overview | Settings | DB connection | DB structure | Archiving & Filtered export | **QueryForms** | Help | Report a Bug

Debugging: WARNING. Logs at /biocase/log system settings

Wrapper:

PLEASE ENTER SOME BIOCASE PROTOCOL XML

Replace form with templates for a :

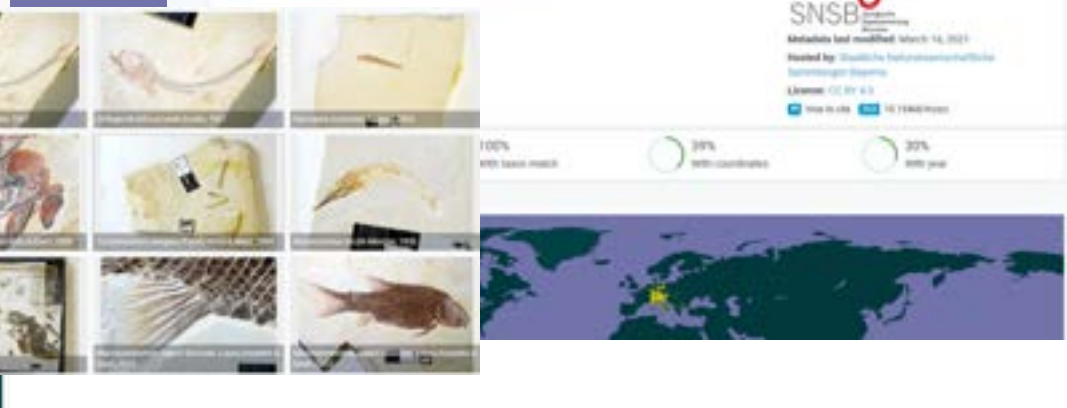
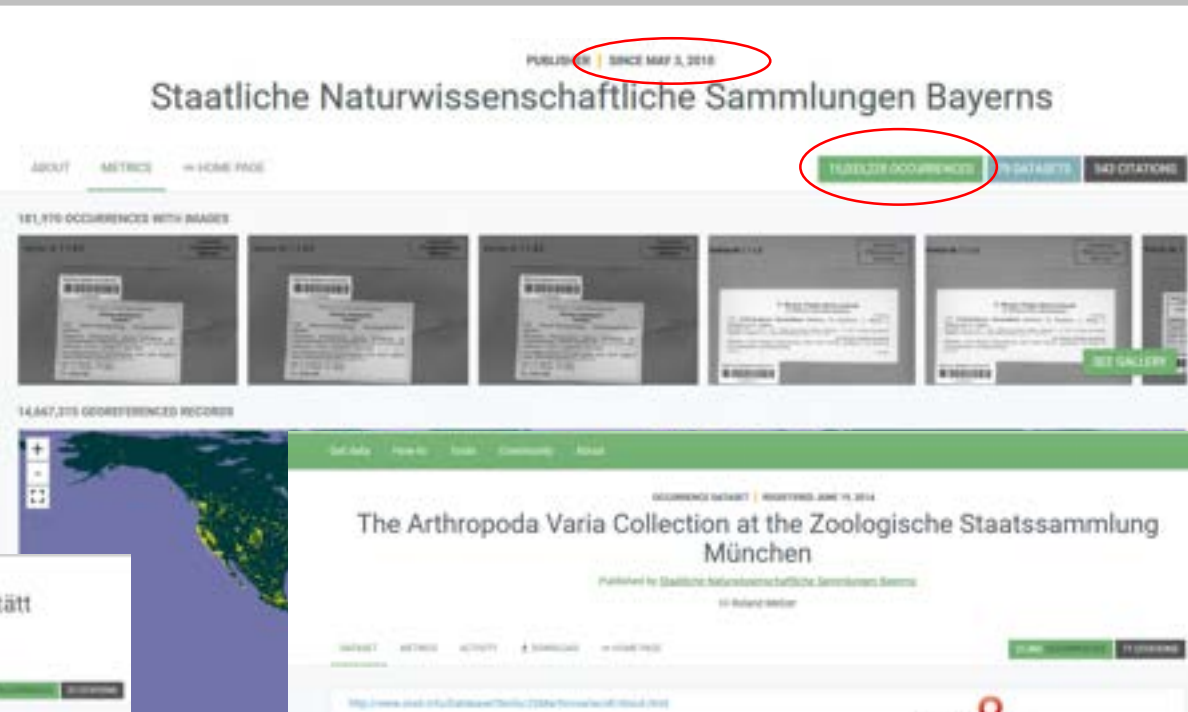
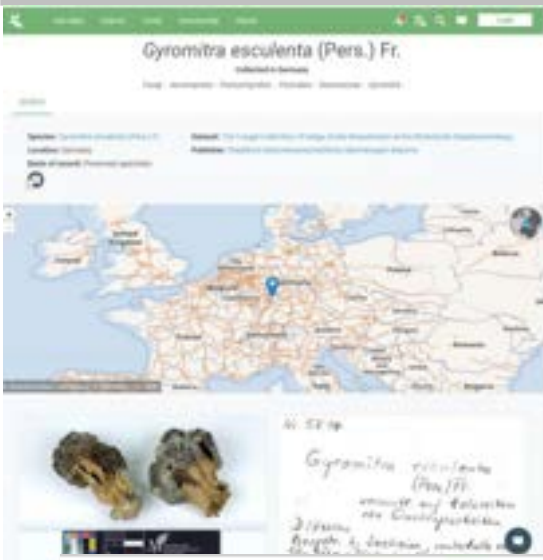
ABCD scan, ABCD search, ABCD2 scan, ABCD2 search, ABCD2.1 scan, ABCD2.1 search, DWK Scan, DWK Search, ABCD2 search, ABCD2 search, TCS 1.01 Scan, TCS 1.01 Search, SPICE-1 Scan, SPICE-1 Search, SPICE-2 Scan, SPICE-2 Search, SPICE-4 Search, SPICE-5 Search, GCP Passport 1.03 Scan, GCP Passport 1.03 Search, GCP Passport 1.04 Scan, GCP Passport 1.04 Search



