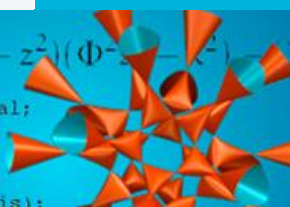


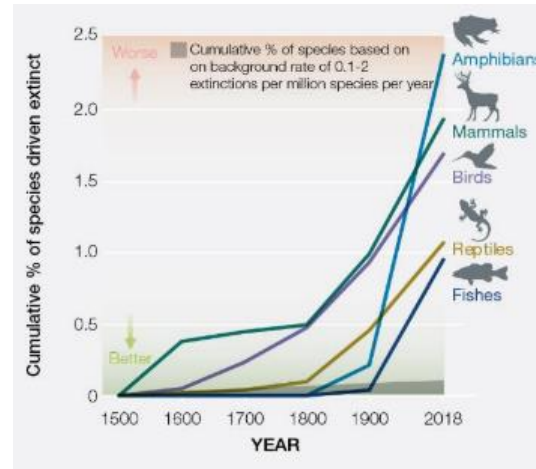
FAIRe Daten und FAIRe Software in der Biodiversitätsforschung

Bernhard Seeger
Philipps-Universität Marburg

```
|(Φ²x² - y²)|(Φ²y² - z²)|(Φ²z² - x²) - 1  
perspective=central;  
spec_c=150.0;  
radius=10.0;  
sextic=rotate(  
    sextic,-0.1,xAxis);
```

A 3D visualization of a complex, multi-lobed surface, likely a sextic surface, rendered in orange and blue. The surface is composed of many small, interconnected lobes, creating a complex, crystalline structure.

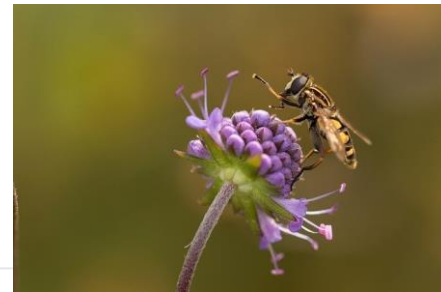
Motivation



- Díaz, Sandra, et al. "Summary for policymakers of the global assessment report on biodiversity and ecosystem services of **IPBES**." (2019).

Motivation

- The diversity is declining faster than at any time in human history.
- Most of nature's contributions to people are not fully replaceable, and some are irreplaceable.
- More than 75 per cent of global food crop types rely on **animal pollination**. (Ollerton et al. 2011)



What is biodiversity about?

- Biological diversity is the variety of life on Earth and the relationship among living items.
 - foundation for human well-being (food, water, shelter)
 - ecosystem services (Brauman et al. 2007)
 - lineages reflecting millions of years of evolutionary diversification (Vucetich et al. 2015)

Biodiversity: mostly unknown

- 86% of existing species on Earth and 91% of species in the ocean are not detected yet (Mora et al. 2011)
- knowledge of the known species remains uneven and irregular, especially across the Tropics.

The link to spatio-temporal databases

- Biodiversity research relies on large spatio-temporal databases.
 - occurrence of a particular **species** at **a place** at **a point in time**
- There is also a fundamental change in (biodiversity) research:
 - We must all accept that science is data and that data are science,... (Hanson et al. 2011)
 - Large spatio-temporal databases have been built up and integrated over the last decade.

Projects



- Consortium of 19 Partners
- Funded by DFG
- see (Diepenbroek et al. 2014)

Natur 4.0
Sensing Biodiversity



- Consortium with 10 partners
- Funded by State Hessen

Overview

- Data in the Biodiversity Realm
- FAIR (Data) Principles
- NFDI4Biodiversity

2. Data in the Biodiversity Realm

- Categories of data
 - Observational
 - occurrence of animals and plants
 - weather
 - Computational
 - generated by a computer model
 - Experimental
 - field experiments
 - Metadata
 - context information

