

Sammlungsdaten am ZFMK

Data Exchange and Facilitation of Work Processes through Diversity Collection

Peter Grobe

Zoological Research Museum Alexander Koenig





Components

Management and Storage of Collection Data

- Collection/Occurrence/Field data: DiversityWorkbench
- Media: easyDB
- Sequences: Geneious

Publication/Sharing:

- Collection Portal: collections.zfmk.de
- German Barcode of Life: bolgermany.de
- BioCAsE: biocase.zfmk.de/biocase
- GFBio: gfbio.org



Requirements

Must haves:

- Occurrences, person names, objects, taxonomic assignments

Additional:

- Methods, relations, history, measurement, multimedia, identifier

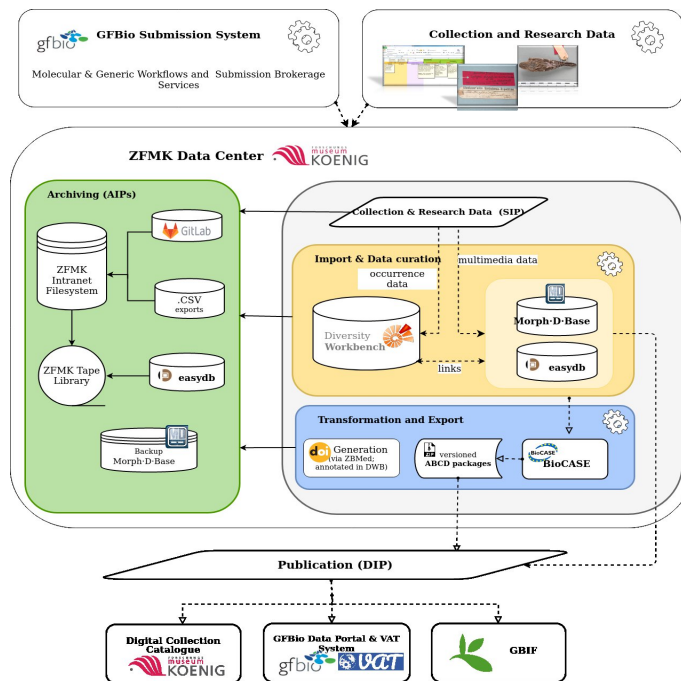
Metadata:

- Provenance, data sets, licenses, access rights, moratoria

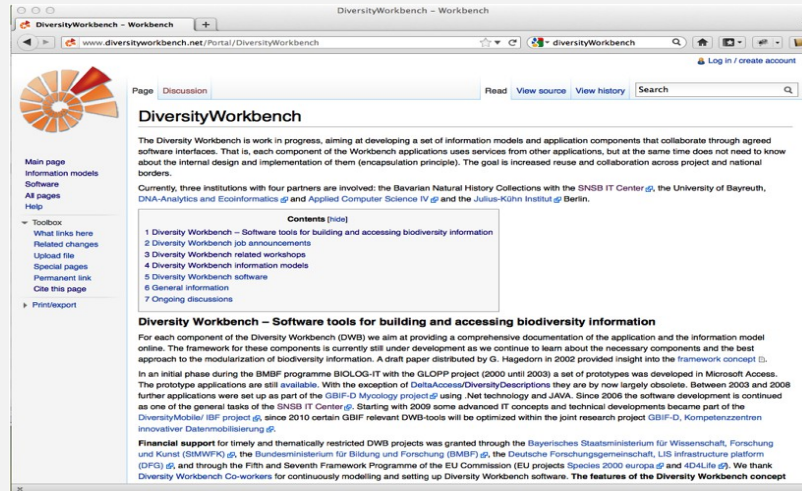


Benefits

- Easy sharing of data with other researchers
- Long term availability of structured data
- Publishing



Diversity Workbench at ZFMK



The screenshot shows the DiversityWorkbench website. The main heading is "DiversityWorkbench". Below it, a paragraph states: "The Diversity Workbench is work in progress, aiming at developing a set of information models and application components that collaborate through agreed software interfaces. That is, each component of the Workbench applications uses services from other applications, but at the same time does not need to know about the internal design and implementation of them (encapsulation principle). The goal is increased reuse and collaboration across project and national borders." It then lists the current partners: "Currently, three institutions with four partners are involved: the Bavarian Natural History Collections with the SNSB IT Center [@](#), the University of Bayreuth, DNA-Analytics and Ecoinformatics [@](#) and Applied Computer Science IV [@](#) and the Julius-Kühn Institut [@](#) Berlin."

