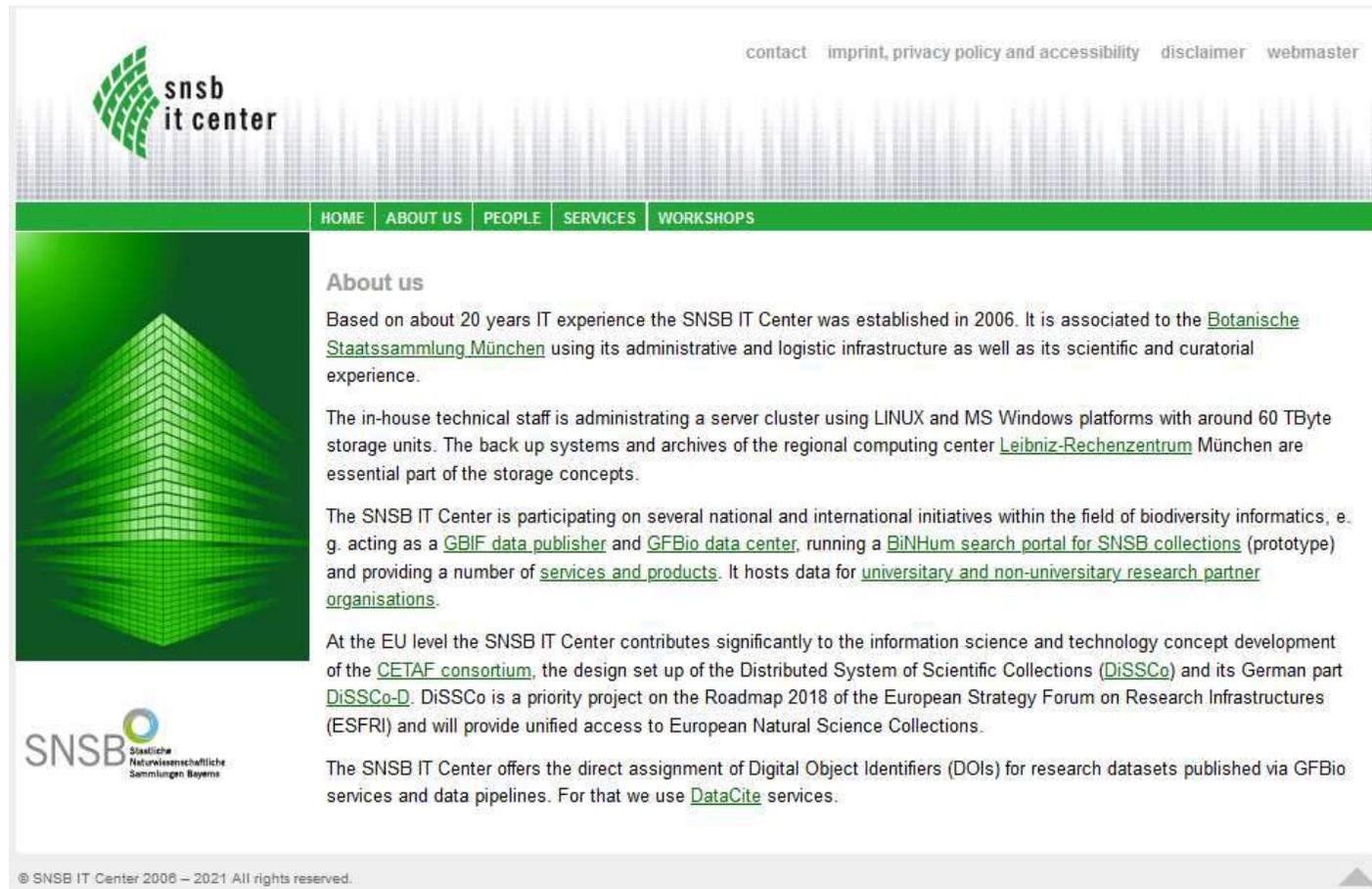


# The SNSB data pipelines for publishing occurrence data via GBIF are appropriate for human osteological collections

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 is titled 'About us' and contains three paragraphs of text. The first paragraph states that the SNSB IT Center was established in 2006 and is associated with the Botanische Staatssammlung München. The second paragraph describes the in-house technical staff and their infrastructure. The third paragraph details the center's participation in national and international initiatives. At the bottom left of the content area is the SNSB logo. At the bottom of the page 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](#). It hosts data for [university and non-university research partner organisations](#).

