

Knowledge Graph(s) for NFDI4BIOIMAGE

Carsten Fortmann-Grote

October 16 2025

This work is published under Creative Commons Attribution License CC-BY-4.0 International.



Data driven lifescience is multimodal by nature

- Experimental
 - Imaging
 - Microscopy (>150 file formats), X-ray, Nuclear Magn. Resonance, Electron-Micr., ...
 - OMICS
 - Genomics (DNA), Transcriptomics (RNA), Proteomics (AA), ...
 - Flow cytometry
 - ...
- Theoretical
 - Simulations
 - Mathematical models
- Computational
 - High throughput data processing
 - Structural biology
 - AI

Data driven lifescience is multimodal by nature

- Experimental
 - Imaging
 - Microscopy (>150 file formats), X-ray, Nuclear Magn. Resonance, Electron-Micr., ...
 - OMICS
 - Genomics (DNA), Transcriptomics (RNA), Proteomics (AA), ...
 - Flow cytometry
 - ...
- Theoretical
 - Simulations
 - Mathematical models
- Computational
 - High throughput data processing
 - Structural biology
 - AI

Standardization of data and metadata is very challenging



Life Sciences

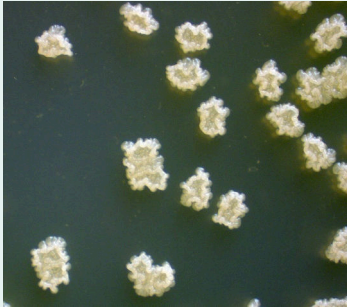
- [DataPLANT](#): Plant research data
- [FAIRagro](#): FAIR Data Infrastructure for Agrosystems
- [NFDI4Immuno](#) – National Research Data Infrastructure for Immunology
- [GHGA](#): German Human Genome-Phenome Archive
- [NFDI4Biodiversity](#): Biodiversity, Ecology and Environmental Data
- [NFDI4BIOIMAGE](#): National research data infrastructure for microscopy and bioimage analysis
- [NFDI4Health](#): NFDI personal health data
- [NFDI4Microbiota](#): NFDI for Microbiota Research

www.nfdi.de/consortia

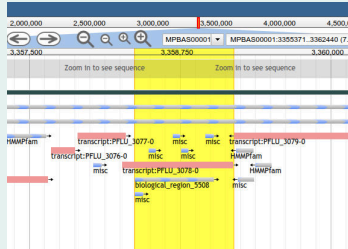
- Domain specific consortia X data specific consortia

Inhouse data

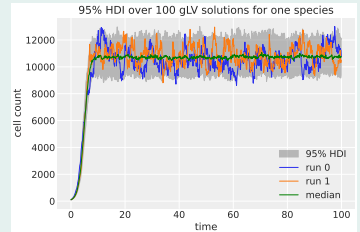
Image data



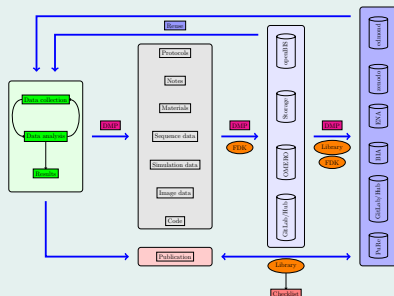
Sequence data



Code & Simulation data



Policy



MPI-EvolBio Data Policy

- Dedicated solutions per data type
- pre- and post publication repositories
- DMP per group and scientist (pilot phase)
- Monitored by library and data coordinator (FDK)

Repositories

Type	Documentation	Internal Storage (unpublished)	External Archive (post published)
Seq. Data	Electronic Lab Notebook	Institute storage	European Nucleotide Archive
Images	OMERO Image Database	Institute storage	Biolmage Archive
Code	Gitlab	Gitlab	Gitlab/Github



Questions we would like to be able to answer

- Find me images of cells with genotype XY and their DNA sequences!
- Annotate images with organism, genotype, phenotype metadata!
- Provide raw image and sequence data from paper doi:10.1234/abcdef



