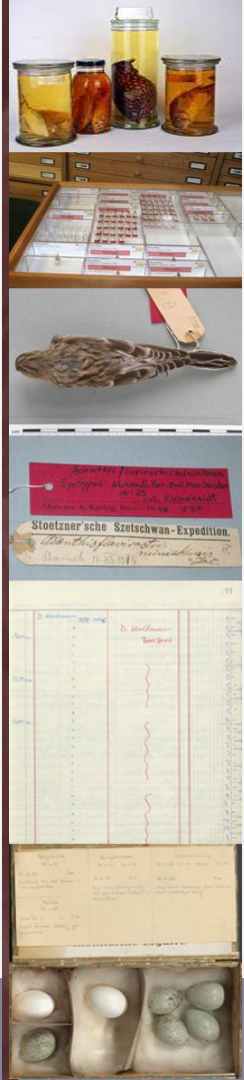
Sammlungsdaten am ZFMK

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/Occurence/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:

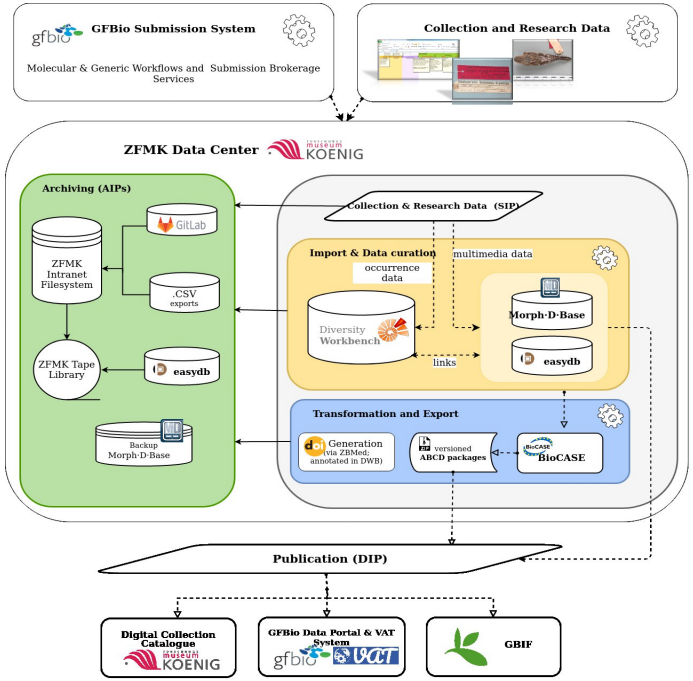
- Methods, relations, history, measurement, multimedia, identifier

Metadata:

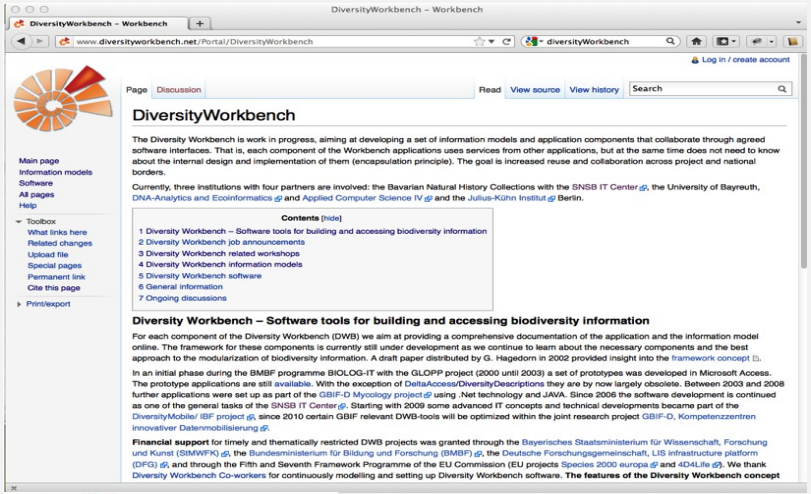
- Provenance, data sets, licenses, access rights, moratoria