Observational Data

- Three-dimensional data points
 - **time** (day, month, year, ...),
 - **place** (country, stateProvince, county, Latitude, Longitude, coordinateUncertainty, coordinatePrecision, ...).
 - **taxon** (genus, subgenus, specificEpithet, infraspecificEpithet, taxonRank)

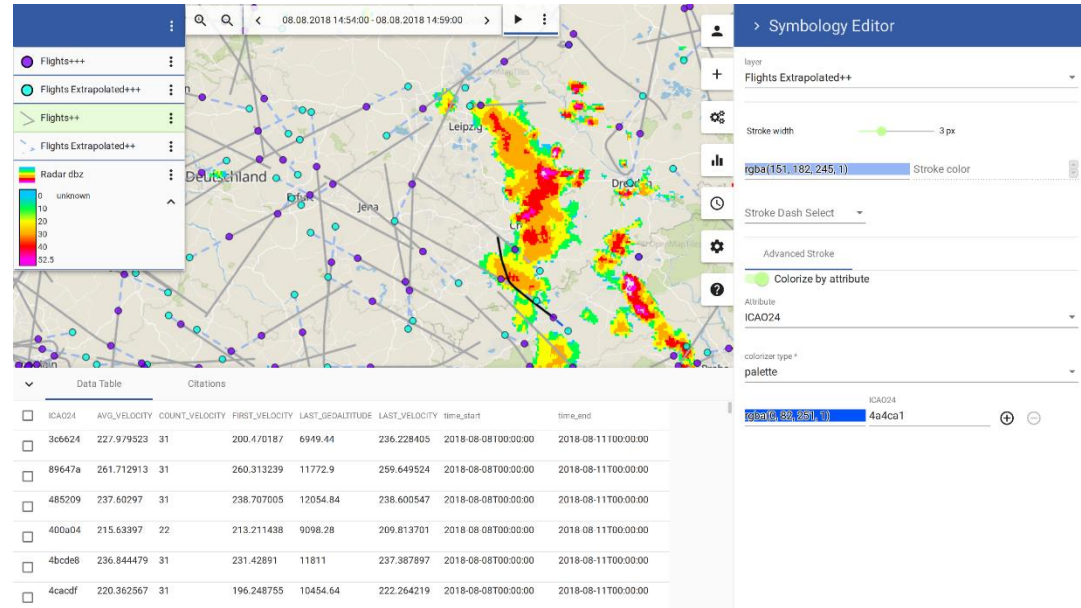
- Best source (GBIF)



- 1'391'799'117 records (March 10, 2020)
- It only contains 6-10% of the specimen

Real-Time Tracking

- It is quite common in biodiversity
 - Moose
 - Birds
 - Bats
 - ...
- Animals and other objects are sensors for environmental data



Remote Sensing (1)

- Satellite
 - Geostationary e.g. weather
 - Sunstationary e.g. land observation
- Meteosat 2nd generation
 - Spatio-temporal resolution: 3 km x 3 km x 15 minutes
 - Multispectral camera
 - 13 channels
 - 3 RGB-channels
- Impressive pictures from Landsat
 - Okjökull Remembered



<https://earthobservatory.nasa.gov/images/145439/okjokull-remembered>

Remote Sensing (2)

- UAV (drone)
- plane/helicopter

- Instruments
 - Cameras
 - LIDAR
 - RADAR
 - ...



Additional Data Requirements

- Management of large raster data sets with multiple channels
 - Time-series of raster
- Multimedia data
 - Pictures
 - Movies
 - Audios
- 3-dimensional point sets (LIDAR)
 - Measures about the forest canopy

Microbiological Data

- Microorganisms and bacteria
 - “You never walk alone”
- Barcode of Life
 - Latitude and longitude, date of collection,...
 - Coverage to 2.5 million species by 2026
- Biological networks