Questions we would like to be able to answer

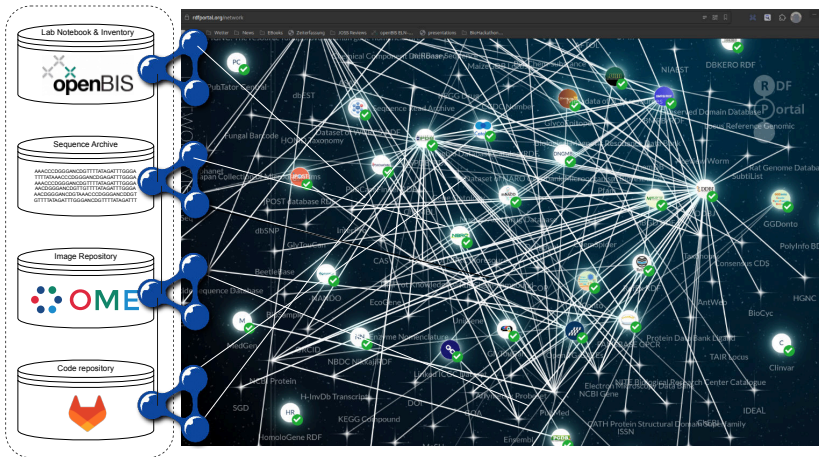
- Find me images of cells with genotype XY and their DNA sequences!
- Annotate images with organism, genotype, phenotype metadata!
- Provide raw image and sequence data from paper doi:10.1234/abcdef

Requirements

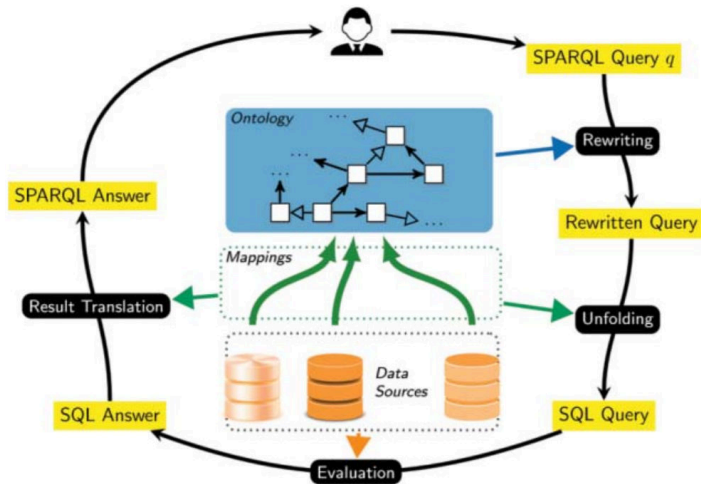
- In-house data integration
- Integration of external (meta)data sources



Data integration by Relational-to-RDF mapping



Knowledge Graph Virtualization



Reproduced from Xiao et al., 2019

Virtualization with Ontop-VKG (Calvanese et al., 2015)

SQL table omero.image

image.id	image.name	image.date	image.owner
1	img01.tif	20250101	1
2	T00-S01-RGB.czi	20241224	2

Mappings

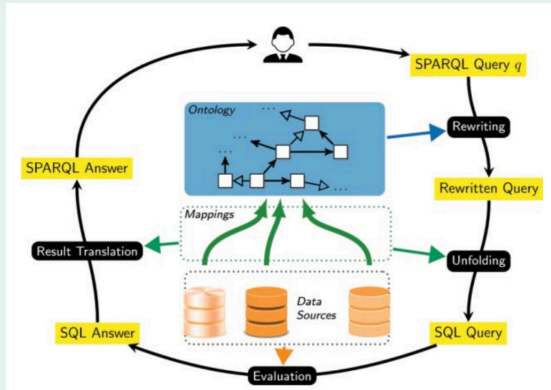
```
PREFIX dcterms: <http://purl.org/dc/terms/>
```

```
PREFIX core: <https://ld.openmicroscopy.org/core/>
```

```
:image/{image.id} a ome_core:Image;  
                   dcterms:title {image.name} ;  
                   dcterms:date {image.date} .
```

```
select image.id, image.name, image.date from image where image.owner='1'
```