Benefits

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



Diversity Workbench at ZFMK



snsb
it center

- 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

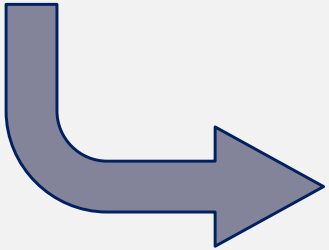
220,203 Analyses

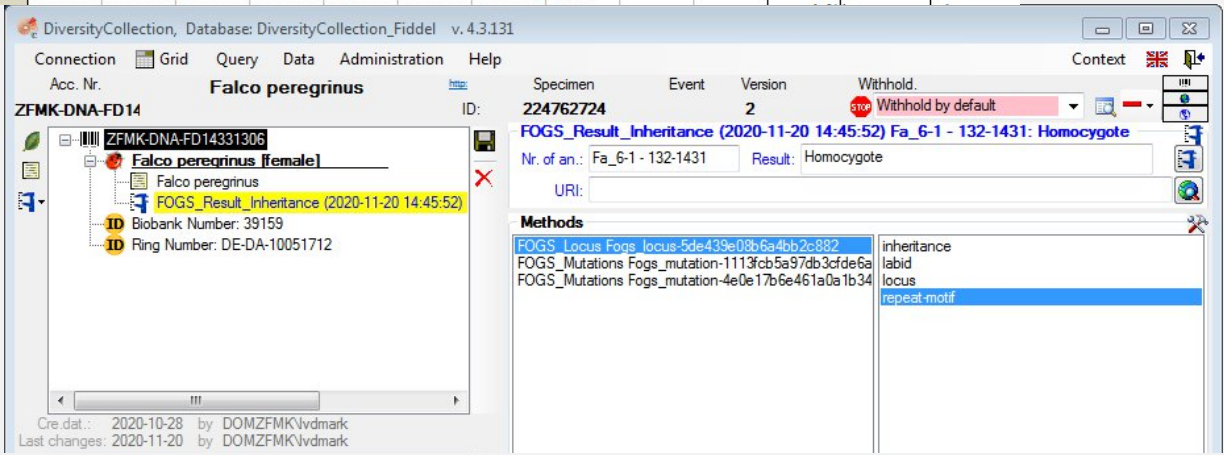
4,583,828 Parameter

- 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

								Wanderfalken (n=6)												
Species	Locus	Amp size	Repeat motif	Ref allele	NCBI No	Scaffold ID	NGS Val	Haplotype									Anteil*	inheritance	Allele Nr.
	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	-	AI9-Ref	145G	AI5-Ref	-	-	-	30%		
								-	-	-	-	AI9-Ref	-	AI5-Ref	-	-	-	4%		
								-	-	-	-	AI9-Ref	-	AI5-Ref	-	DEL-159-A	-	4%		
			(GATG)$_n$G					35T	-	85A	-	10.1	-	AI5-Ref	-	-	-	28%		
			(GATG)$_n$G					35T	-	85A	-	12.1	-	AI5-Ref	DEL-158-C	-	-	11%		
Falco perep	Fa-9	369	(GATG)$_n$G	9, 5	GCA_000	scaffold9_1	132-1431	T	-	A	-	11.1	-	5	-	-	-	42%	he	11b
Falco perep	Fa-9	369	(GATG)$_n$G	9, 5	GCA_000	scaffold9_1		T	-	A	-	12.1	-	5	GAGT	-	-	29%	he	12a
Falco perep	Fa-9	369	GATG	9, 5	GCA_000	scaffold9_1	133-1432	T	G	A	-	8	-	DEL-(GAT	-	-	A	41%	he	8a
Falco perep	Fa-9	369	GATG	9, 5	GCA_000	scaffold9_1		T	-	A	-	9	-	4.1	GAGT	-	-	29%	he	9c