Powered by: 

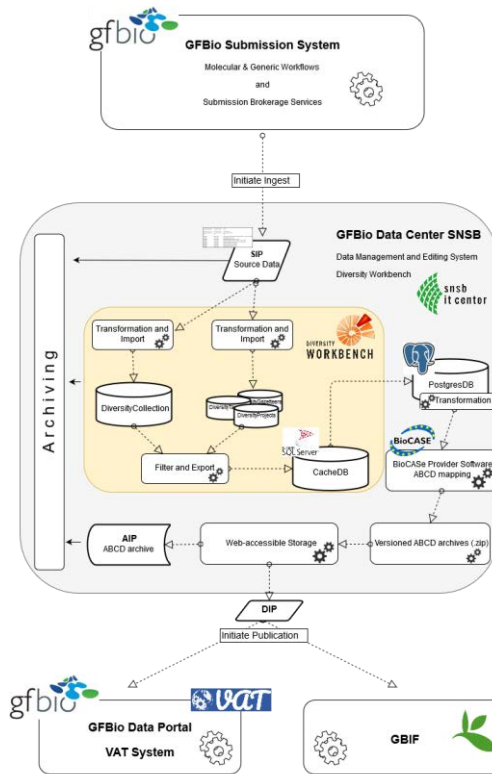
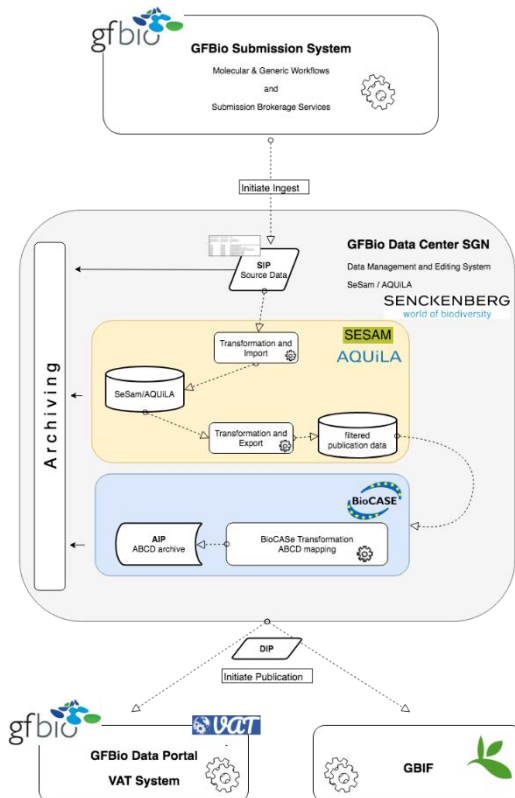
Imprint and Privacy Policy