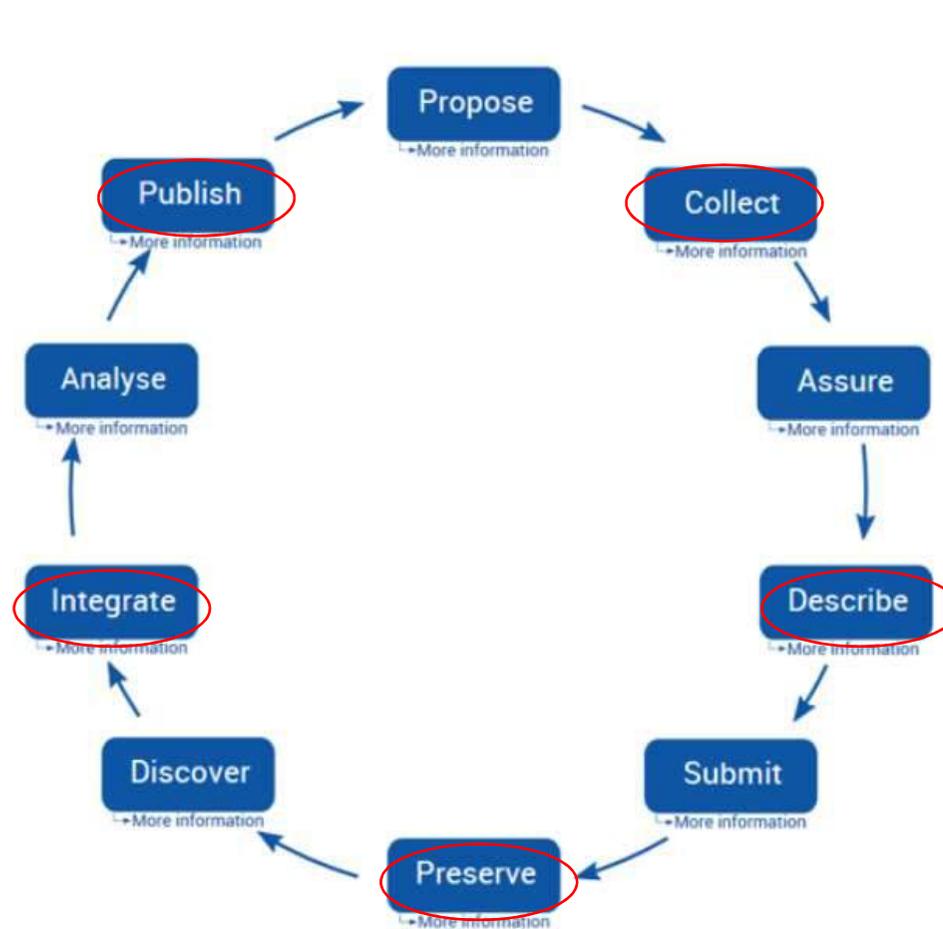
At the EU level the SNSB IT Center contributes significantly to the information science and technology concept development of the [CETAF consortium](#), the design set up of the Distributed System of Scientific Collections ([DiSSCo](#)) and its German part [DiSSCo-D](#). DiSSCo is a priority project 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 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.