A "Contents [hide]" section lists the following items:

- 1 Diversity Workbench – Software tools for building and accessing biodiversity information
- 2 Diversity Workbench job announcements
- 3 Diversity Workbench related workshops
- 4 Diversity Workbench information models
- 5 Diversity Workbench software
- 6 General information
- 7 Ongoing discussions

Below this is the section "Diversity Workbench – Software tools for building and accessing biodiversity information". It explains the goal: "For each component of the Diversity Workbench (DWB) we aim at providing a comprehensive documentation of the application and the information model online. The framework for these components is currently still under development as we continue to learn about the necessary components and the best approach to the modularization of biodiversity information. A draft paper distributed by G. Hagedorn in 2002 provided insight into the framework concept." It mentions the initial phase during the BMBF programme BIOLOG-IT with the GLOPP project (2000 until 2003) and the development of prototypes in Microsoft Access. It notes that prototype applications are still available but largely obsolete between 2003 and 2008, and that further applications were set up as part of the GIBIF-D Mycology project [@](#) using .Net technology and JAVA. Since 2006, the software development is continued as one of the general tasks of the SNSB IT Center [@](#). Starting with 2009, some advanced IT concepts and technical developments became part of the Diversity/Mobile/IBF project [@](#), since 2010 certain GIBIF-relevant DWB-tools will be optimized within the joint research project GIBIF-D, Kompetenzzentren innovativer Datenmobilisierung [@](#).

The "Financial support" section states: "Financial support for timely and thematically restricted DWB projects was granted through the Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst (StMWFK) [@](#), the Bundesministerium für Bildung und Forschung (BMBF) [@](#), the Deutsche Forschungsgemeinschaft, LIS infrastructure platform (DFG) [@](#), and through the Fifth and Seventh Framework Programme of the EU Commission (EU projects Species 2000 europa [@](#) and 4D4Life [@](#)). We thank Diversity Workbench Co-workers for continuously modelling and setting up Diversity Workbench software. The features of the Diversity Workbench concept

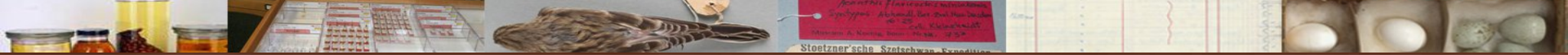
- Since 2011 at ZFMK
- Central CMS
- Adaptation to different requirements
- Mapping of work flows to data:
 - in the collection
 - in the laboratory
 - in projects
- Data exchange between collections





Features (important for us)

- Data management via relational database and 170 tables (!)
- Easy expandability
- Integrated versioning and management of data history
- Flexible quality assurance mechanisms, search algorithms
- Integrated upload and download functions
- Integration of generic services (standard thesauri, gazetteers, GIS)
- Transactions



Numbers

1,165,481 Collection Specimens

1,020,653 Organisms

1,169,303 Identifications

220,203 Analyses

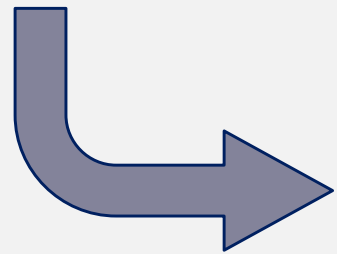
4,583,828 Parameter

- Nearly all features of Div. Coll. in use
- Live projects (FOGS, AMMOD, ...)
- Deeply nested data
- Management of
 - Insect drawers,
 - Malaise bottles,
 - Single objects,
 - Barcode sequences,
 - Biobanking
- ... and relations between all