SNSB: GBIF Data Publisher



Benefit in Using ABCD

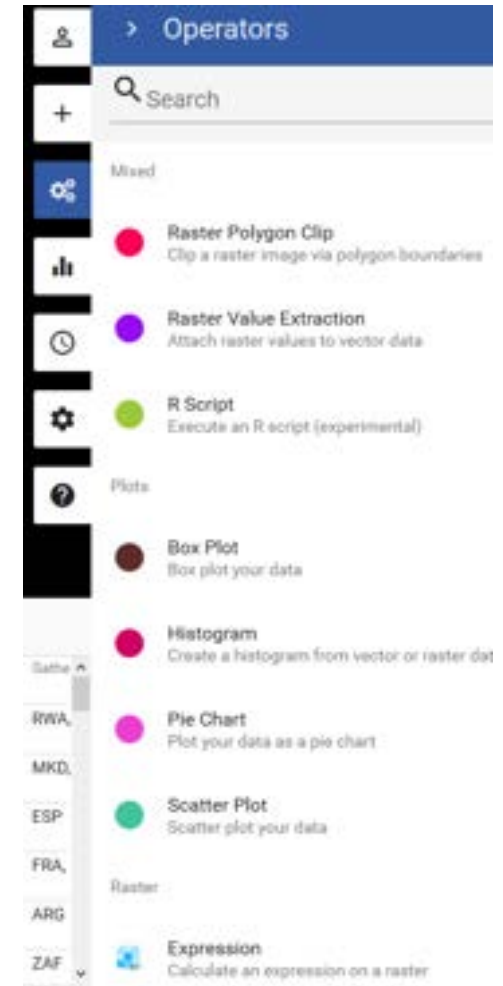
- ABCD structured data in national/ international data portals, free for download
→ GBIF, GFBio and NFDI4BioDiversity
- ABCD data pipelines at all major Natural History Collections in Germany



ABCD Data in GFBio and NFDI



<https://www.gfbio.org/data/visualizeandanalyze>



Anthropology Data in GBIF

Species | ACCEPTED
Homo sapiens Linnaeus, 1758
Published by Linnaeus, C. (1758). Systema Naturae per regna tria naturalia, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. 2380a decima, reformata (7th revised edition), vol. 1: 824 pp. Laurentius Salvus: Holmiae.
source: The Integrated Taxonomic Information System

Merch in German

11,800 OCCURRENCES | 11 REFERENCES

OVERVIEW | TREATMENT | METRICS | REFERENCE TAXON

139 OCCURRENCES WITH IMAGES

2,328 OCCURRENCED RECORDS

Classification

Kingdom: Animalia
Phylum: Chordata
Class: Mammalia
Order: Primates
Family: Hominidae
Genus: Homo Linnaeus, 1758