Knowledge Graph Virtualization pros and cons



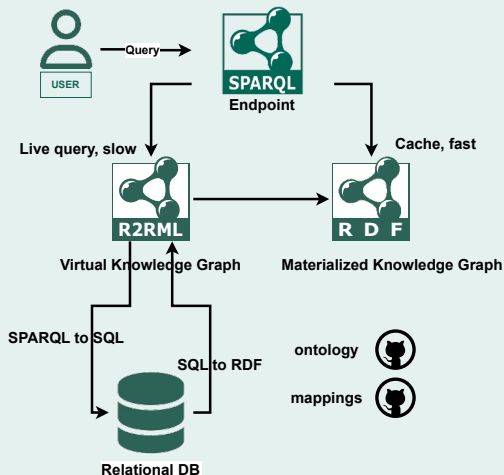
Reproduced from Xiao et al., 2019

- + Single source of truth
- + Instant synchronization
- – Requires access to DB backend
- – Performance not so great

Query response time is a concern for VKGs




Caching strategy to buffer slow response from VKGs



- Live response from virtual KG
- can be slow to answer complex queries
- RDF dump at regular intervals ($\approx 1/d$)
- SPARQL frontend serves dumped RDF
- fast but lags behind live data

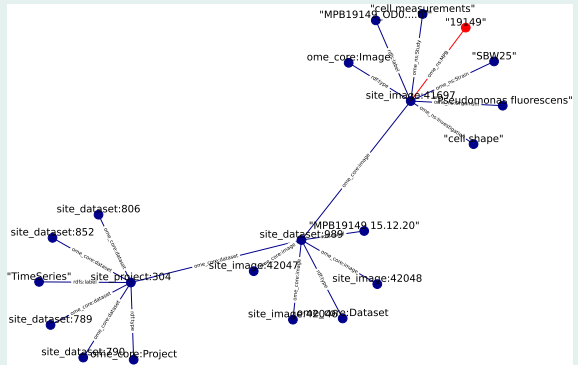
Annotated images in Omero



Key-Value Pairs 1

Organism	Pseudomonas fluorescens
Strain	SBW25
MPB	19149
Parent	0
Investigation	cell shape
Study	cell measurements
Assay	overnight culture in LB at 28°C with shaking; next day measured OD; 1 ml spin down; washed with M9; resuspended and inoculated in M9 (5 ml); incubated 3 h; measured OD again OD=0.3; agarose pad with M9; channel was cut only in the blue sticker part; 2 µl of culture on the pad

Derived RDF Graph



Samples collection

Collection: MPB Strains

Page 10 of 10

10 C. 5.100 al 59.150 5 10 Roma per pagina 180 = 0029999 FILTERS ESPOSURE

Four	Four	Five	+	Four	Four
------	------	------	---	------	------

[illegible]

Downloaded from <http://ajphaphapublications.org/> on 11/11/2015

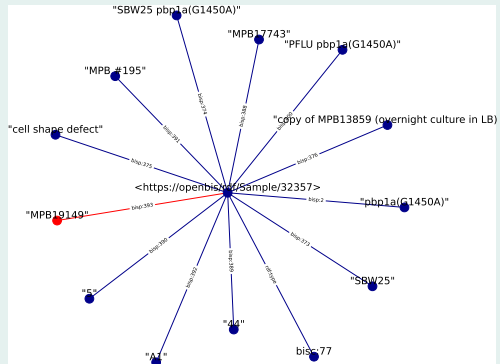
0146643	MF042712	Buanocone	dx.doi.org/10.1002/chem.200700150	small molecule
---------	----------	-----------	--	----------------

DOI: 10.1002/poc

bioRxiv preprint doi: <https://doi.org/10.1101/242350>; this version posted May 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

arsten Fortmann-Grote

openBIS Graph

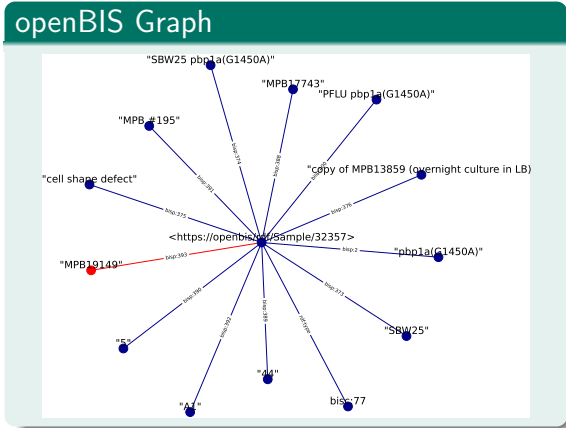


OMERO Graph

The graph illustrates the relationships between various datasets and images in an OMERO system. The nodes are labeled with their names and IDs, and the edges represent the relationships between them.