Forensic Genetics for Species Protection: Analysis for DNA extraction

Species	Locus	Amp size	Repeat motif	Ref allele	NCBI No	Scaffold ID	NGS Val	Wanderfalken (n=6)						Anteil*	inheritance	Allele Nr.					
								Haplotype													
	Fa-9	369	GATG	9, 5	GCA_000	scaffold9_1		35 C->T	58 A->G	85 C->A	89 C->T	Msat 126	145 A->G	Msat 158	158 DEL	159 DEL	161 T->A				
							Peregrine falcon (n=6)	35T	-	85A	-	Al9-Ref	145G	Al5-Ref	-	-	-	-	30%		
								-	-	-	-	Al9-Ref	-	Al5-Ref	-	-	-	-	4%		
								-	-	-	-	Al9-Ref	-	Al5-Ref	-	DEL-159	-	-	4%		
			(GATG) ₉ G					35T	-	85A	-	10.1	-	Al5-Ref	-	-	-	-	28%		
			(GATG) ₈ G					35T	-	85A	-	12.1	-	Al5-Ref	DEL-158	-	-	-	11%		
Falco peregrinus	Fa-9	369	(GATG) ₉ G	9, 5	GCA_000	scaffold9_1	132-1431	T	-	A	-	11.1	-	5	-	-	-	-	42%	he	11b
Falco peregrinus	Fa-9	369	(GATG) ₈ G	9, 5	GCA_000	scaffold9_1		T	-	A	-	12.1	-	5	GAGT	-	-	-	29%	he	12a
Falco peregrinus	Fa-9	369	GATG	9, 5	GCA_000	scaffold9_1	133-1432	T	G	A	-	8	-	DEL-(GATG)	-	-	A	-	41%	he	8a
Falco peregrinus	Fa-9	369	GATG	9, 5	GCA_000	scaffold9_1		T	-	A	-	9	-	4.1	GAGT	-	-	-	29%	he	9c



DiversityCollection, Database: DiversityCollection_Fiddel v. 4.3.131

Connection Grid Query Data Administration Help

Acc. Nr. **Falco peregrinus** ID: 224762724

Specimen Event Version Withhold. 2 Withhold by default

ZFMK-DNA-FD14 FOGS_Result_Inheritance (2020-11-20 14:45:52) Fa_6-1 - 132-1431: Homocygote

Nr. of an.: Fa_6-1 - 132-1431 Result: Homocygote

URI:

Methods

FOGS_Locus Fogs_locus-5de433e08b6a4bb2c882 inheritance

FOGS_Mutations Fogs_mutation-1113fcb5a97db3cfd66a labid

FOGS_Mutations Fogs_mutation-4e0e17b6e461a0a1b34 locus

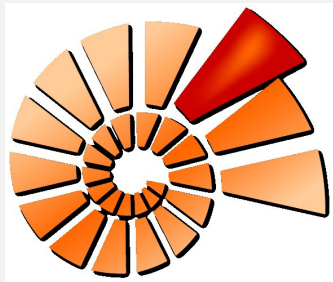
repeat-motif

Cre.dat.: 2020-10-28 by DOMZFMK\vdmark

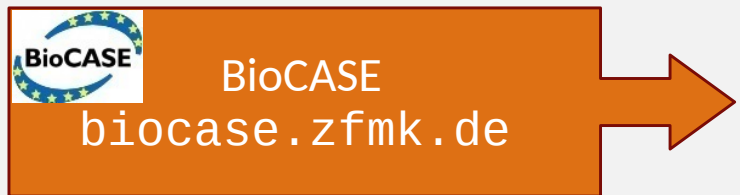
Last changes: 2020-11-20 by DOMZFMK\vdmark