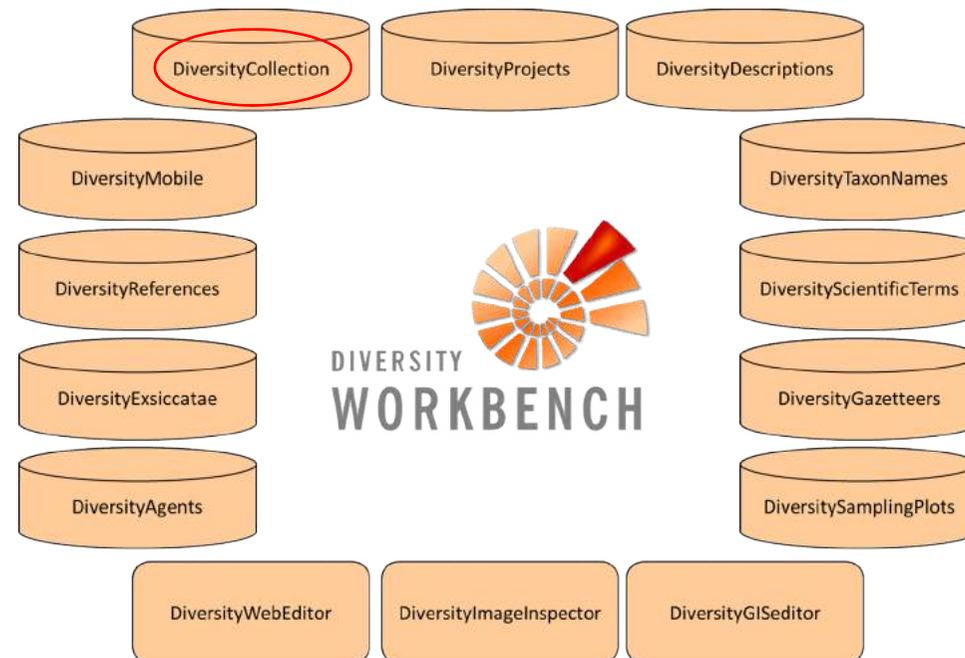
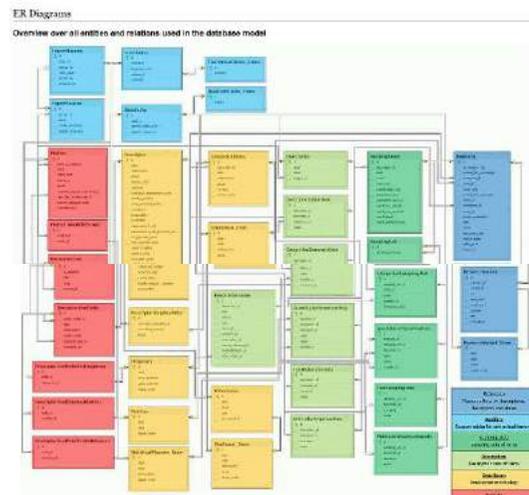


- Data management of all repositories is done in own installations of DWB databases
- Administration of 200 DWB accounts
- For collection and occurrence data: DiversityCollection