Nodes and their connections:

- site_dataset:806** is connected to **site_dataset:852** (via **ome_core:dataset**), **TimeSeries** (via **ome_core:dataset**), **site_dataset:789** (via **ome_core:dataset**), **site_dataset:790** (via **ome_core:dataset**), and **site_project:304** (via **ome_core:dataset**).
- site_dataset:852** is connected to **site_dataset:806** (via **ome_core:dataset**).
- TimeSeries** is connected to **site_dataset:806** (via **ome_core:dataset**).
- site_dataset:789** is connected to **site_dataset:806** (via **ome_core:dataset**).
- site_dataset:790** is connected to **site_dataset:806** (via **ome_core:dataset**).
- site_project:304** is connected to **site_dataset:806** (via **ome_core:dataset**).
- site_dataset:889** is connected to **site_image:42047** (via **ome_core:dataset**), **site_image:42048** (via **ome_core:dataset**), **site_image:42049** (via **ome_core:dataset**), **site_image:42050** (via **ome_core:dataset**), and **site_image:42051** (via **ome_core:dataset**).
- site_image:42047** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:42048** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:42049** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:42050** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:42051** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:41697** is connected to **site_dataset:889** (via **ome_core:dataset**).
- site_image:41697** is connected to **ome_core:Image** (via **ome_core:Image**).
- site_image:41697** is connected to **"MPB19149_OD0000000000"** (via **ome_core:Image**).
- site_image:41697** is connected to **"19149"** (via **ome_core:Image**).
- site_image:41697** is connected to **"SBW25"** (via **ome_core:Image**).
- site_image:41697** is connected to **"cell_measurements"** (via **ome_core:Image**).
- site_image:41697** is connected to **"cell_shape"** (via **ome_core:Image**).
- site_image:41697** is connected to **"Pseudomonas fluorescens"** (via **ome_core:Image**).



Example

How many images are in OMERO for openBIS sample MPB19149?

How many images are in OMERO for openBIS sample MPB19149?

Virtual KG system supports cross-database queries

Number of images (OMERO) grouped by sample (openBIS)

The screenshot shows a SPARQL query interface with a text editor containing the following query:

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 prefix ome_core: <http://www.openmicroscopy.org/rdf/2016-06/ome_core>
3 prefix mpiebkg: <http://lod.evolbio.mpg.de/>
4 prefix ome_ns: <http://www.openmicroscopy.org/ns/default>
5 prefix bisp: <http://www.semanticweb.org/grotec/ontologies/2025/8/openbis-vkg#>
6 prefix bisci: <http://www.semanticweb.org/grotec/ontologies/2025/8/openbis-vkg/>
7
8 SELECT ?sample (count(distinct ?img) as ?n_img) WHERE {
9
10   graph mpiebkg:ome {
11     ?img a ome_core:image;
12     ome_ns:MPB ?mpb_nz .
13     bind(concat("MPB", ?mpb_nz) as ?mpb_string)
14   }
15
16   graph mpiebkg:evolbis {
17     ?sample a bisci:??;
18     bisp:?? ?mpb_string .
19   }
20
21 }
22 group by ?sample
23 order by desc(?n_img)
24 LIMIT 10
25
```

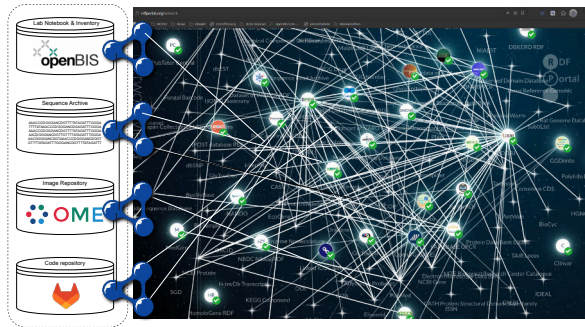
Below the query editor are buttons for Execute, Download, and Share. To the right are links for Index Information, Backend Information, and Shortcuts / Help. Below the buttons are links for Format, Clear cache, Analysis, and Examples. A dropdown menu for Context sensitive suggestions is also present.

Query results:

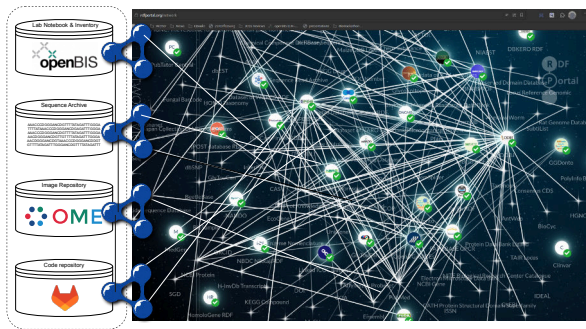
10 lines found | 31ms in total | 30ms for computation | 1ms for resolving and sending

	?sample	?n_img
1	26032	712
2	32357	623
3	27919	431
4	27918	260

Connecting local resources to the semantic web



Connecting local resources to the semantic web



Leverage NFDI to

- make graphs discoverable (F)
- ensure aligned terminologies (I)

NFDI4BIOIMAGE Knowledge Graph(s)