Collection Data, Dataflow

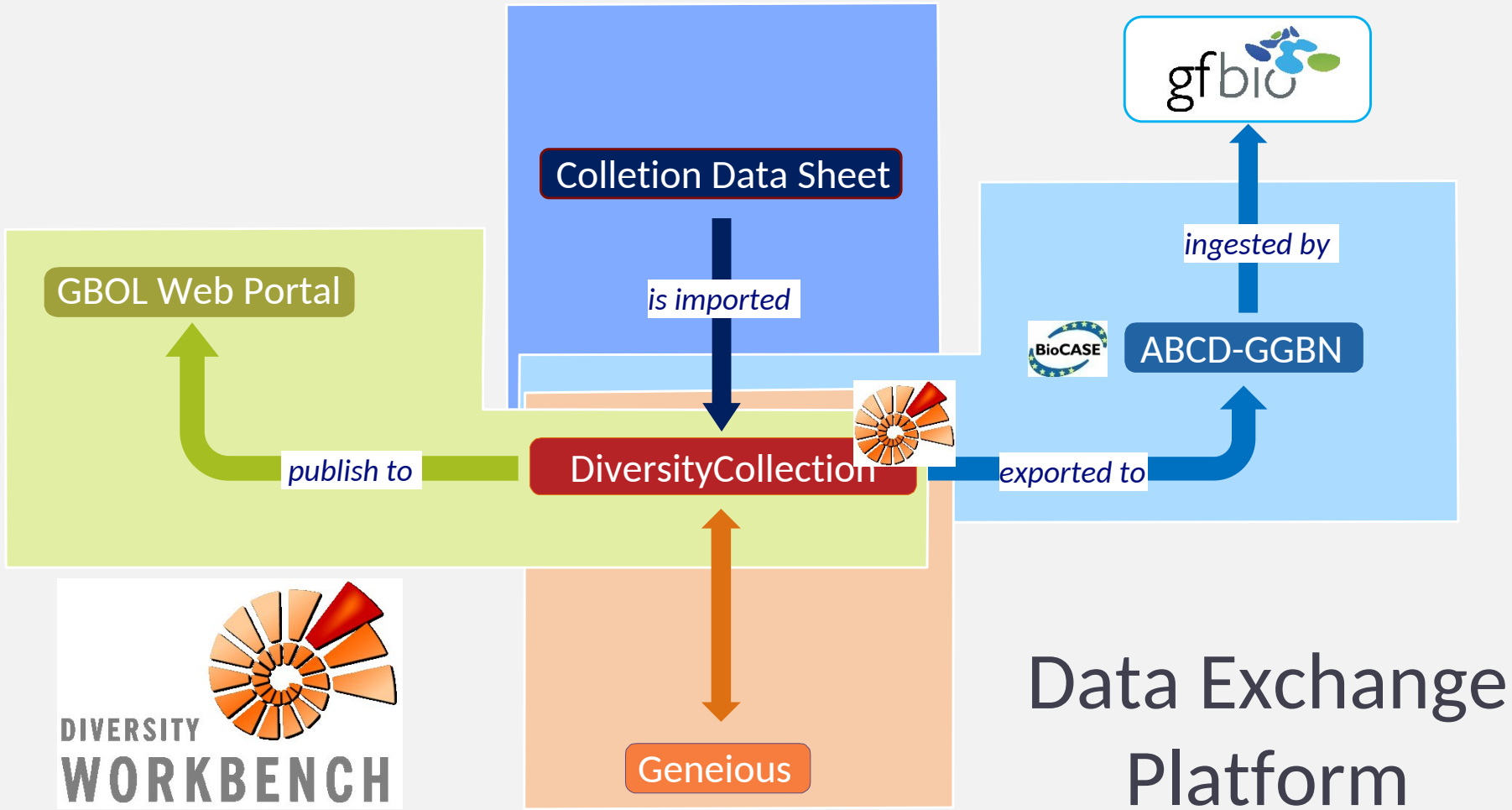


API, e.g.
https://id.zfmk.de/collection_ZFMK/

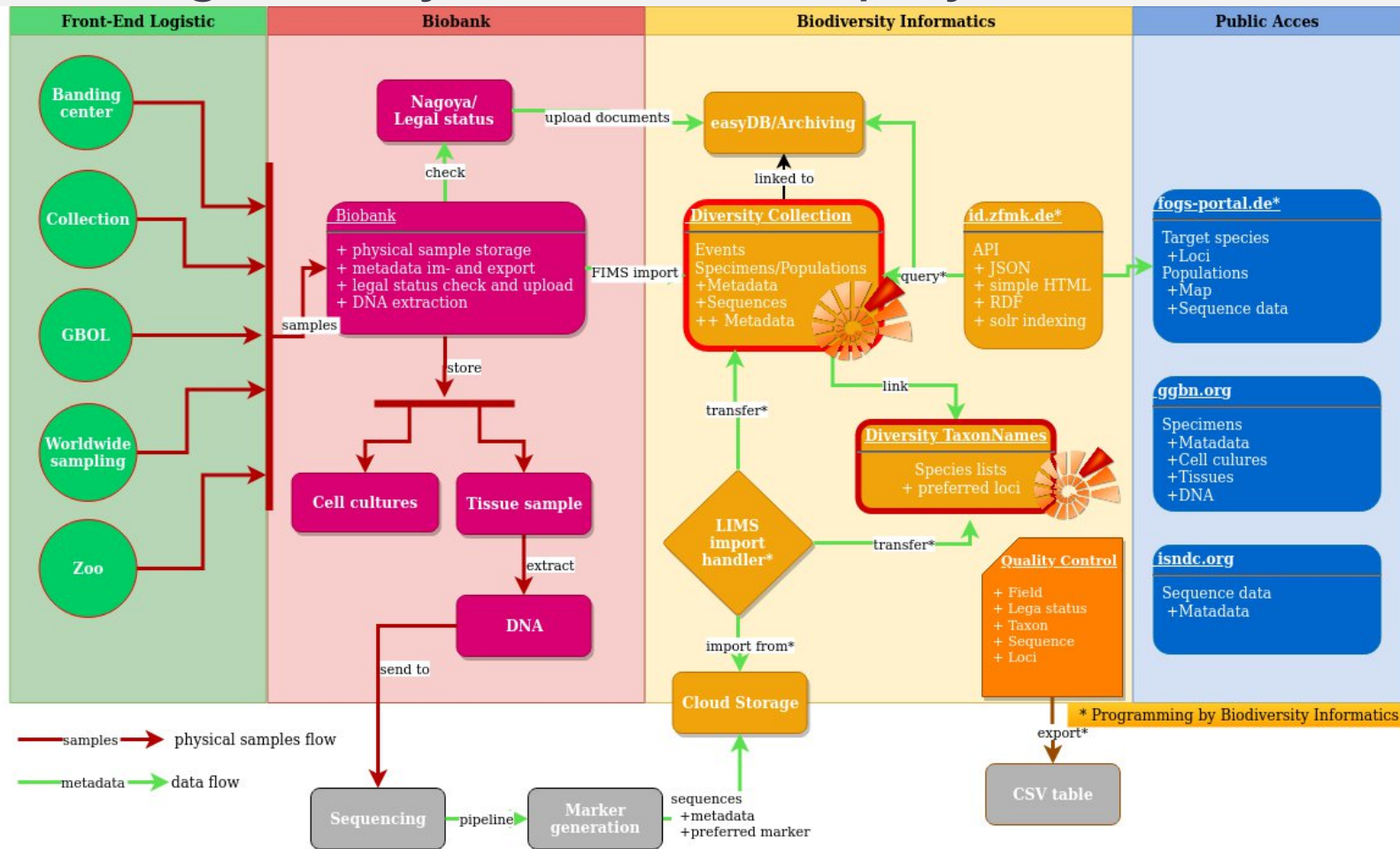


*

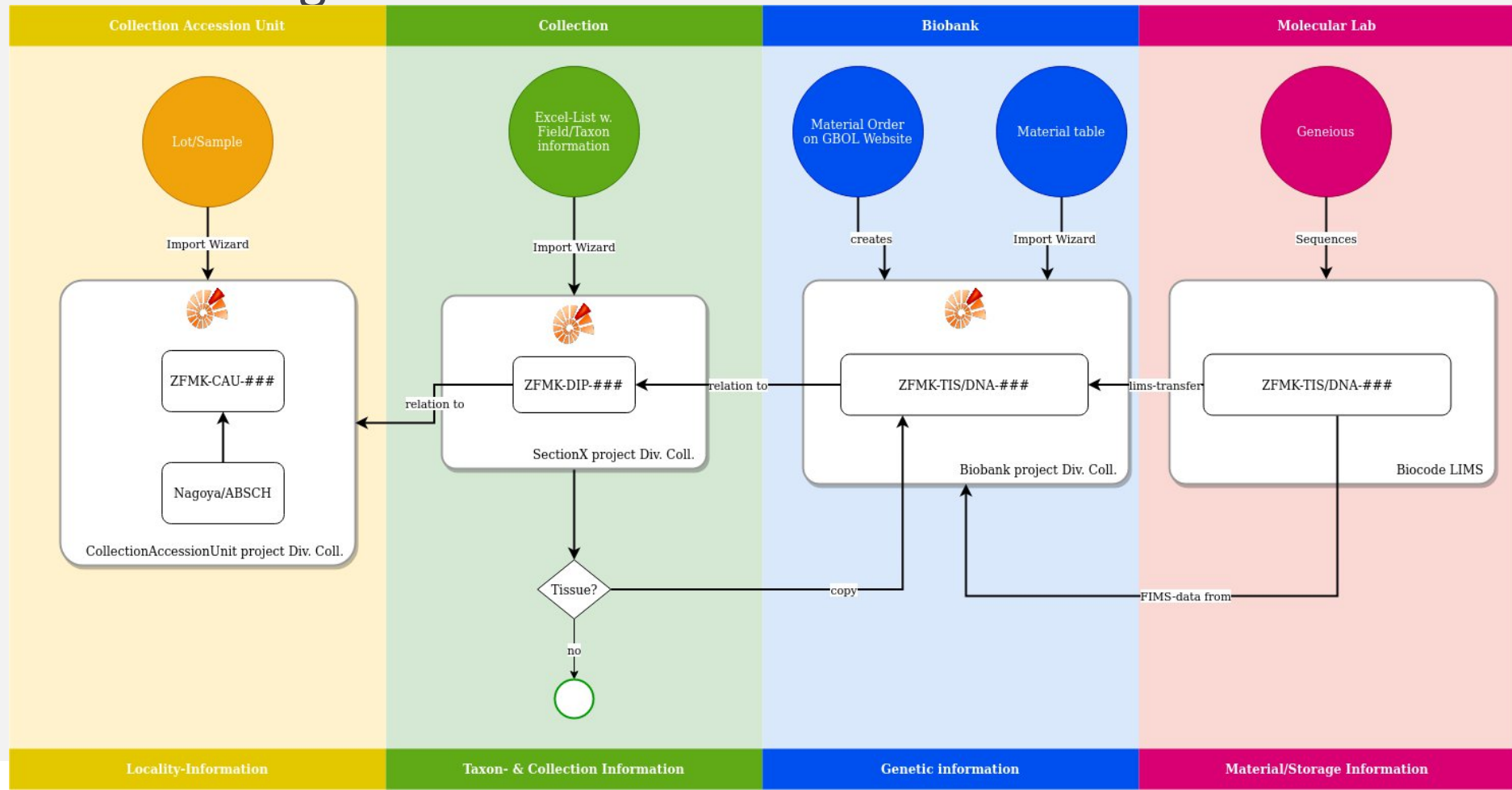
Transfer or Export: https://datacenter.zfmk.de/gitlab/BioCASE/biocase_media

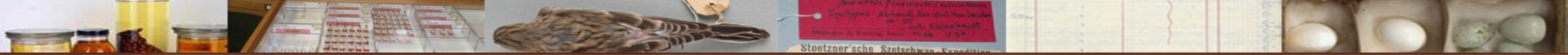


Data Exchange in Projects: The FOGS project



Data Exchange between Collections





Data Exchange between Collections



- Involved sections:
 - Biobank (GBOL)
 - Laboratory
 - Morphological Collection

The screenshot displays the GBOL database interface for specimen ZFMK-115-10605. The main view shows a photograph of the specimen, a tree view of the collection hierarchy, and a detailed data panel on the