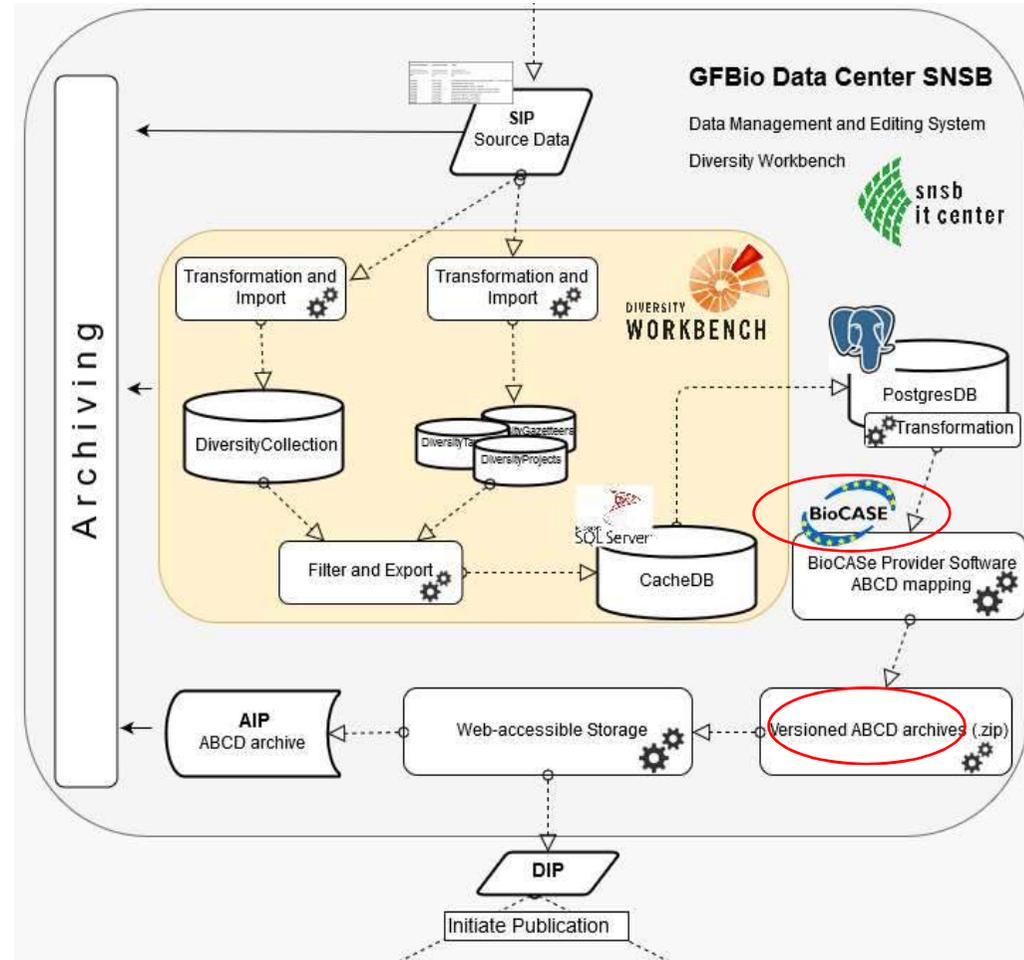
DiversityDescriptions Information Model (version 3.0.15, 11 July 2016)

Author:	In: Hagedorn, A., Pflanz, A., Lenz, G., Sperber, K. D., Thaler, D. (2016)
License:	CC-BY-NC-ND
Organization:	G. Hagedorn, A. Pflanz, A. Lenz, G. Sperber, K. D., Thaler, D. (2016). DiversityDescriptions Information Model (Version 3.0.15). <a href="https://www.snsb.de/dwb/descriptions">https://www.snsb.de/dwb/descriptions</a>
Notes:	This model is currently under development. It is not yet ready for use in production. It is not yet ready for use in production. It is not yet ready for use in production. It is not yet ready for use in production.

iVivo Database Scheme  
This information model is available as iVivo database scheme with both single database and data query reference tables or conceptual to view table and personal URL.



# 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

technical specification 2005 standard 2005

Home / 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

#### 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 high supporting data from a wide variety of databases: several existing data standards. Parallel structure both) atomised data and free-text can be accessed and 2.06 are currently in use with the GBIF (Global Information Facility) and BioCAsE (Biological Collections for Europe) networks. Apart from the GBIF and potential for the application of ABCD extends to in-house legacy data access (e.g. datasets from shall not be converted and integrated into an i be kept separately, though easily accessible). Between terms, ABCD is a step towards an ontology collections.

#### Creator

Access to Biological Collections Data task group  
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>