Collection Data, Dataflow

DiversityCollection

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

*

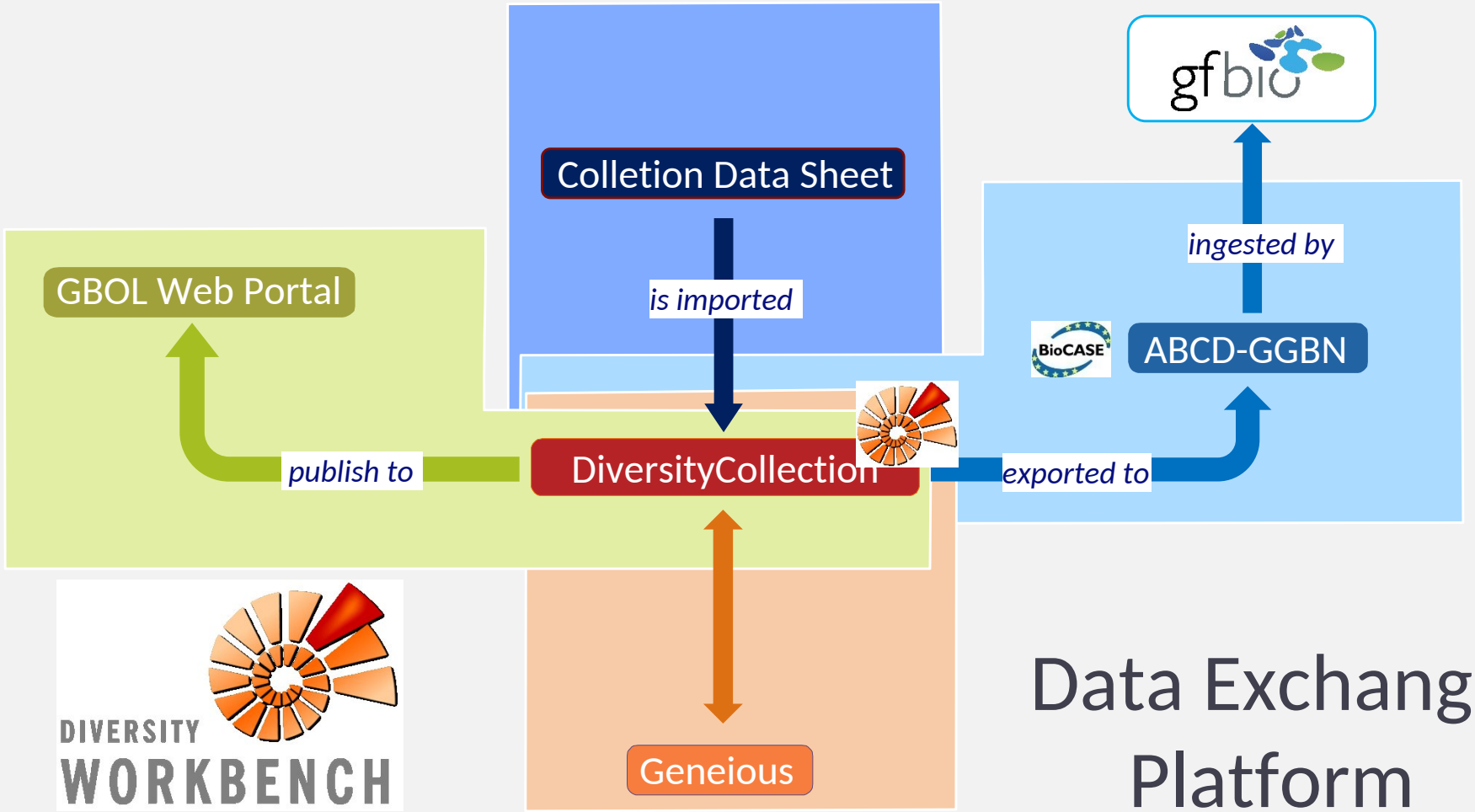
BioCASE
biocase.zfmk.de

Global Biodiversity Information Facility
Free and Open Access to Biodiversity Data