right. The data panel includes fields for Title, Link, Part, Ctype, Notes, Int.N., Weib., and D.ord. Below this, there is a 'Barcode_Analyse' section with 'Nr. of an.: 3953' and 'Result: GACCTATATTTGTTATTTGGGGTTTGACGA'. The 'Methods' section lists 'Sequencing 2' and 'Sequencing 3'. At the bottom, a 'Parameter value' section shows a sequence alignment with a 'Mark: 9' indicator.

Morphological Collection

Collection specimen

Number: ZFMK-TIS-10605

Depositor: Mattem, Dirk

Source: GBOL-ZFMK

Acc. date: 17_6_2013

Label

Title:

Trans: Type:

Notes:

- complete = all data from label transferred in database
- Complete GBOL = GBOL Voucher ready for copy
- copied = data copied by script
- copy error = status after copy has failed
- copy skipped = copy not done because target entry already present
- curator review required = curator for transcription / review required
- extern review required = extern curator for transcription / review required
- final curator review = specimen has been finally reviewed by curator
- first curator review = specimen has been reviewed by curator for first time
- incomplete = transcription started by technical staff
- not started = transcription not started
- Ready for Biobank = Morphological voucher ready for copy to Biobank

Set status: **Complete GBOL**

Collection specimen

Number: ZFMK-TIS-10605

Depositor: Mattem, Dirk

Source: GBOL-ZFMK

Acc. date: 17_6_2013

Label

Title:

Trans: copied = data copied by script

Type:

Notes: OK: CopyCollectionSpecimenID=224189402

New status: **copied**

ByCollection, Database: DiversityCollection_ZFMK, v. 4.3.102

Connection: Grid Query Data Administration Help

Acc. No: ZFMK C01504

Specimen: 224189402

Event: 6709

Version: 8

Withold:

Collection specimen

Number: ZFMK C01504

Depositor: Mattem, Dirk

Source: GBOL-ZFMK

Acc. date: 17_6_2013

Suppl.:

State:

Projects

Notes

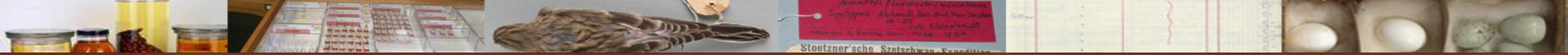
Original

Additional: GBOL_Ext_059 76/96

Internal: Sonnenbobl_Mattem_10576-10623

Problems

 MS SQL procedure: **Copy_Specimen_and_Set_Relation**



Data Exchange with Outside World

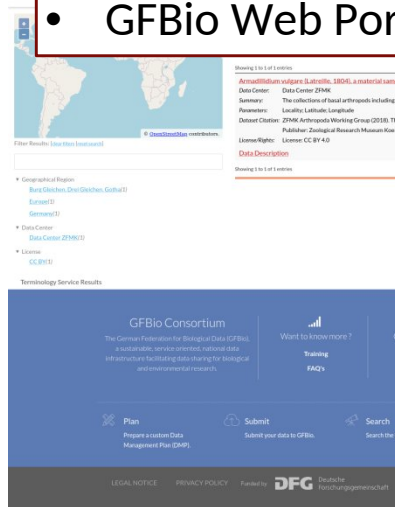
```

<?xml version="1.0" encoding="UTF-8"?>
<abcd21:DataSets>
<abcd21:DataSet>
<abcd21:DataSetGUID>https://id.zfmk.de/dataset_ZFMK/26798/<abcd21:DataSetGUID>
<abcd21:Unit>
<abcd21:UnitGUID>https://id.zfmk.de/collection_ZFMK/10605/496819/67984/<abcd21:UnitGUID>
<abcd21:SourceInstitutionID>ZFMK</abcd21:SourceInstitutionID>
<abcd21:SourceID>TIS/<abcd21:SourceID>
<abcd21:UnitID>ZFMK-TIS-10605/<abcd21:UnitID>
<abcd21:UnitIDNumeric>10605/<abcd21:UnitIDNumeric>
<abcd21:RecordBasis>MaterialSample/<abcd21:RecordBasis>
<abcd21:KindOfUnit>tissue sample/<abcd21:KindOfUnit>
<abcd21:MultiMediaObject>
<abcd21:MultiMediaObject>
<abcd21:fileURI>https://media.zfmk.de/earlts/image/82367/full/original/0/ZFMK_10605.tif/<abcd21:fileURI>
<abcd21:createDate>2018-02-02T15:08:04</abcd21:createDate>
</abcd21:MultiMediaObject>
</abcd21:MultiMediaObject>
<abcd21:Associations>
<abcd21:UnitAssociation>
<abcd21:UnitGUID>http://id.zfmk.de/collection_ZFMK/22418940/<abcd21:UnitGUID>
<abcd21:SourceInstitutionCode>ZFMK-Collections-v2/<abcd21:SourceInstitutionCode>
<abcd21:SourceName>CrustaceaColl Collection</abcd21:SourceName>
<abcd21:UnitID>ZFMK-Cr1504/<abcd21:UnitID>
<abcd21:AssociationType>tissue sample/<abcd21:AssociationType>
<abcd21:DataSetAccessPoint>https://biocase.zfmk.de/biocase/pywrapper.cgi?dsa=ZFMK-Collections-v2/<abcd21:DataSetAccessPoint>
</abcd21:UnitAssociation>

```

BioCASE Provider Software

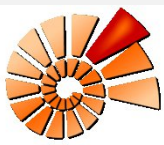
GFBio Web Portal



GBOL Web Portal

CETAF Stable Identifier:

https://id.zfmk.de/collection_ZFMK/json/10605/496819



Vielen Dank!

Das Team

Björn Quast

Birgit Klasen

Didem Cifci

Sandra Meid

Karl-Heinz Klameth

Wiebke Walbaum

All die guten Leute vom zte

Und natürlich das Team vom SNSB IT Zentrum