Species: Homo sapiens Linnaeus, 1758

- Homo aethiopicus Bory de St-Vincent, 1825
- Homo americanus Bory de St-Vincent, 1825
- Homo arabicus Bory de St-Vincent, 1825
- Homo australasicus Bory de St-Vincent, 1825
- Homo cafer Bory de St-Vincent, 1825
- Homo capensis Broom, 1917
- Homo columbianus Bory de St-Vincent, 1825
- Homo cro-magnonensis Gregory, 1921
- Homo denehmii KleinSchmidt, 1931
- Homo floresiensis Brown et al., 2004

SEARCH OCCURRENCES | 139 WITH IMAGES

TABLE | GALLERY | MAP | TAXONOMY | METRICS | DOWNLOAD

- more than 52.000 occurrences
- from major natural history collections, e.g. MfN-Berlin, NHM-London

Anthropology Data in GBIF

OCCURRENCE | 1 JANUARY 1863

Homo sapiens Linnaeus, 1758

Mensch in German Collected in France

Animalia · Chordata · Mammalia · Primates · Hominoidea · Homo

DETAILS

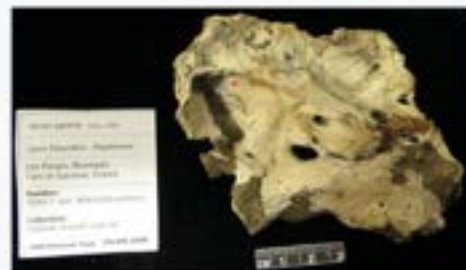
Species: *Homo sapiens* Linnaeus, 1758
 Location: Europe · France
 Basis of record: Fossil specimens
 Dataset: Natural History Museum (London) Collections
 Publisher: Natural History Museum
 Issues: [institution main issue](#) [institution problem](#)



Title: NHMUK_PA_EM3978_1_M_1.jpg
 Record ID: <http://dx.doi.org/10.26434/chem:11461>
 Rights holder: The Trustees of the Natural History Museum, London
 Identifier: <https://www.nhm.ac.uk/collections/preview>
 Suggested attribution: "NHMUK_PA_EM3978_1_M_1.jpg" - *Homo sapiens* Linnaeus, 1758 collected in France by The Trustees of the Natural History Museum, London (licensed under <http://creativecommons.org/licenses/by/4.0/>)



Title: NHMUK_PA_EM3978_2_M_1.jpg
 Record ID: <http://dx.doi.org/10.26434/chem:11462>
 Rights holder: The Trustees of the Natural History Museum, London
 Identifier: <https://www.nhm.ac.uk/collections/preview>
 Suggested attribution: "NHMUK_PA_EM3978_2_M_1.jpg" - *Homo sapiens* Linnaeus, 1758 collected in France by The Trustees of the Natural History Museum, London (licensed under <http://creativecommons.org/licenses/by/4.0/>)



Title: NHMUK_PA_EM_3978

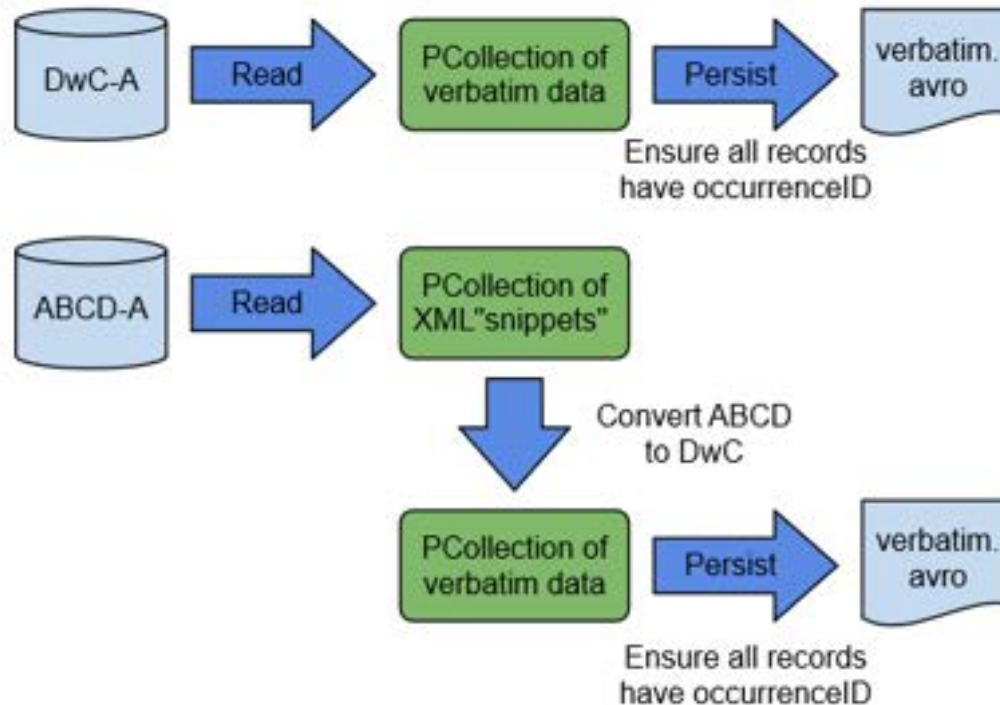
Record

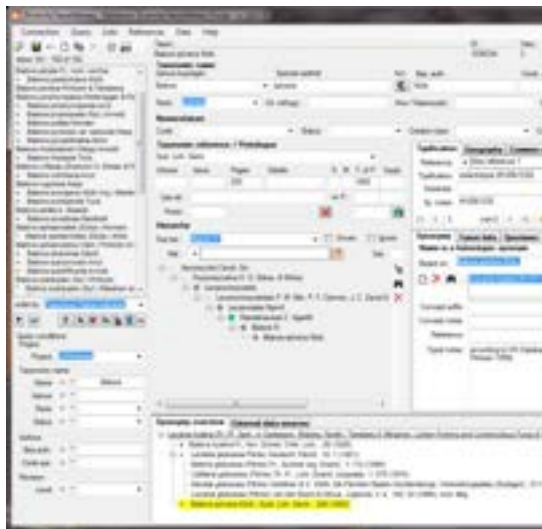
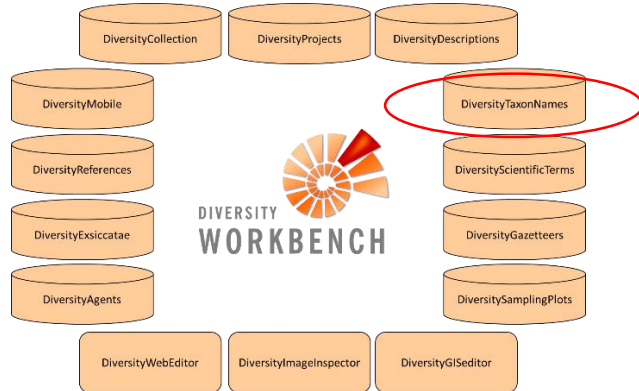
Term	Interpreted	Original	Remarks
Dynamic properties	["chronostratigraphy": "Quaternary, Pleistocene", "cataloguedescription": "Skull, Homo sapiens, Bruniquet Cave, France, Magdalenian period", "gbifissue": "GEOLOGIC_DATUM_ASSUME_D_WG584"], "created": 1380125005000, "donor-name": "F de Lastic", "associatedmediacount": 3, "determinationnames": "Homo sapiens Linnaeus, 1758", "subdepartment": "Anthropology", "gbifid": 1944428234}	["chronostratigraphy": "Quaternary, Pleistocene", "cataloguedescription": "Skull, Homo sapiens, Bruniquet Cave, France, Magdalenian period", "gbifissue": "GEOLOGIC_DATUM_ASSUME_D_WG584"], "created": 1380125005000, "donor-name": "F de Lastic", "associatedmediacount": 3, "determinationnames": "Homo sapiens Linnaeus, 1758", "subdepartment": "Anthropology", "gbifid": 1944428234}	
Institution code	NHMUK Natural History Museum, London	NHMUK	institution main institution problem
Basis of record	Fossil specimen	FossilSpecimen	
Collection code	PAI	PAI	
Occurrence			
Term	Interpreted	Original	Remarks
Catalogue number	PA EM 3978	PA EM 3978	
Occurrence ID	D26a491c-b124-4341-afb4-58a66283c2a	D26a491c-b124-4341-afb4-58a66283c2a	
Occurrence status	PRESENT	present	
Other catalogue numbers	NHMUK:ecatalogue:2827575	NHMUK:ecatalogue:2827575	
Recorded by	F de Lastic	F de Lastic	
Event			
Term	Interpreted	Original	Remarks
Year	1863	1863	
Event date	1863-01-01T00:00:00		Inferred