GGBN

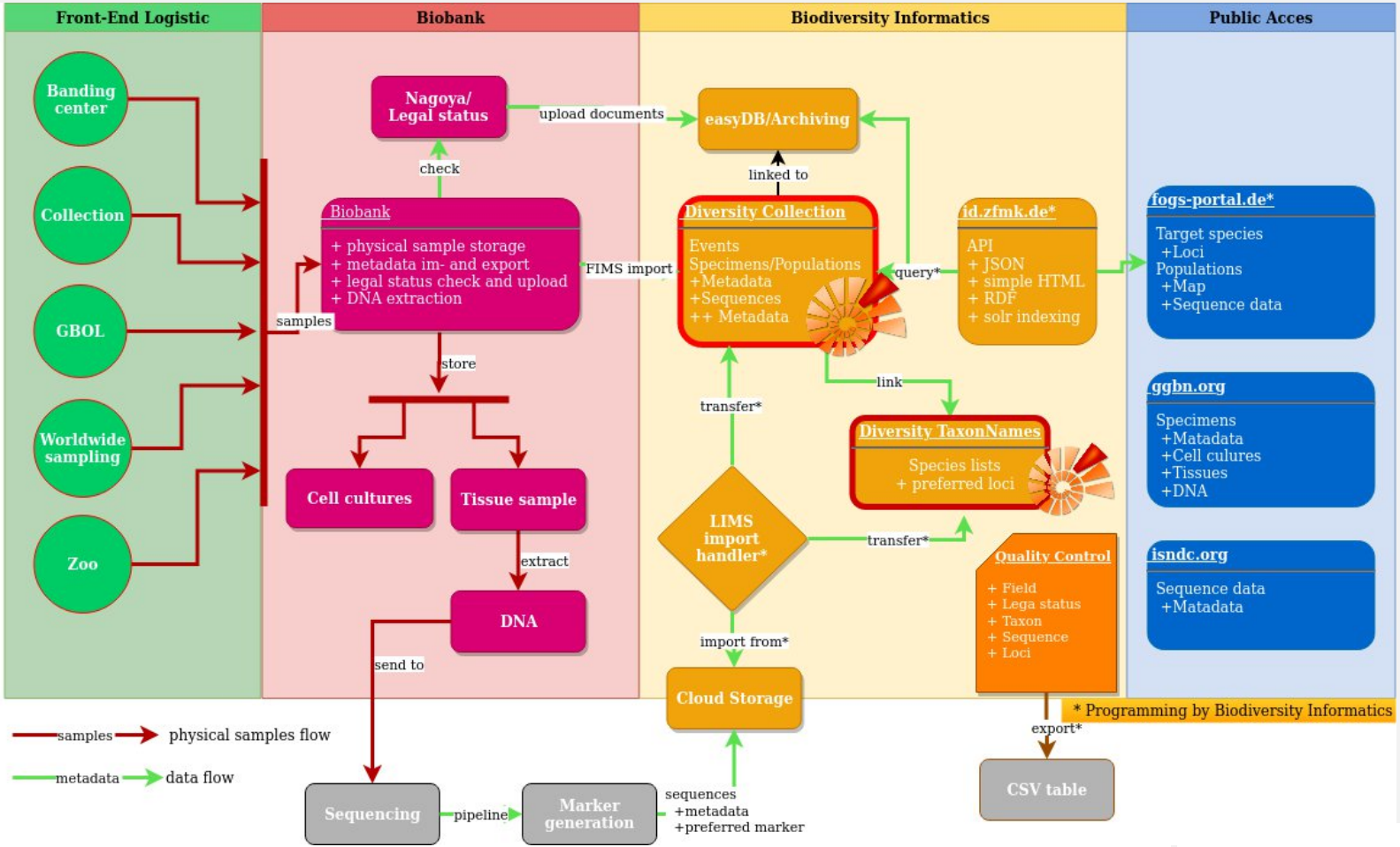
gfbio
GERMAN FEDERATION FOR BIOLOGICAL DATA