```
https://kg.nfdi4bioimage.de/#/datasets/N4BIKG/query
```

```
<<prefixes>>
```

```
select distinct ?graph (count(?s) as ?triples) where {  
  graph ?graph {?s ?p ?o}  
}  
group by ?graph  
order by desc(?triples)
```

graph	triples
https://nfdi4bioimage.de/kg/owl	13887
https://kg.nfdi4bioimage.de/n4bikg/owl	12382
https://kg.nfdi4bioimage.de/n4bikg/n4bi_zenodo_community	3482
https://kg.nfdi4bioimage.de/n4bikg/core	19
https://kg.nfdi4bioimage.de/n4bikg/services	5

`n4bikg:owl` Duplicates employed ontologies
to support local lookup

`n4bikg:zenodo` Metadata harvested from
the "NFDI4BIOIMAGE" zenodo
community

`n4bikg:core` Core data on
NFDI4BIOIMAGE consortium

`n4bikg:services` Web services and RDF
resources maintained by
NFDI4BIOIMAGE



n4bikg:core :: Core data about NFDI4BIOIMAGE

s	p	o
n4bi:node	rdf:type	nfdicore:0001039
n4bi:node	dc:description	Nationale Forschungsdateninfrastruktur für Mikroskopie und Bildanalyse
n4bi:node	nfdicore:0001006	wd:Q113500855
n4bi:node	nfdicore:0001008	https://nfdi4bioimage.de
n4bi:node	nfdicore:0010015	N4BI
n4bi:node	nfdicore:0010015	NFDI4BIOIMAGE
n4bi:node	wdp:P121	https://omero-nfdi.uni-muenster.de
n4bikg:core	rdf:type	wd:Q31386861
n4bikg:core	ndficore:0000142	wd:Q18199165
n4bikg:owl	rdf:type	wd:Q31386861
n4bikg:owl	ndficore:0000142	wd:Q18199165
n4bikg:services	rdf:type	wd:Q31386861
n4bikg:services	ndficore:0000142	wd:Q18199165
https://omero-nfdi.uni-muenster.de	rdf:type	nfdicore:0000001
https://omero-nfdi.uni-muenster.de	rdf:type	obo:NCITC15426
https://omero-nfdi.uni-muenster.de	nfdicore:0000201	https://omero-nfdi.uni-muenster.de/vkg/sparql
https://omero-nfdi.uni-muenster.de/vkg/sparql	rdf:type	wd:Q33002955



s	p	o
nfdicore:0001039	rdf:type	http://www.w3.org/2002/07/owl#Class
nfdicore:0001039	rdfs:label	nfdi consortium
nfdicore:0001039	rdfs:comment	An NFDI Consortium is a collaborative organizational ...
nfdicore:0001039	rdfs:comment	The National Research Data Infrastructure (NFDI) is intended to ...
nfdicore:0001039	iao:0000119	https://www.dfg.de/en/research-funding/funding-initiative/nfdi
nfdicore:0001039	iao:0000119	https://www.nfdi.de/
nfdicore:0001039	iao:0000118	National Research Data Infrastructure
nfdicore:0001039	iao:0000118	Nationale Forschungsdateninfrastruktur
nfdicore:0001039	rdfs:subClassOf	nfdicore:0000006
...

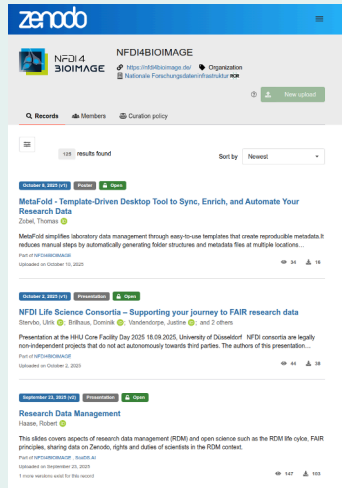


n4bikg:zenodo_community :: Zenodo records in NFDI4BIOIMAGE community