Ingress

Ingress is from [Darwin Core Archive](#) (zip files of one or more delimited text files) or [ABCD Archives](#) (compressed XML) only[1]. During ingress data is converted from its native format and stored as [Avro](#) files containing Darwin Core compliant data.

This is depicted below:





Regionalised and Domain-specific Taxon Lists

The DWB REST Webservice for Taxon Lists is part of a Diversity Workbench (DWB) services network. It is delivering basic information on taxon names in use, synonyms, classification and German vernacular names of a number of groups of animals, fungi and plants.

The current focus is on domain-specific lists (checklists, taxon reference lists, red lists) from Germany under curation by experts on taxonomy or floristics and faunistics. Each regionalised and domain-specific taxon list has its own history and objectives, is managed completely separately and has its own hierarchical classification. The lists may include additional taxon-related data useful, e. g., for regions.

For more information please check [How to use the DWB REST Webservice for Taxon Lists Services](#).

Logos for gfbio and GBOL are visible in the top right corner.

Regionalised and Domain-specific Taxon Lists

Der DWB REST Webservice für Taxon Lists ist Teil eines Netzwerkes von Diensten, die Informationen über anerkannte Taxonomien, Synonymen, Klassifikationen und deutsche Namen liefern.

Der Fokus liegt dabei zur Zeit auf domän-spezifischen Listen (Checklisten oder floristisch bzw. faunistisch arbeitenden Experten). Jede regionalisierte und domän-spezifische Taxonliste hat ihre eigene Geschichte und Zielvorstellungen, wird unabhängig voneinander verwaltet und hat ihre eigene hierarchische Klassifikation. Die Listen können zusätzliche taxon-relevante Daten enthalten, z. B. für Regionen.

Hinweise zum Einsatz der REST API finden sich unter [How to use the DWB REST Webservice for Taxon Lists Services](#). Weitere Informationen zum Einsatz der REST API finden Sie unter [REST API Documentation](#).

Regionalised and Domain-specific Taxon Lists

For further information and how to use this service please see the [documentation](#).

Overview on Published Lists

Animalia

- Taxon list of Annelida from Germany compiled in the context of the GBOL project
 - TaxRefCode: TaxRef_GBOL_Annelida_DE
- Taxon list of Araneae from Germany compiled in the context of the GBOL project
 - TaxRefCode: TaxRef_GBOL_Araneae_DE
 - Curator:
 - Höfler, Hubert, Dr.
- Taxon list of Bryozoa from Germany compiled in the context of the GBOL project
 - TaxRefCode: TaxRef_GBOL_Bryozoa_DE
- Taxon list of Chilopoda from Germany compiled in the context of the GBOL project
 - TaxRefCode: TaxRef_GBOL_Chilopoda_DE
 - Curators:
 - Speldi, Jörg, Dr.
 - Wiesner, Thomas, Dr.

DwCAHowToGuide

Matt Blissett edited this page on 8 Feb - 58 revisions

Darwin Core Archives – How-to Guide

Version 2.1

Table of Contents

- What is Darwin Core Archive (DwC-A)?
 - DwC-A Components
- DwC-A Data Publishing Solutions
 - Publishing DwC-A using the IPT
 - Registering your Dataset using IPT
 - Publishing DwC-A using GBIF Spreadsheet Templates
 - Publishing DwC-A Manually
- Validation of DwC-As
- Registration of DwC-As with GBIF
- Annex: Preparing Your Data
 - Required and recommended terms
 - Character Encoding
 - Data From a Database
 - DwC-A Examples



INSTALLATION

DiversityTaxonNames Server

Installation type: Http installation
Hosted by: Staatliche Naturwissenschaftliche Sammlungen Bayerns
Registered: November 4, 2015

32 DATASETS

Taxon list of Araneae from Germany compiled in the context of the GBOL project	Checklist dataset
http://www.diversitymobile.net/wiki/About_%22Taxon_list_of_Araneae_from_Germany_compiled_in_the_context_of_the_GBOL_project%22	
1,647 records	
Taxon list of Hymenoptera from Germany compiled in the context of the GBOL project	Checklist dataset
http://www.diversitymobile.net/wiki/About_%22Taxon_list_of_Hymenoptera_from_Germany_compiled_in_the_context_of_the_GBOL_project%22	
11,471 records	
Taxon list of Orthoptera (Grashoppers) from Germany compiled at the SNSB	Checklist dataset
http://www.diversitymobile.net/wiki/About_the_%22Taxon_list_of_Orthoptera_(Grashoppers)_from_Germany_compiled_at_the_SNSB%22	
208 records	



INSTALLATION

DiversityTaxonNames Server

Installation type: Http installation
Hosted by: Staatliche Naturwissenschaftliche Sammlungen Bayerns
Registered: November 4, 2015

CHECKLIST DATASET | REGISTERED METRICS

Taxon list of Jurassic Pisces of the Tethys Palaeo-Environment compiled at the SNSB-JME

Published by Staatliche Naturwissenschaftliche Sammlungen Bayerns

Marin Ebert

1,363 records

DATASET | TAXONOMY | METRICS | DOWNLOAD | HOME PAGE

http://www.diversitymobile.net/wiki/About_Taxon_List_of_Jurassic_Pisces_of_the_Tethys_Palaeo-Environment_compiled_at_the_SNSB-JME

Metadata last updated: 2015-11-04

Hosted by: Staatliche Naturwissenschaftliche Sammlungen Bayerns

License: CC BY

How to cite

736 Accepted names

64 Synonyms

72% Overlap with GBIF Backbone

21 Overlap with Catalogue of Life

Description

http://www.diversitymobile.net/wiki/About_Taxon_List_of_Jurassic_Pisces_of_the_Tethys_Palaeo-Environment_compiled_at_the_SNSB-JME



Taxon list of Jurassic Pisces of the Tethys Palaeo-Environment compiled at the SNSB-JME

Published by Staatliche Naturwissenschaftliche Sammlungen Bayerns

ACCEPTED NAMES | SYNONYMS | METRICS | DOWNLOAD | HOME PAGE

1,363 records

OVERLAP METRICS

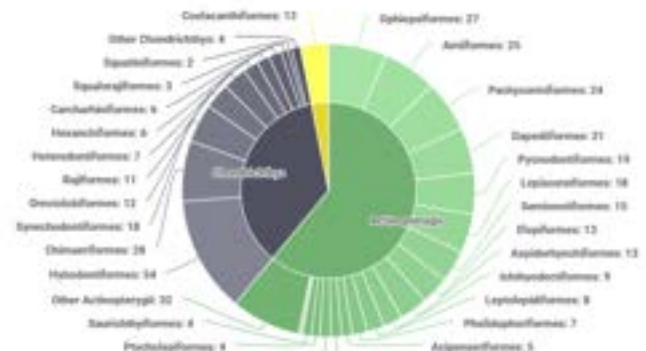
737 Accepted names

651 Synonyms

93% Overlap with GBIF Backbone

1% Overlap with Catalogue of Life

NUMBER OF ACCEPTED SPECIES BY HIGHER TAXON





GBIF Tools

GBIF Tools

An index to tools brought to you by the Global Biodiversity Facility:

[GBIF Excel Templates](#) [Darwin Core Archive Assistant](#) [Darwin Core Archive Validator](#) [Name Parser](#)

GBIF Excel Templates

The GBIF Excel Templates are MS Excel spreadsheets that support biodiversity data entry in a standardized format. There is one template for each of the three classes of biodiversity data:

1. [Checklist Data](#)
2. [Occurrence Data](#)
3. [Sampling Event Data](#)

To publish the data through the GBIF network, upload the templates to the GBIF Integrated Publishing Toolkit (IPT). To enter dataset metadata, use the IPT's built-in metadata editor. If you require an account on an IPT, it is highly recommended that you save yourself time and money by requesting an account on a [trusted data hosting centre](#) located in your country. If you need assistance, please contact the [GBIF Helpdesk](#) for assistance.

Darwin Core Archive Assistant

The Darwin Core Archive Assistant is a web application that presents a simple interface for describing the data elements a data publisher wishes to serve to the GBIF network as basic text files and composes the appropriate XML descriptor file as defined in the Darwin Core Text Guidelines to accompany them. It communicates with the GBIF registry to provide an up-to-date listing of all relevant Darwin Core terms and available extensions and presents these in a simple checklist format.

The [Darwin Core](#) is a body of standards that include a set of terms relating to taxa and their occurrence in nature, and a set of practices regarding the use of these terms in the publication of biodiversity data and information. GBIF has adopted a text-based solution for using Darwin Core that both simplifies and extends the publication of species and species-occurrence data. This format is referred to as a Darwin Core Archive (DWCA) and provides a relatively non-technical option for publishing biodiversity data that does not require complicated installations of data publication software. Darwin Core Archives can be published via a simple web address or URL.

Darwin Core Archives support the publication of enriched data types that extend the core terms while retaining the relatively simple, text-based data format. These extensions, however, require the inclusion of an XML descriptor file (meta.xml) that serves as a map to the different files and data elements in the archive. Many biologists and data managers find working with XML challenging while otherwise finding the technical threshold for producing Darwin Core Archives quite low.

Darwin Core Archive Validator

The validator is a tool to test Darwin Core Archives as specified in the [Darwin Core Text Guidelines](#). Due to the simplicity of the archives GBIF encourages publishers to create them using simple custom scripts. Therefore the need arises to provide a testing framework for developers to make sure GBIF and others can read the information as expected.

The validator uses the official XML schema to validate the meta.xml descriptor, but additionally it uses the [Darwin Core Archive Reader](#) java library to validate the content against the known extensions and terms registered within the GBIF network for sharing biodiversity data. GBIF runs a production and a development registry that keeps track of extensions, both of which are used by this validator.

GBIF recommends to bundle an [Ecological Metadata Language \(EML\)](#) xml file with an archive. As EML is a rather large and complex schema GBIF has specified a [GBIF profile](#) that uses a subset of EML 2.1.1 and also declares specific additions to EML within the generic additionalMetadata section of EML. Every valid GBIF profile document should therefore always be valid according to the official EML schema. The EML validation is done according to these two xml schemas.

.... become part of the GBIF Network

Get data | How to | Tools | Community | About

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES | SPECIES | DATASETS | PUBLISHERS | RESOURCES

WHAT IS GBIF? | ABOUT GBIF GERMANY

Subergorgia suberosa observed in the Persian Gulf near Bahrain by robin_slwe_diva (CC BY-NC 4.0)

Occurrence records 1.878.260.089	Datasets 60.354	Publishing institutions 1.695
--	---------------------------	---

Peer-reviewed papers using data
5.882

GBIF forecast: increasing chance of clouds for species occurrence data

BID programme funds 18 new projects in sub-Saharan Africa

Using GBIF-mediated data with Apache Spark on Amazon Web Services