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

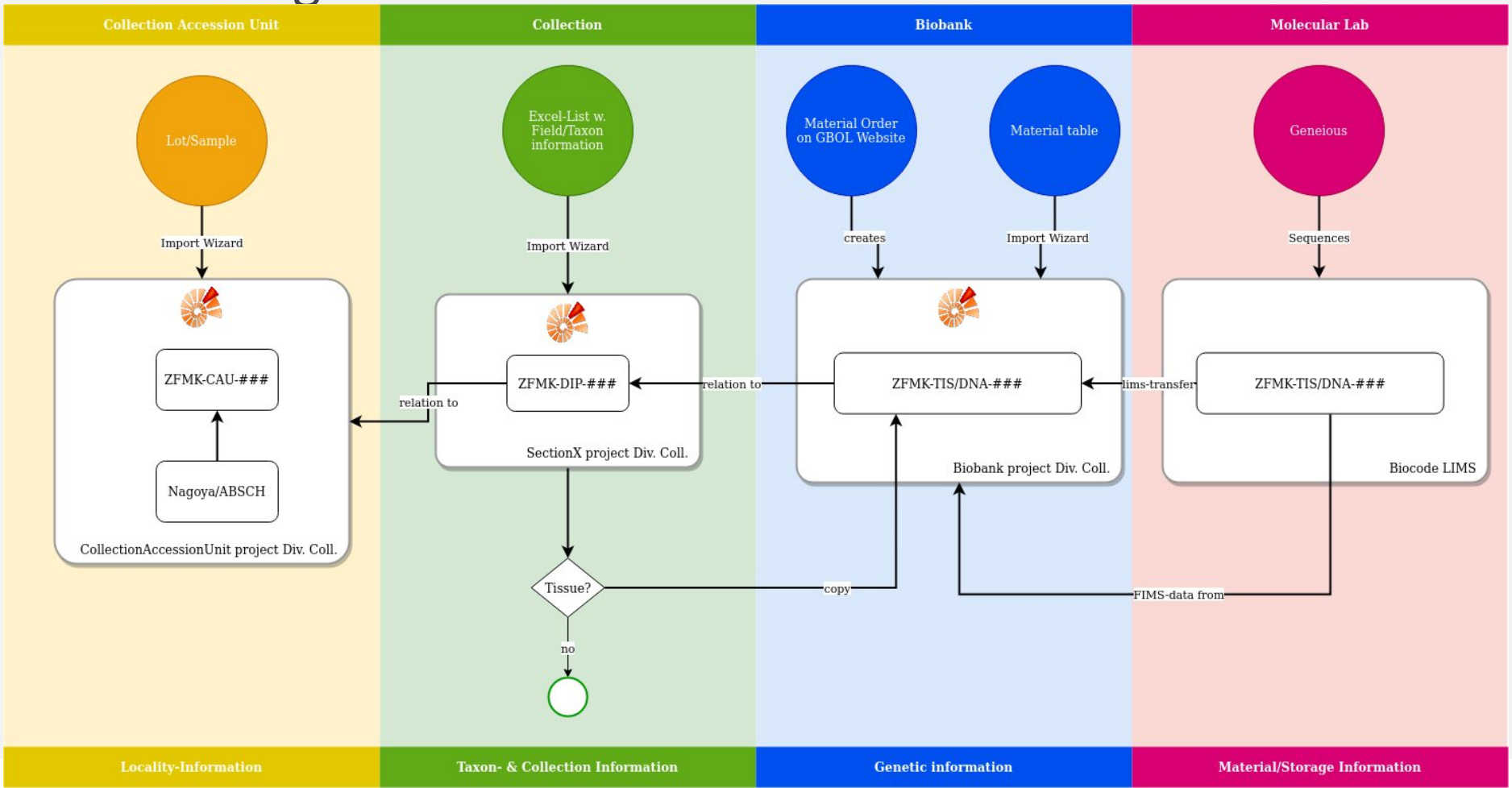


Data Exchange Platform

Data Exchange in Projects: The FOGS project

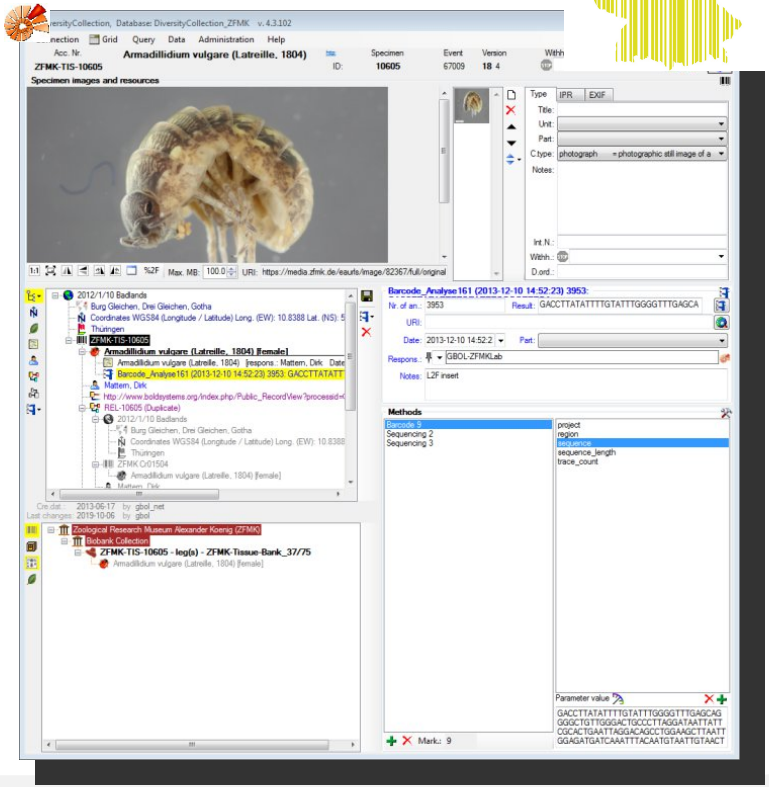


Data Exchange between Collections

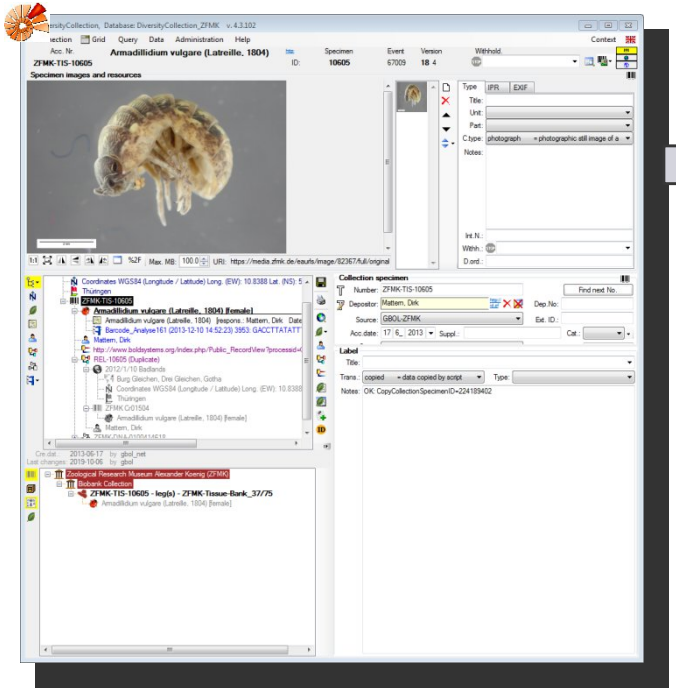


Data Exchange between Collections

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

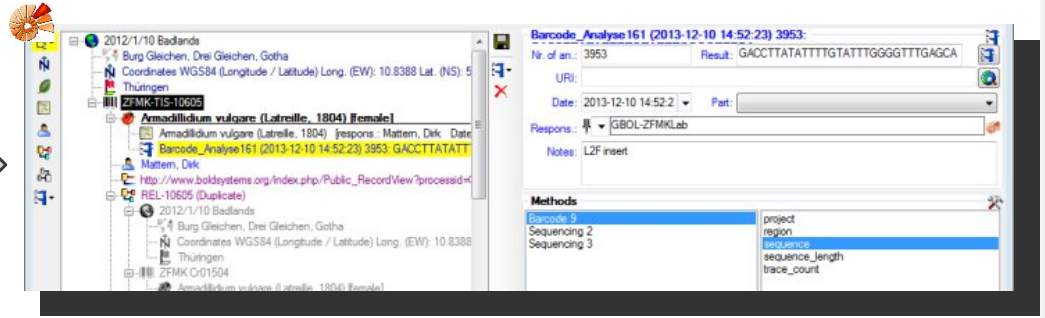


Occurrence data



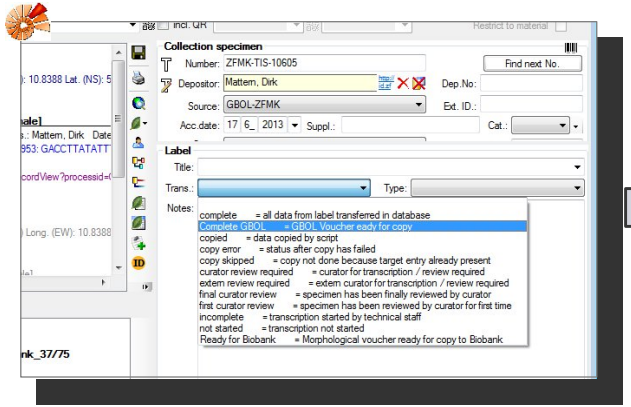
Biobank/GBOL

Sequences

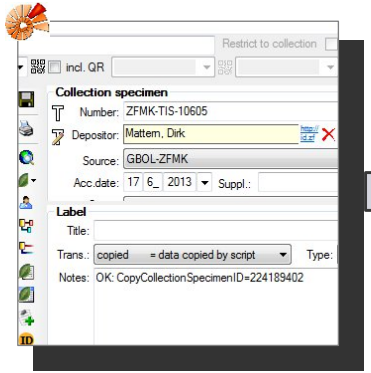


Laboratory

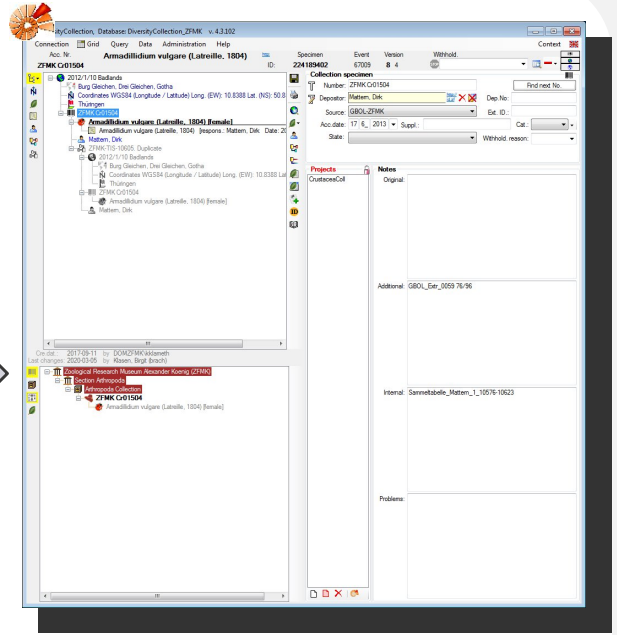