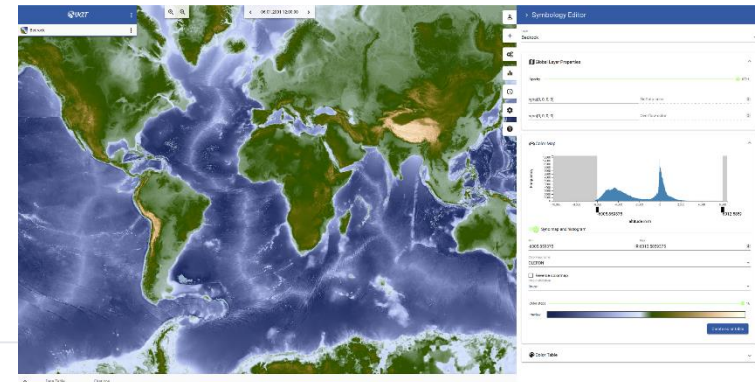


Grafik: Anke Becker

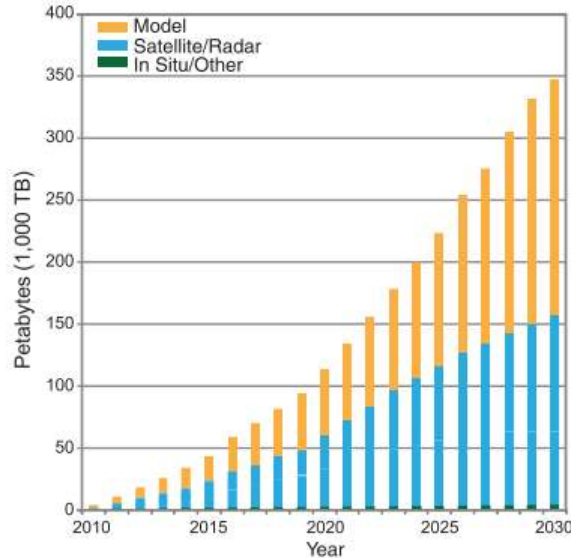


Computational Data

- Scientists use a model to compute a new data set from observations
 - Output is generally a raster file.
- Examples
 - point set \rightarrow raster
 - Given a few measurements.
 - Return an area-covering map.
 - channels of a raster \rightarrow raster
 - Computation of the drop size in a cloud



Data Volume in Biodiversity



- Overpeck, J., Meehl, G., Bony, S., and Easterling, D. (2011). Climate data challenges in the 21st century. *Science*, 331(6018), 700.

3. FAIR (Data) Principles

- Overarching Goal
 - **Research resources** should be
 - Findable
 - Accessible
 - Interoperable
 - Reusable



- The main focus of FAIR is the reuse of valuable research objects.
 - It is **NOT** about data only, but addresses code, workflows as well.

To be Findable

- F1 (meta)data are assigned a globally **unique and persistent identifier**
- F2 data are described with **rich metadata** (defined by R1 below)
- F3. metadata clearly and explicitly include the **identifier** of the data it describes
- F4. (meta) data are registered or **indexed in a searchable resource**

To be Accessible

- A1. (meta)data are **retrievable by their identifier** using a standardized communications protocol
- A1.1 the **protocol is open**, free, and universally implementable
- A1.2 the protocol allows for an **authentication and authorization** procedure, where necessary
- A2. **metadata are accessible**, even when the data are no longer available

To be Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation**.
- I2. (meta)data use **vocabularies that follow FAIR principles**
- I3. (meta)data include **qualified references** to other (meta)data

To be Reusable

- R1. meta(data) are richly described with a plurality of **accurate and relevant attributes**
- R1.1. (meta)data are released with a clear and accessible **data usage license**
- R1.2. (meta)data are associated with **detailed provenance**
- R1.3. (meta)data meet domain-relevant **community standards**

Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud

Barend Mons^{a,b,c+}, Cameron Neylon^d, Jan Velterop^e, Michel Dumontier^f, Luiz Olavo Bonino da Silva Santos^{h,g} and Mark D. Wilkinson^h

FAIR is not

... a data standard

- It is about **properties of a data source** for making it findable for reuse

... a semantic data model (e.g. Spatial RDF)

- It is about utilization of **machine-readable frameworks**
- There are a multitude of FAIR formats

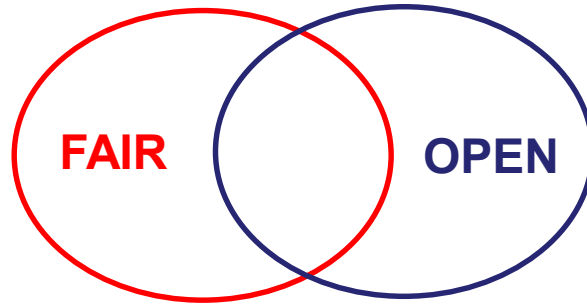
... made for human-machine interaction

- It is made for **autonomously interacting computers**

... is not limited to data

- **It applies to digital resources like data sets, code, workflows and research objects**

FAIR is not ...



... open

- There are good reasons to **protect data and services from public access**
 - Privacy, national security, ...
- Provide comprehensible **rules for their reuse** (e.g. a machine readable licence)
- Participation of a **broad range of data producers and various stakeholders for reusing** the data

4. NFDI4BioDiversity



Biodiversity – an artistic rendition by @microbiobac

- Large consortium of biologists, ecologists, and computer scientists
 - Part of the NFDI (National Research Data Infrastructure) initiative of the federal government
 - systematically exploit, sustainably preserve and make accessible the data collections of science and research.

Objectives of NFDI4Biodiversity

1. Promoting research data management as an integral part of biodiversity research
2. Enabling FAIRness of data
3. Consolidating FAIRness with quality
4. Embedding NFDI4BioDiversity into the (inter-)national landscape
5. Addressing NFDI-wide cross-cutting topics
6. Promoting collaborative governance, viability and sustainability of data infrastructure services and science

Task Areas of NFDI4Biodiversity

1. Community engagement (2involve)

- 23 use case with community-driven topics and cross-domain user groups.

2. National and international networking (2connect):

- cross-domain activities of NFDI4BioDiversity with other NFDIs and international partners

3. Long-term data, tools and service preservation, certification (2consolidate):

- Consolidation of existing German infrastructures and data pipelines for data preservation, archiving and publication.

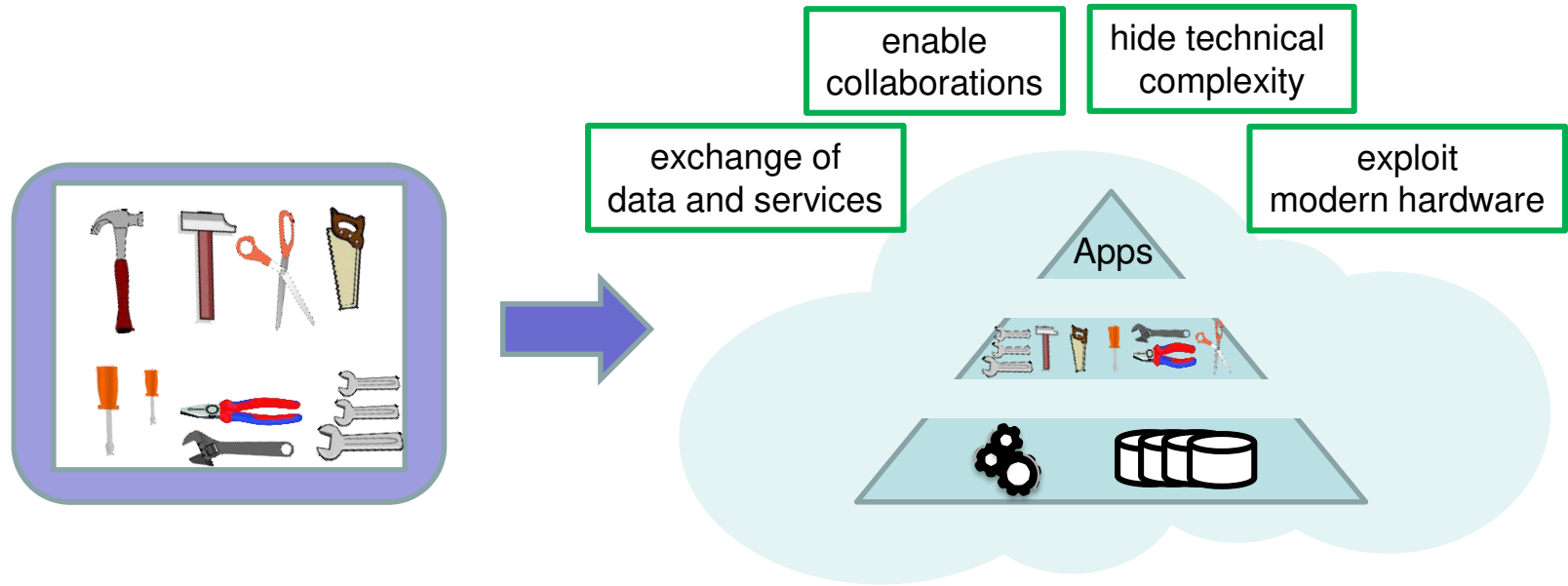
4. The NFDI-Research Data Commons (NFDI-RDC, 4all & 4future):

- Development of a Cloud-based research infrastructure to foster exchange of data and services

5. Task Area 5 – Coordination, collaborative governance & sustainability

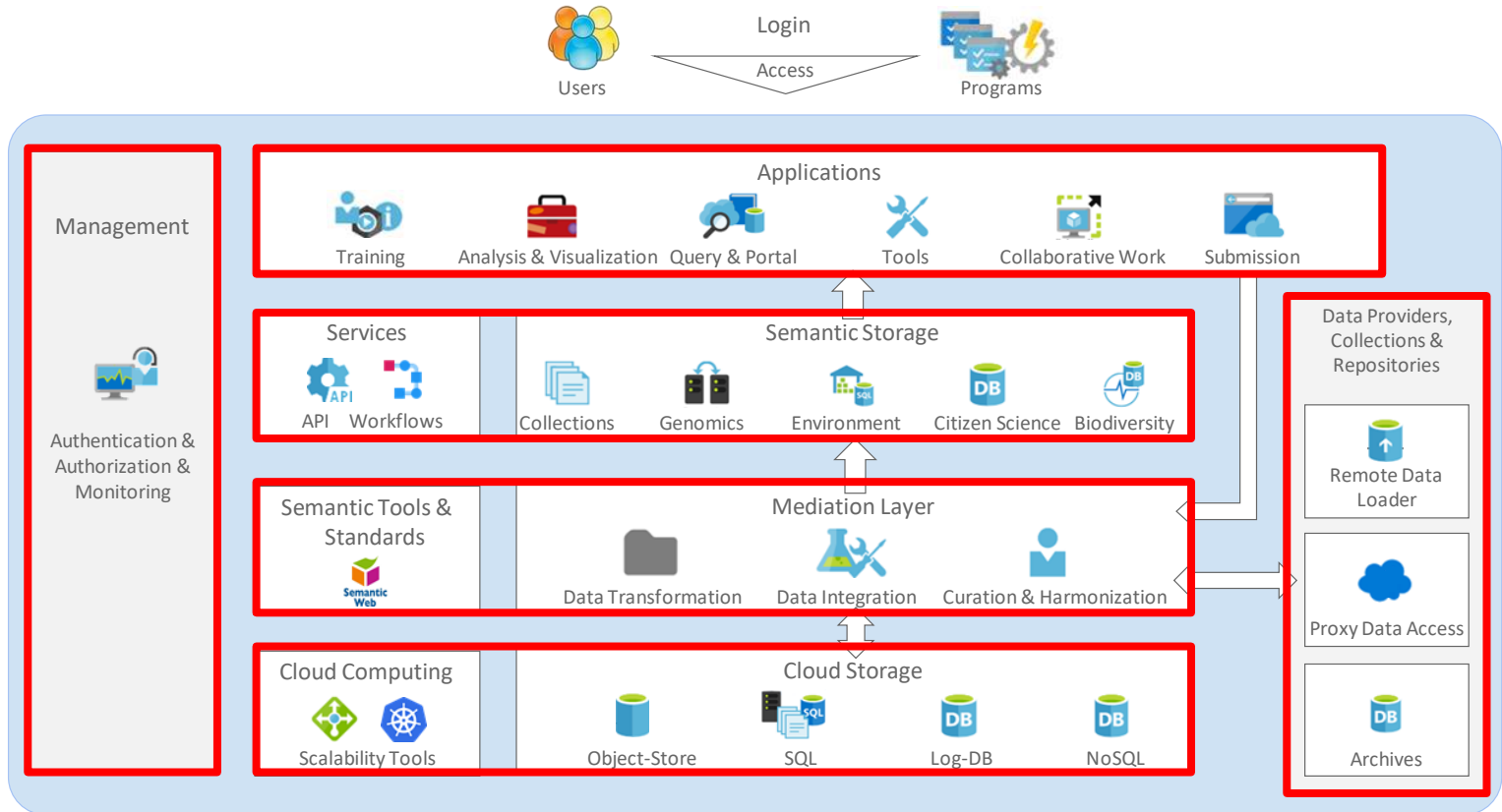
- Governance model and coordination of NFDI4BioDiversity

NFDI-Research Data Commons



- NFDI Research Data Commons (**NFDI-RDC**)
 - cloud-based infrastructure to support a FAIR data-driven research in biodiversity and beyond

Architecture of NFDI-RDC



Summary

- Biodiversity needs to be addressed
- Specific requirements of biodiversity on spatio-temporal databases
- FAIR Principles for research objects like data, code, and workflows
- NFDI4Biodiversity will provide a unique research platform that obey the FAIR principles for data and software.

Call to the spatial open-source community

- Development of FAIR open-source tools to preserve biodiversity.
- An open-source GEE would be very much appreciated, see vat.gfbio.org.

Literatur (1)

- Authmann, Christian, et al. "VAT: a system for visualizing, analyzing and transforming spatial data in science." *Datenbank-Spektrum* 15.3 (2015): 175-184.
- Brauman, Kate A., et al. "The nature and value of ecosystem services: an overview highlighting hydrologic services." *Annu. Rev. Environ. Resour.* 32 (2007): 67-98.
- Díaz, Sandra, et al. "Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services." (2020).
- Diepenbroek, Michael, et al. "Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio)." *Informatik 2014* (2014).
- Hanson, B., Sugden, A. & Alberts, B.: „Making data maximally available. *Science* 331,649 (2011).
- Hardisty, Alex R., et al. "The Bari Manifesto: An interoperability framework for essential biodiversity variables." *Ecological informatics* 49 (2019): 22-31.
- Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." *Information Services & Use* 37.1 (2017): 49-56.

Literatur (2)

- Mora, Camilo, et al. "How many species are there on Earth and in the ocean?." *PLoS biology* 9.8 (2011).
- Ollerton, Jeff, Rachael Winfree, and Sam Tarrant. "How many flowering plants are pollinated by animals?." *Oikos* 120.3 (2011): 321-326.
- Overpeck, Jonathan T., et al. "Climate data challenges in the 21st century." *science* 331.6018 (2011): 700-702.
- Vucetich, John A., Jeremy T. Bruskotter, and Michael Paul Nelson. "Evaluating whether nature's intrinsic value is an axiom of or anathema to conservation." *Conservation Biology* 29.2 (2015): 321-332.
- Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016).

Web-Quellen (letzter Zugriff am 11.03.2020)

- GBIF: <https://www.gbif.org/>
- GFBio: <https://www.gfbio.org/>
- NFDI: <http://www.rfii.de/de/nationale-forschungsdateninfrastruktur-nfdi/>
- NFDI4Biodiversity: <https://www.nfdi4biodiversity.org/>
- VAT-System: <https://vat.gfbio.org>