## 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

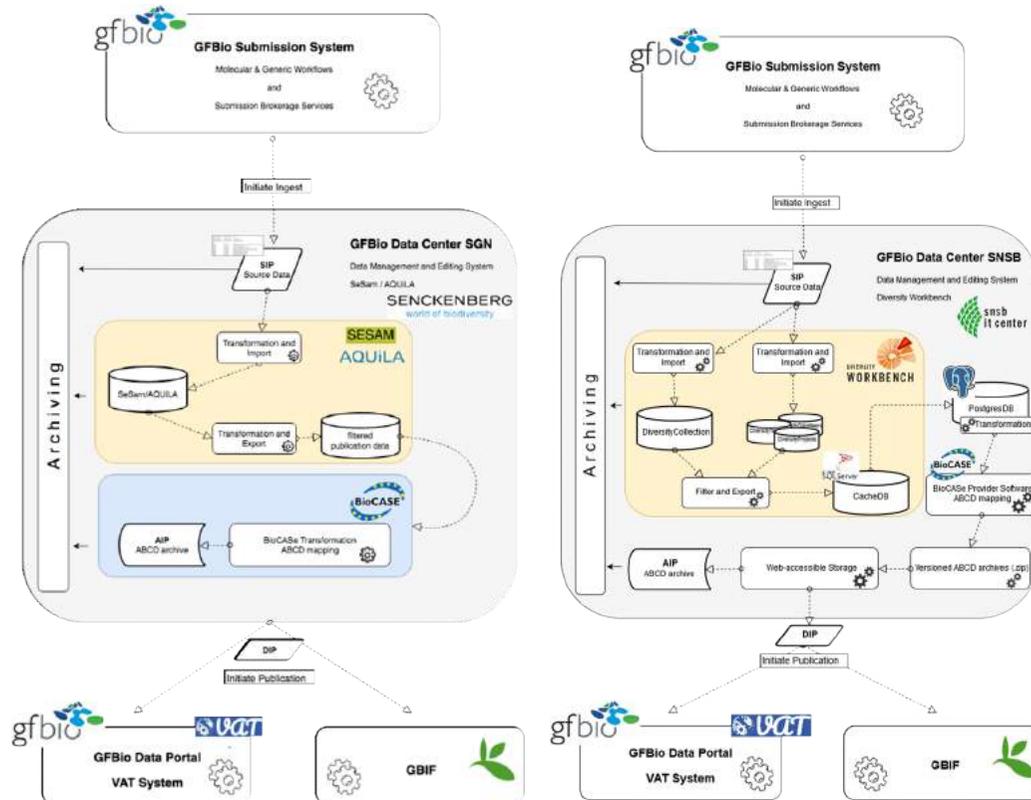
The ABCD Standard is now described as an XML Schema and an Ontology. This allows the access of the standard through semantic queries, encourages element reuse and serves as basis for future software and services in the area of semantic web. The XML Schema is linked to the Ontology via SAWSDL Annotations.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





# 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>



**> Operators**

- Search**
- Mixed**
  - Raster Polygon Clip**  
Clip a raster image via polygon boundaries
  - Raster Value Extraction**  
Attach raster values to vector data
  - R Script**  
Execute an R script (experimental)
- Plots**
  - Box Plot**  
Box plot your data
  - Histogram**  
Create a histogram from vector or raster data
  - Pie Chart**  
Plot your data as a pie chart
  - Scatter Plot**  
Scatter plot your data
- Raster**
  - Expression**  
Calculate an expression on a raster

# Anthropology Data in GBIF

The screenshot shows the GBIF species page for *Homo sapiens* Linnaeus, 1758. The page is divided into several sections:

- Classification:** A vertical sidebar on the right lists the taxonomic hierarchy: Kingdom: Animalia, Phylum: Chordata, Class: Mammalia, Order: Primates, Family: Hominidae, Genus: *Homo* Linnaeus, 1758.
- Species List:** Below the classification, a list of subspecies is provided, including *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 columbicus* Bory de St. Vincent, 1825, *Homo cro-magnonensis* Gregory, 1921, *Homo drennani* Kleinschmidt, 1931, and *Homo floresiensis* Brown et al., 2004.
- Species Information:** The main content area displays the species name, its publication details (Linnaeus, C. (1758). *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata* [10th revised edition], vol. 1: 824 pp. Laurentius Salvius: Holmiae. source: The Integrated Taxonomic Information System), and its German name "Mensch".
- Occurrences:** A green bar indicates "52,985 OCCURRENCES" and "11 INFRASPECIES".
- Georeferenced Records:** A world map shows the global distribution of records, with a concentration in Europe and Africa.
- Images:** A row of four images shows various human specimens, including a skull, a skeleton, and a person.