```
zenodo:11031747
  datacite:IsPartOf <https://zenodo.org/communities/mpievolbio/> ,
                    <https://zenodo.org/communities/nfdi4bioimage/>;
  datacite:IsVersionOf zenodo:11031746;
  datacite:Issued "2024-04-22"^^xsd:date;
  datacite:Updated "2024-07-06"^^xsd:date;
  datacite:creator orcid:0000-0002-2579-5546;
  datacite:description "This presentation was given at ...";
  datacite:resourceType "Presentation";
  datacite:right "Creative Commons Attribution 4.0 International";
  datacite:title "NFDI4BIOIMAGE" .

orcid:0000-0002-2579-5546
  datacite:affiliation ror:0534re684;
  datacite:familyName "Fortmann-Grote";
  datacite:givenName "Carsten";
  datacite:identifier "https://orcid.org/0000-0002-2579-5546"^^xsd:anyURI .

ror:0534re684 rdf:type ontology:NFDI_0000003;
  rdfs:label "Max Planck Institute for Evolutionary Biology";
  datacite:identifier "https://ror.org/0534re684"^^xsd:anyURI .
```



```
<https://omero-nfdi.uni-muenster.de>  
  rdf:type          obo:NCIT_C15426 , ontology:NFDI_0000001;  
  ontology:NFDI_0000201 <https://omero-nfdi.uni-muenster.de/vkg/sparql> .  
  
<https://omero-nfdi.uni-muenster.de/vkg/sparql>  
  rdf:type  wd:Q33002955 .  
  
<https://nfdi4bioimage.de/rdf/node>  
  wdt:P121  <https://omero-nfdi.uni-muenster.de> .
```



dataset	title	publisher
kgi:KGR83	FactGrid	wd:Q1438165
kgi:KGR80	GESIS KG	wd:Q1485220
kgi:KGR77	NFDI4BIOIMAGE KG	wd:Q168426
kgi:KGR77	NFDI4BIOIMAGE KG	wd:Q1493545
kgi:KGR77	NFDI4BIOIMAGE KG	wd:Q116959508
kgi:KGR72	Semantic Kompakkt	wd:Q2399120
kgi:KGR7	Culture Knowledge Graph	wd:Q1388737
kgi:KGR69	Wikidata	wd:Q8288



- Roll out virtualization scheme on all public OMERO instances
- Automate workflows
- Graphical UIs

Thank you

- NFDI4BIOIMAGE: Josh Moore, Susanne Kunis
- German Bioimaging: Tiago Lubiana + Andra Waagmeester and others
- MPI EvolBio: Paul Rainey, Daniel Olvera
- FDM-SH: Inken Wohlers (Research Center Borstel)
- NFDI Section Metadata KG Working Group: Heike Fliegl, Lozana Rossenova, Fidan Limani, ...



Thank you

- NFDI4BIOIMAGE: Josh Moore, Susanne Kunis
- German Bioimaging: Tiago Lubiana + Andra Waagmeester and others
- MPI EvolBio: Paul Rainey, Daniel Olvera
- FDM-SH: Inken Wohlers (Research Center Borstel)
- NFDI Section Metadata KG Working Group: Heike Fliegl, Lozana Rossenova, Fidan Limani, ...
- FIZ Karlsruhe for the kind invitation
- everyone for your attention!