Morpholocial Collection



Set status: **`Complete GBOL`**



New status: **`copied`**



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/26790</abcd21:DatasetGUID>
  <abcd21:Units>
   <abcd21:Unit>
    <abcd21:UnitGUID>https://id.zfmk.de/collection_ZFMK/10605/496819/67904</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:MultiMediaObjects>
     <abcd21:MultiMediaObject>
      <abcd21:fileURI>https://media.zfmk.de/eaurls/image/82367/full/original/0/ZFMK_10605.tif</abcd21:fileURI>
      <abcd21:createDate>2018-02-02T15:08:04</abcd21:createDate>
     </abcd21:MultiMediaObject>
    </abcd21:MultiMediaObjects>
    <abcd21:Associations>
     <abcd21:UnitAssociation>
      <abcd21:UnitGUID>http://id.zfmk.de/collection_ZFMK/224189402</abcd21:UnitGUID>
      <abcd21:SourceInstitutionCode>ZFMK-Collections-v2</abcd21:SourceInstitutionCode>
      <abcd21:SourceName>CrustaceaColl Collection</abcd21:SourceName>
      <abcd21:UnitID>ZFMK Cr01504</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

DIVERSITY WORKBENCH

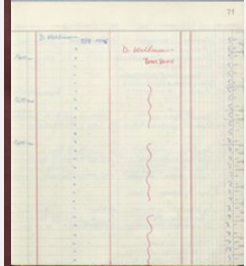
FORSCHUNGS museum KOENIG

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



Dec. 1st 2020