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

The screenshot shows the GBIF search results page for *Homo sapiens* Linnaeus, 1758. The page is divided into several sections:

- Search Occurrences:** A green bar at the top indicates "139 WITH IMAGES".
- Navigation:** A horizontal bar contains tabs for TABLE, GALLERY, MAP, TAXONOMY, METRICS, and DOWNLOAD.
- Image Gallery:** A grid of 40 small images shows various human specimens, including skulls, skeletons, and bones. Each image is labeled with the species name and the year 1758.
- Filters:** A green bar at the bottom right contains a "FILTERS" button.

# Anthropology Data in GBIF

OCCURRENCE | 1 JANUARY 1863

## Homo sapiens Linnaeus, 1758

Mensch In German Collected in France

Animalia > Chordata > Mammalia > Primates > Homnidae > Homo

DETAILS

Species: *Homo sapiens* Linnaeus, 1758  
 Location: Europe > France  
 Basis of record: Fossil specimen  
 Dataset: Natural History Museum (London) Colli  
 Publisher: Natural History Museum  
 Issues: [institution match fuzzy](#) [instrubon collectio](#)



Title: NHM-UK\_PA\_EM3978\_1\_M\_1.jpg  
 Record license: <http://creativecommons.org/licenses/by/4.0/>  
 Rights holder: The Trustees of the Natural History Museum, London  
 Identifier: <https://www.nhm.ac.uk/content/preview>  
 Suggested attribution: 'NHM-UK\_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: NHM-UK\_PA\_EM3978\_2\_M\_1.jpg  
 Record license: <http://creativecommons.org/licenses/by/4.0/>  
 Rights holder: The Trustees of the Natural History Museum, Lond  
 Identifier: <https://www.nhm.ac.uk/content/preview>  
 Suggested attribution: 'NHM-UK\_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

Navigation icons: Home, Search, Filter, List, Menu

### Record

Term	Interpreted	Original	Remarks
Dynamic properties	["(chronostratigraphy": "Quaternary, Pleistocene", "cataloguedescription": "Skull. Homo sapiens. Bruniquel Cave, France. Magdalenian period.", "gbfissue": "[\"GEODETTIC_DATUM_ASSUMED_WGS84\", \"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. Bruniquel Cave, France. Magdalenian period.", "gbfissue": "[\"GEODETTIC_DATUM_ASSUMED_WGS84\", \"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	<a href="#">institution match</a> <a href="#">institution collectio</a>
Basis of record	Fossil specimen	FossilSpecimen	
Collection code	PAL	PAL	

### Occurrence

Term	Interpreted	Original	Remarks
Catalogue number	PA EM 3978	PA EM 3978	
Occurrence ID	f26a491c-b124-4341-afb4-58a666283c2a	f26a491c-b124-4341-afb4-58a666283c2a	
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

# SNSB-SAPM Osteological Data in GBIF

OCCURRENCE DATASET | REGISTERED MAY 5, 2014

## The Pisces Collection at the Staatssammlung für Anthropologie und Paläoanatomie München

Published by Staatliche Naturwissenschaftliche Sammlungen Bayerns

OCCURRENCE DATASET | REGISTERED OCTOBER 29, 2010

## The Mammalia Collection at the Staatssammlung für Anthropologie und Paläoanatomie München

Published by Staatliche Naturwissenschaftliche Sammlungen Bayerns

DATASET METRICS ACTIVITY DOWNLOAD

<http://www.snsb.info/DatabaseClients/SAPMpisc>

4,096 Occurrences

2,544 GEOREFERENCED RECORDS



DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

<http://www.snsb.info/DatabaseClients/SAPMmammaliacol/about.jsp>

3,047 Occurrences 100% With taxon match

45 GEOREFERENCED RECORDS



Occurrences

Search all fields

Simple Advanced

Occurrence status

License

Scientific name

Basis of record

Location

Administrative areas (gadm.org)

Coordinate uncertainty in meters

Year

Month

Dataset

The Pisces Collection at the Staatssammlung für Anthropologie und Paläoanatomie München

Country or area

Continent

Issues and flags

Media type

Publisher

Institution code

SEARCH OCCURRENCES | 3,920 WITH IMAGES

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD



# GBIF Pipelines for Occurrence Data

## 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:

