biodiversa+
European Biodiversity Partnership

2nd Workshop on biodiversity monitoring data interoperability and harmonisation

4th of November– 9am to 12pm CEST

REC

The plenary sessions of this meeting will be recorded and shared on the Biodiversa+ website and Youtube channel

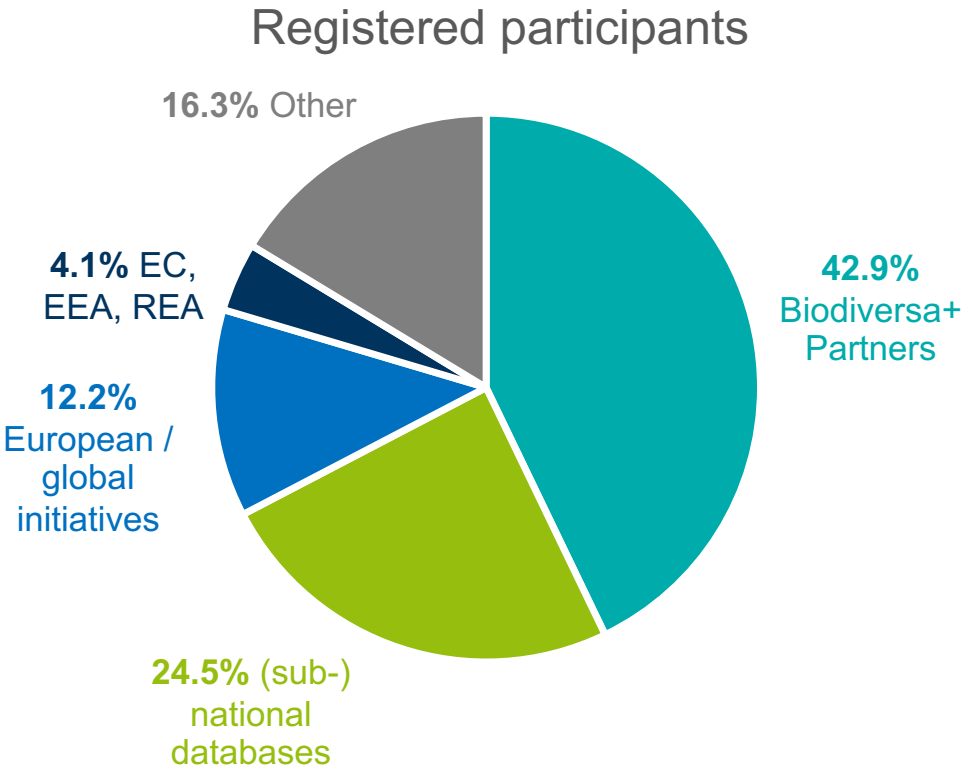
biodiversa+
European Biodiversity Partnership

Welcome words

By Alberto Basset, MUR

Welcome to the workshop on biodiversity monitoring data interoperability and harmonisation

+ 45
Registered participants



Agenda of the workshop

- ✓ Introduction of the meeting and scene setting *by Alberto Basset, MUR and Petteri Vihervaara, MoE_FI*
- ✓ Keynote speech on data interoperability and harmonisation for biodiversity, *By Hanna Koivula, CSC – IT Center for Science*
- ✓ The “Meetnetten” Flemish webtool – data architecture and data workflows, *by Dimitri Brossens, INBO*
- ✓ German NFDI4Biodiversity – data architecture and data workflows, *by Barbara Ebert, managing coordinator of NFDI4Biodiversity*
- ✓ Split sub-groups, *by Cécile Mandon, FRB-OT*
- ✓ Collaborative discussions in sub-groups
- ✓ Concluding words

Aims of Biodiversa+ in this context

- ✓ Better understanding of data architecture & operating dataflow in national and sub-national initiatives
- ✓ Exchange views with different actors about the level of harmonisation and interoperability the different databases
- ✓ Discuss how Biodiversa+ could launch concrete activities to improve data interoperability and harmonisatio tackling already identified issues, in order to support the overall objective of transnational biodiversity monitoring

To be developed:

✓ In connection with the biodiversity monitoring priorities at the (sub-)national, European, international level;
✓ In connection with the running activities in major international initiatives and projects;
✓In connection with the running activities of the ESFRI landscape ERIs, mainly referring to the Biosphere domain.

biodiversa+
European Biodiversity Partnership

Biodiversa+ ongoing work on data interoperability and harmonisation

By Alberto Basset, MUR

Timeline of ongoing work

1st of September WS

European & global focus

TODAY

4th of November WS

2023 - onwards: launch of activities/ actions to support data interoperability and harmonisation

Survey

Targetting participants of the 2 workshops and all European / global / (sub-)national biodiversity monitoring databases

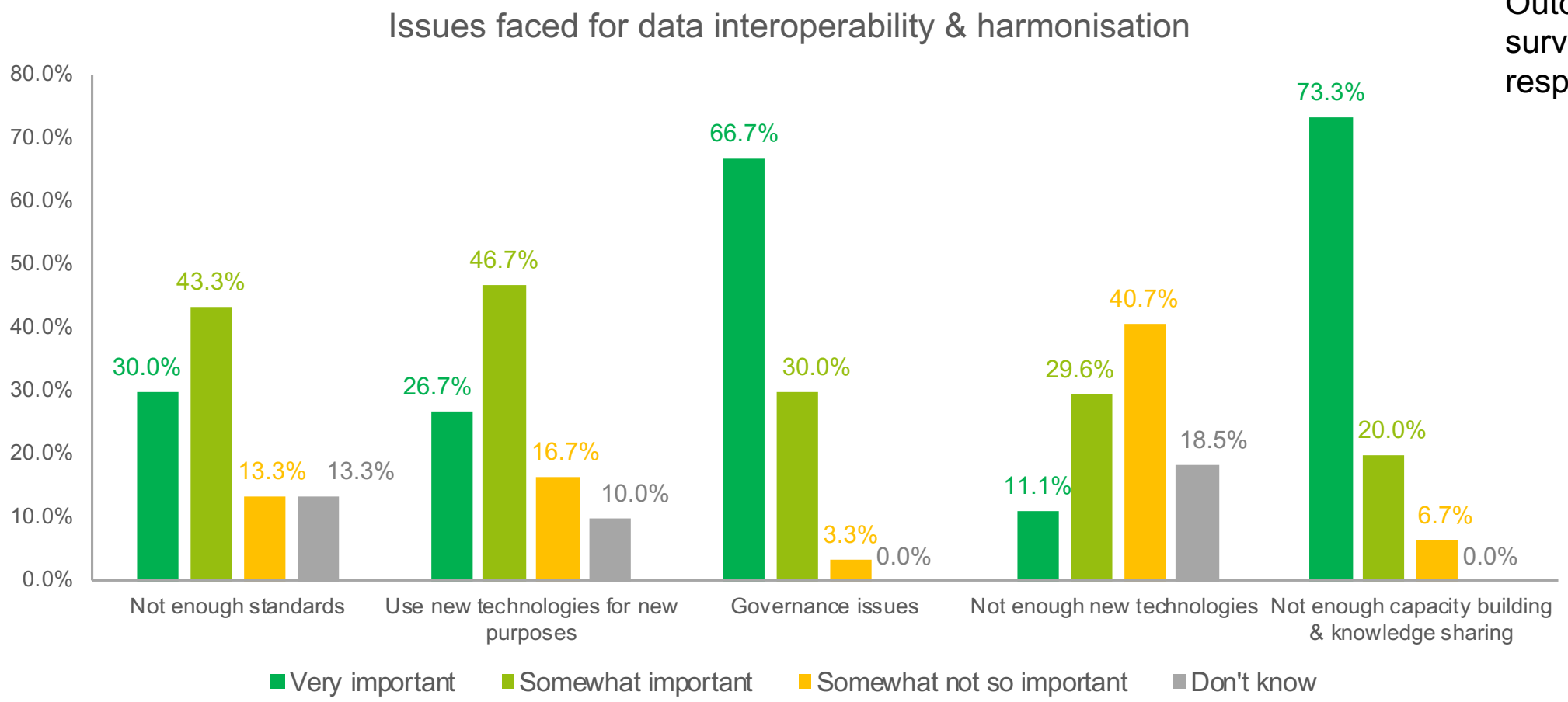
Report on data interoperability and harmonisation

Outcomes of the workshop – Issues to be tackled

Issues	Comments raised by workshop participants
Need for standards  	**1.** **Integration of machine learning** with human interpreted results **has no standards** + +, machine learning processes are not always transparently documented, the proliferation of ontologies which are not easy to cross-walk +. Requires initial human investigation and sense-making to build machine-readable repos. Machine observation has large data storage needs and IT infrastructures are not keeping up. **2.** **Use of new technologies** to tackle interoperability **and** data use & impact. 3. Not enough facilitation of implementation of **domain-specific standards** and **standards commonly used across sectors** **4.** Data collected by different agencies within government apply different standards – so biodiversity data interoperability can be seen as part of the process of mainstream (related to governance issues) 5. Continuous proliferation of metadata standards >> how to **translate this information between standards**? + Need metadata review and validation protocols to ensure the quality and consistency of information workflows. +
Not enough capacity building and knowledge sharing 	1. Too little attention to **capacity building in the poorer areas** of Europe 2. Semantic artefact development requires people who know not only one's institution data and metadata architecture but multiple ones. **Need for best practices in semantic artefact** development/ management/ governance for establishing guidelines. **3.** **Concepts** such as interoperability ontologies and semantics are **still unfamiliar** to many data collectors 4. There are provenance tools such as ORC ID, ROR, DOIs… but there are **not consistently used by the community** **5.** **No clear guidance/mandate from funding agencies** to adhere to specific standards facilitating data interoperability +++ (related with governance issues)
Governance	1. Much existing data is collected in **templates that lose raw data**. Eg. EU reporting. + From the management perspective it is important to see what dataflows are/ can be channelled into official reporting mechanisms + **2.** **Lack of harmonisation** on what to monitor **3.** **No one stop shop** (creating / asking to apply standards, providing a central register for those standards. (related with standards issues) **4.** **No clear roles**: tasks and goals for the different organisations and alignment of these.

How important are the issues preventing data interoperability?

Outcomes of the survey – 31 respondents



Outcomes of the first workshop –Biodiversa+ support

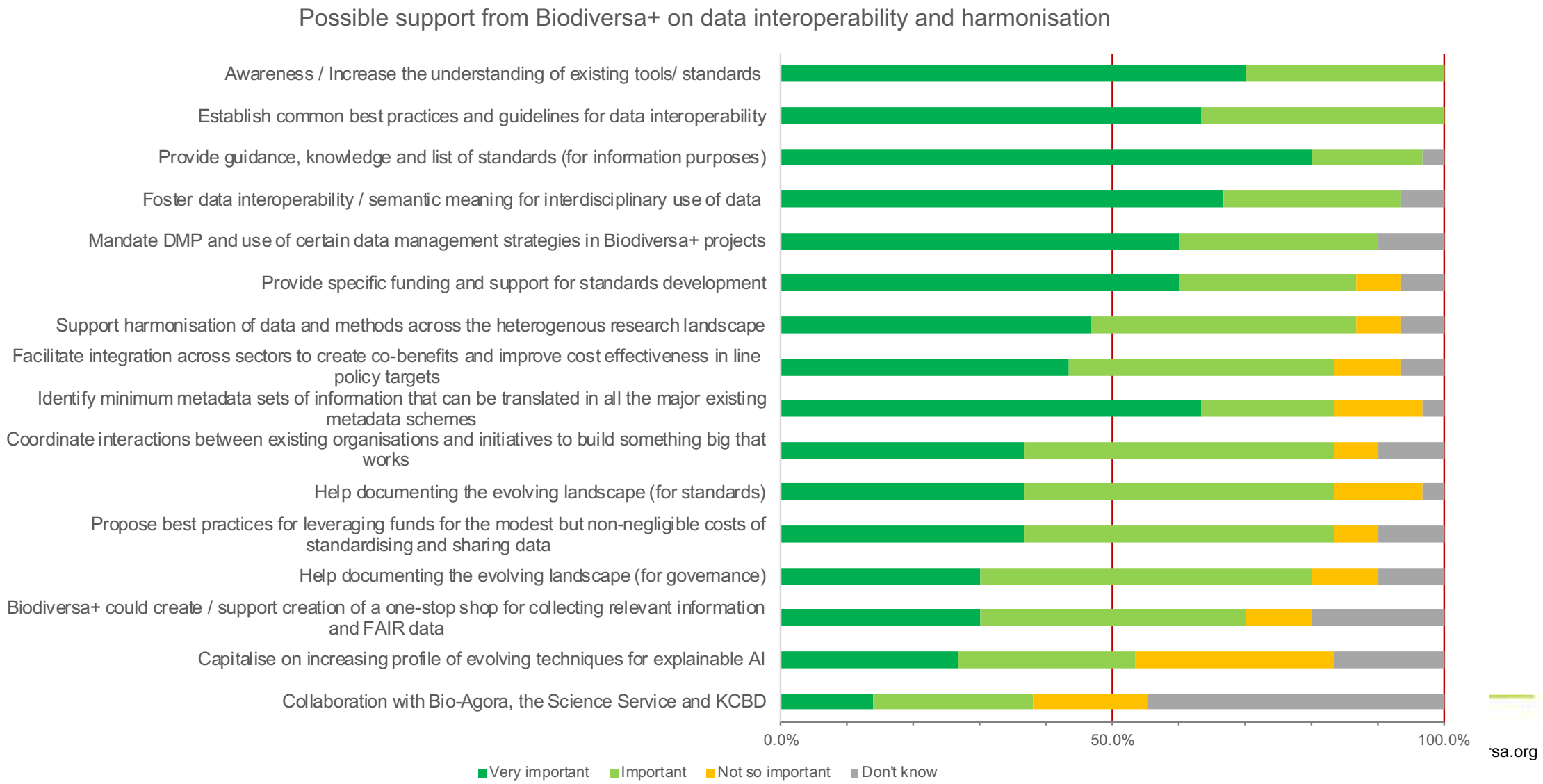
Support	Comments raised by workshop participants
For standards / new technologies	1. **Specific funding** and support for standards development are needed (TDWG, RDA…) ++ 2. **Adherence to the relevance standards** is required in funding programmes, including guidance and a list of standards (for information purposes, the list may not be comprehensive. ++ Foster data interoperability and semantic meaning to allow interdisciplinary use of data + 3. **Mandating data management plans** and use of certain data management strategies within Biodiversa+ projects + Define “playbooks” for data management at different stages of the cycle (collect, analyse, publish…) + 4. Harmonise data and methods across the heterogenous research landscape and link with national, regional (EU) and global institutions ++ 5. Identification of the minimum metadata sets of information that can be translated in all the major existing metadata schemes 6. **New technologies**: Capitalise on increasing profile of evolving techniques for explainable AI with biodiversity and provenance specific funding calls
For best practices, capacity building, knowledge sharing	1. Can propose **best practices for leveraging funds** for the modest but non-negligible costs of standardizing and sharing data, otherwise, this effort is left out of budgets and plans 2. **Establish common best practices and guidelines**. Common guidelines to facilitate **data interoperability** between monitoring data from management and research infrastructures. 3. **Awareness / Increase the understanding of existing tools/ standards** (uptake at the relevant levels) + 4. **Provide knowledge on existing methods and standards** for harmonisation and data interoperability to key actors and stakeholders (technical knowledge) + 5. Biodiversa+ can **provide expertise** in data management and interoperability 6. **Help documenting the evolving landscape**, including documenting key standards (decision tree?), and continue promoting data interoperability and integration (related to standards and governance support)

Outcomes of the first workshop –Biodiversa+ support

Support	Comments raised by workshop participants
Governance	1. Facilitate **integration across sectors** (agri, forest, nature, water…) to create co-benefits and improve cost effectiveness also in line with global and EU policy targets ++++ 2. Biodiversa+ could be a key actor together with the EU science service in make (BioAgora) to see how these infrastructures and biodiversity monitoring could be a pillar under the KCBD. 3. Biodiversa+ could **coordinate interactions between existing organizations** and initiatives to build something big that works. 4. Have a **one-stop shop** for collecting the relevant information needed in terms of monitoring initiatives and FAIR data (related to standards support)

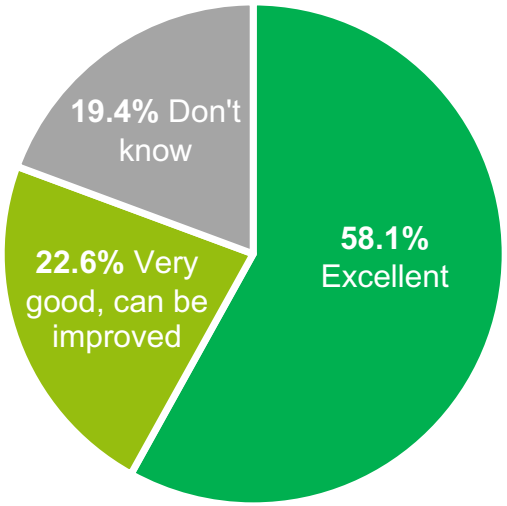
For which activities should Biodiversa+ provide support?

Outcomes of the survey – 31 respondents

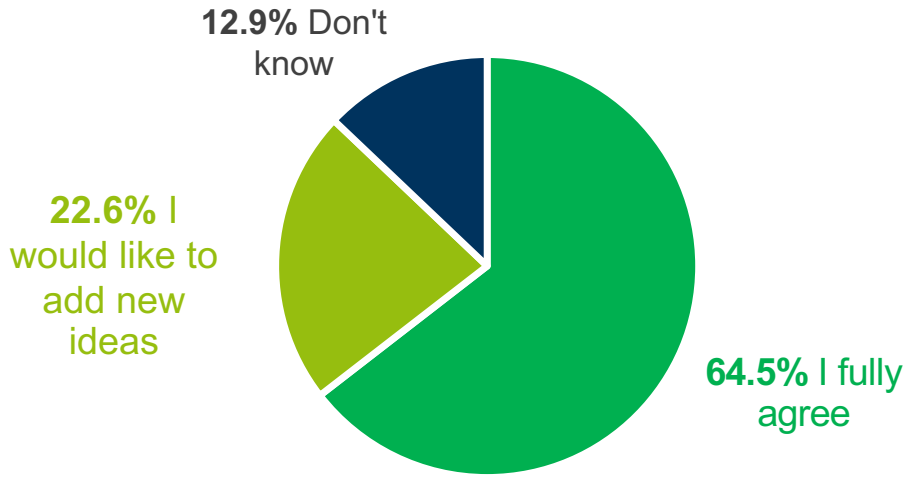


Outcomes of the survey – support for the outcomes of the 1st workshop

How do you consider this summary of data interoperability issues for biodiversity monitoring at the European or international scale?



Do you agree on how Biodiversa+ could provide support to biodiversity monitoring data interoperability?



Next steps after this workshop

- ✓ A public report showcasing these national and global data architectures and workflows will be drawn. The objectives of this report will be to better understand the landscape (including its gaps), identify case studies enabling capacity building and knowledge transfer, identify good practices.

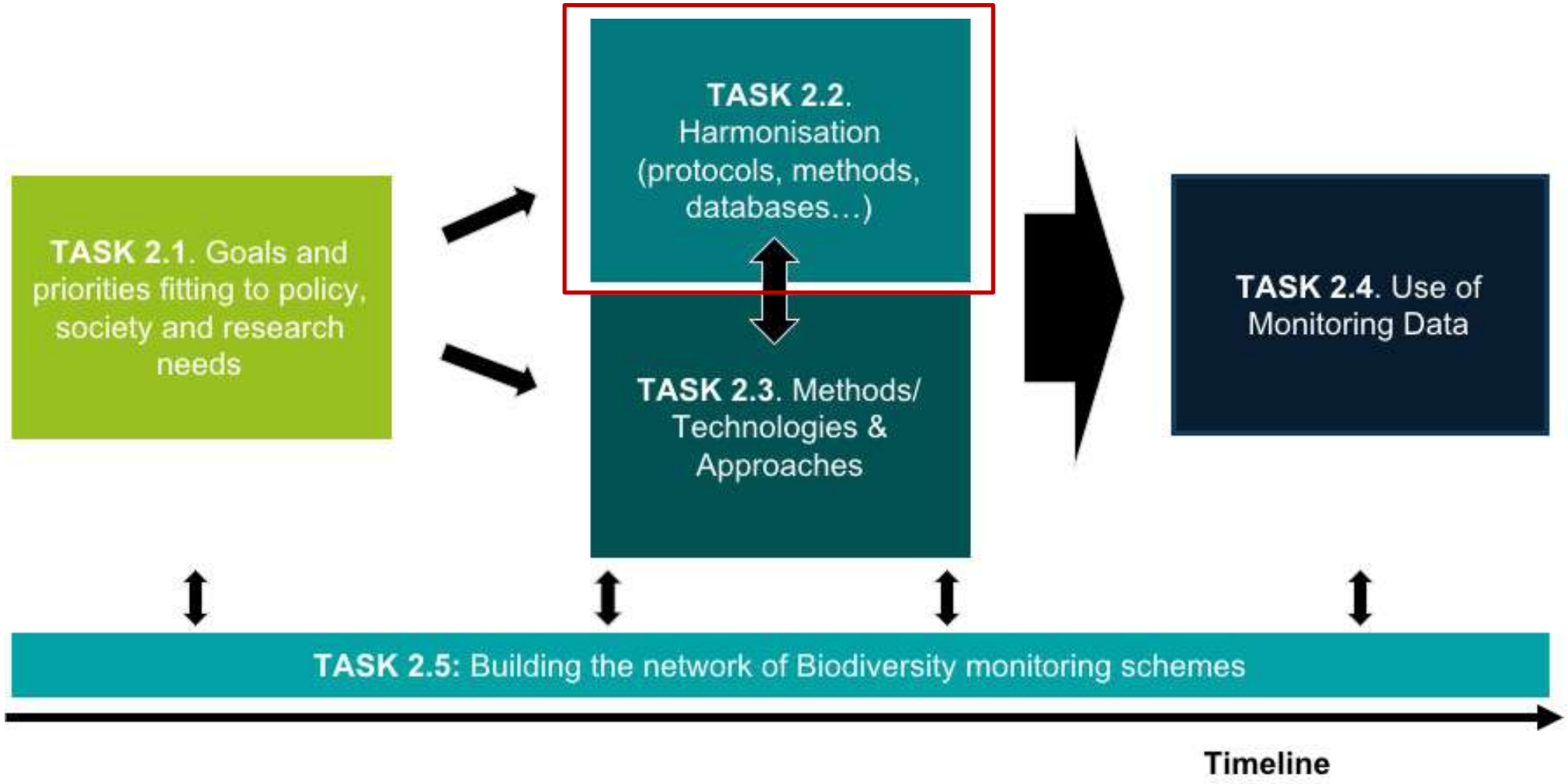
- → Building on relevant EuropaBON activities

biodiversa+
European Biodiversity Partnership

Possible Biodiversa+ support for data interoperability and harmonisation

By Petteri Vihervaara, Biodiversa+ WP2 leader, MoE_FI

Biodiversity monitoring workstreams of activities



Possible support from Biodiversa+

- ✓ Enriching the EuropaBON work in relation to description of data flows (from raw observation, to indicators, to multiple uses);
- ✓ Mapping of (sub)national data architectures => identify best practices;
- ✓ Promoting best practices through capacity building activities;
- ✓ Promoting open data and standards for exchanging information, in line with the EU digitization strategy;
- ✓ Promoting use of European and global research infrastructures; ...
- ✓ Launch of a biodiversity monitoring pilot

biodiversa+
European Biodiversity Partnership

Keynote speech on data interoperability and harmonisation for biodiversity

By Hanna Koivula, CSC – IT Center for Science

CSC

ICT Solutions for Brilliant Minds

Data interoperability and harmonisation for biodiversity

Biodiversa+ WS 2022-11-04

Hanna Koivula

understandable for both, humans and machines

F

FINDABLE

- Essential information described in sufficient detail
- Description page and has a persistent identifier (PID)

A

ACCESSIBLE

- **Can be searched** on the Internet
- **Versioning and life cycle** are documented
- **Tombstone page** if data has been deleted

I

INTEROPERABLE

- Common, documented and **open file formats** are used
- **Data content and constraints** are also interoperable

R

RE-USABLE

- **Data quality is well documented** and understandable
- **Access rights** displayed and machine actionable

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Why FAIR?

- A fundamental principle of the original scientific method i.e. part of so called ”good scientific practice”
- A continnuum that consists of **reproducibility, replicability and re-usability**

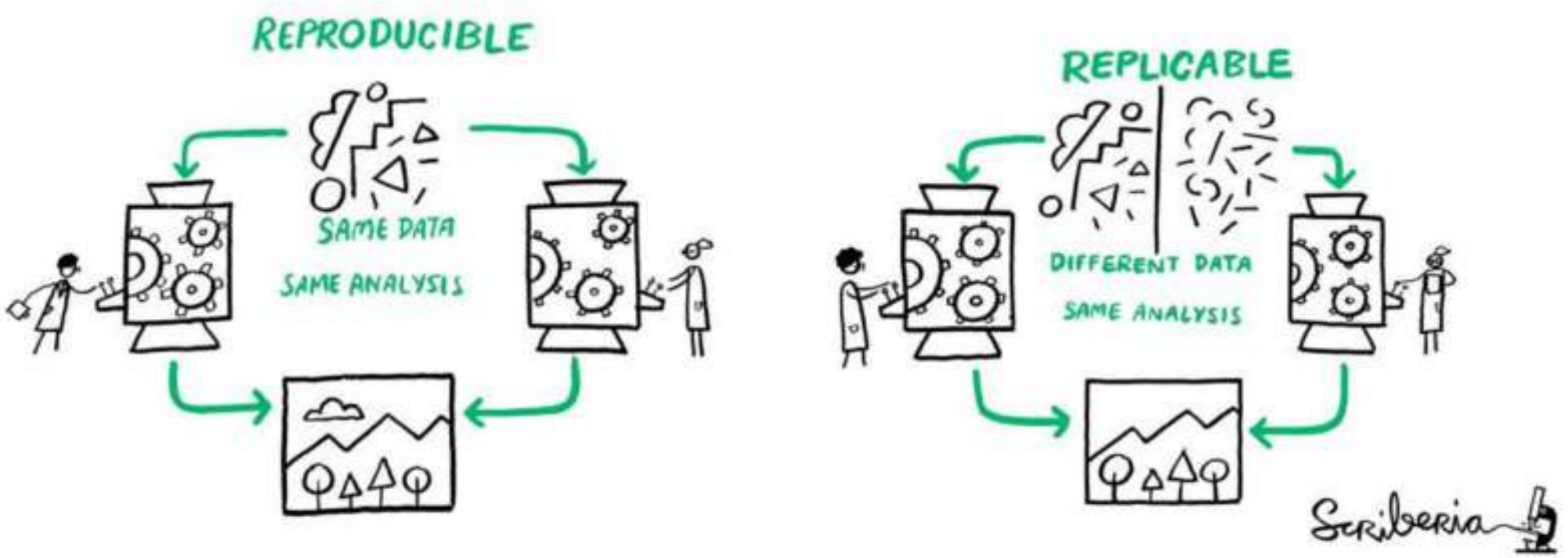
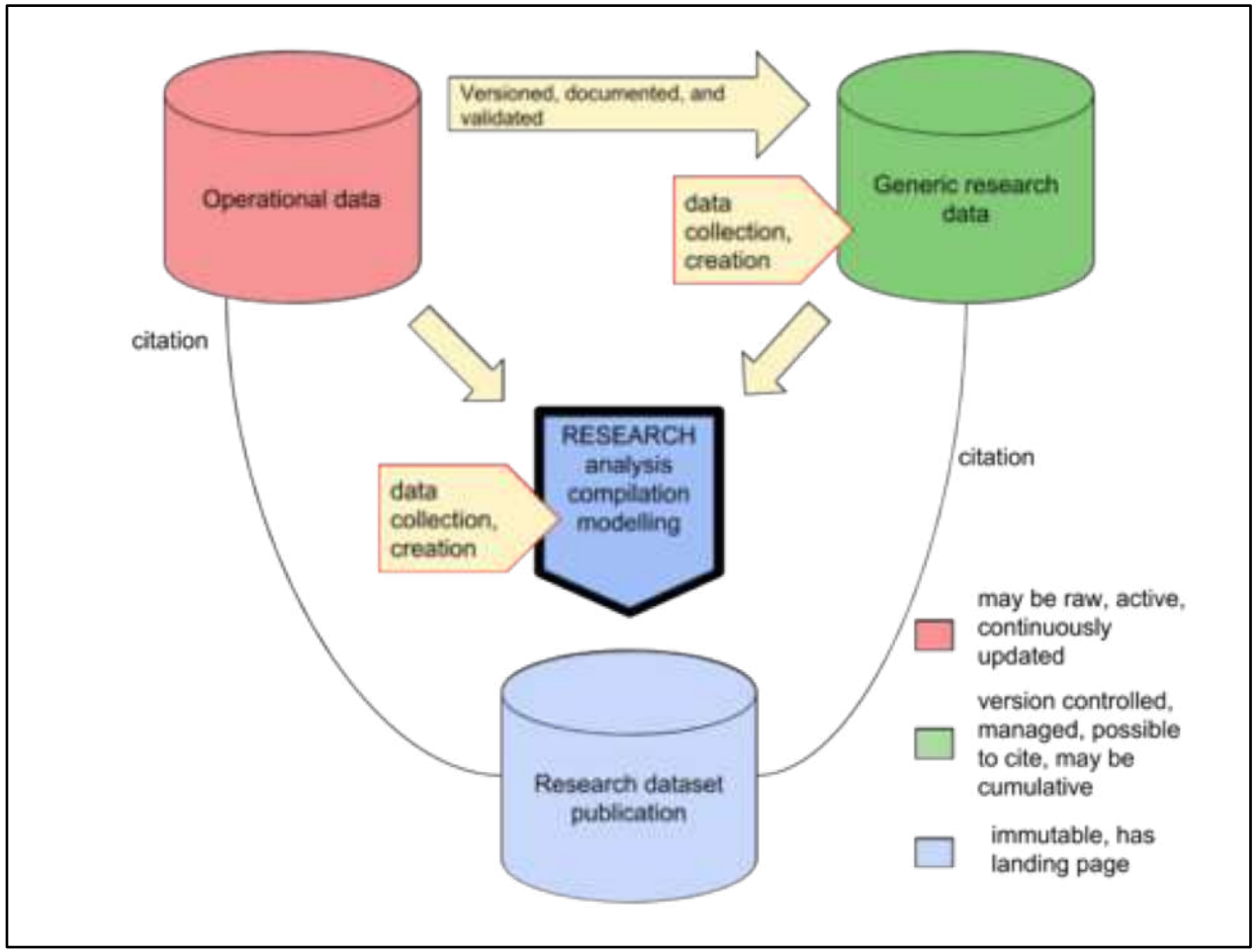


Figure 1. Reproducible and replicable research results[3]

What is research data?



Operational data (original purpose):

- Can be **raw,** active, continnuously updated

Generic research data

- Version controlled, possible to cite, may be cumulative

Research dataset (publication):

- Immutable, has landing page

Data harmonisation and FAIR (Interoperability)

Interoperable data means it can be integrated with other data, applications and workflows.

- This is achieved by using **common metadata and data standards, and *harmonising* data by using semantic artefacts** (i.e. controlled vocabularies, ontologies and thesauri, etc.) **to describe the data variables unambiguously**.
- Likewise, **technical interoperability** is achieved by creating automated workflows **using standards** and **APIs** for data transfer.

Layers of Interoperability

Legal interoperability

Legal framework: Copyright, Licenses, Data Protection, GDPR

Organisational interoperability

Processes, Policies, Governance

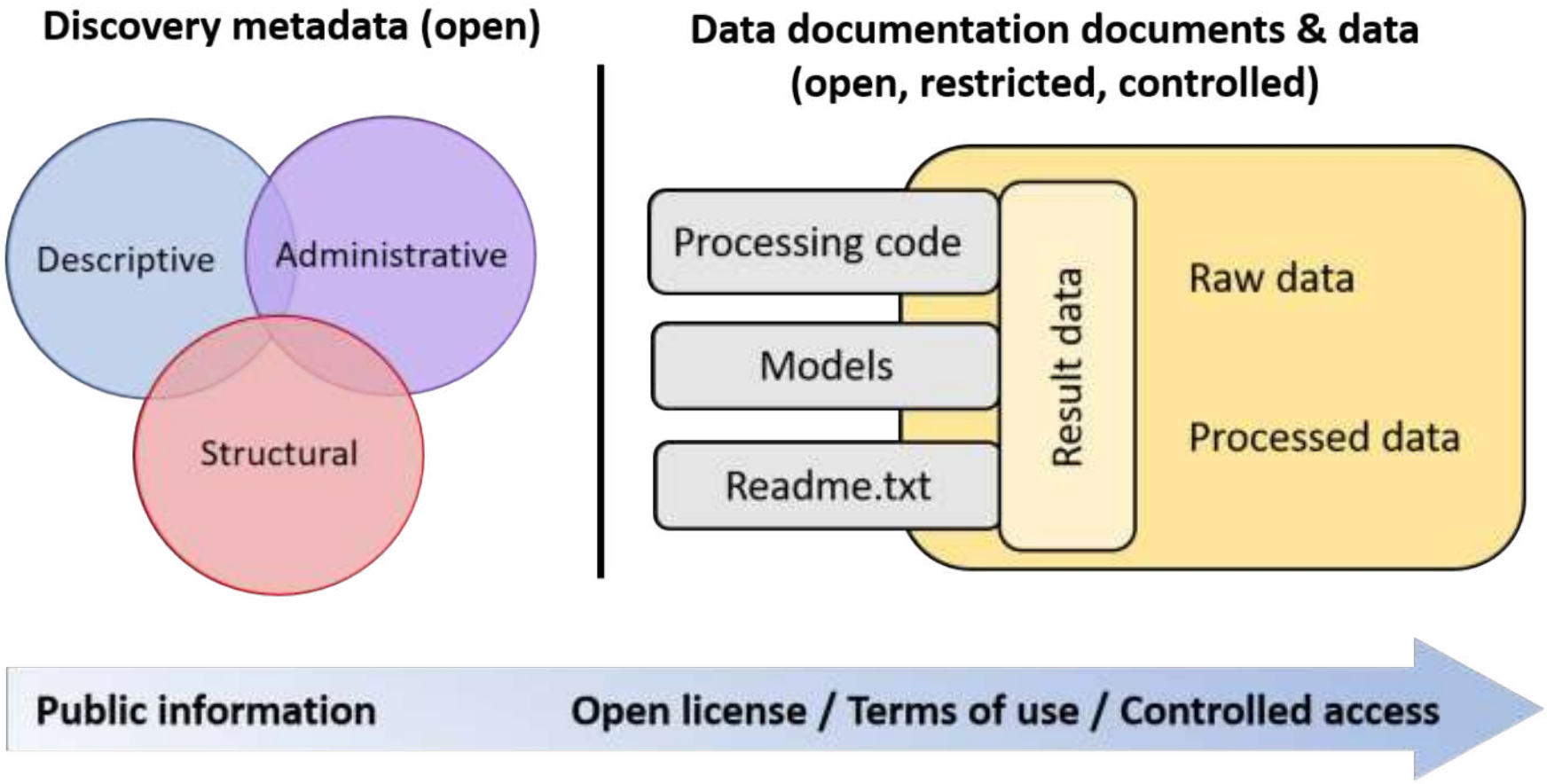
Technical interoperability

Infrastructure, Services, Technologies

Semantic interoperability

Community standards, Metadata & Ontologies, Interpretation of meaning & structure

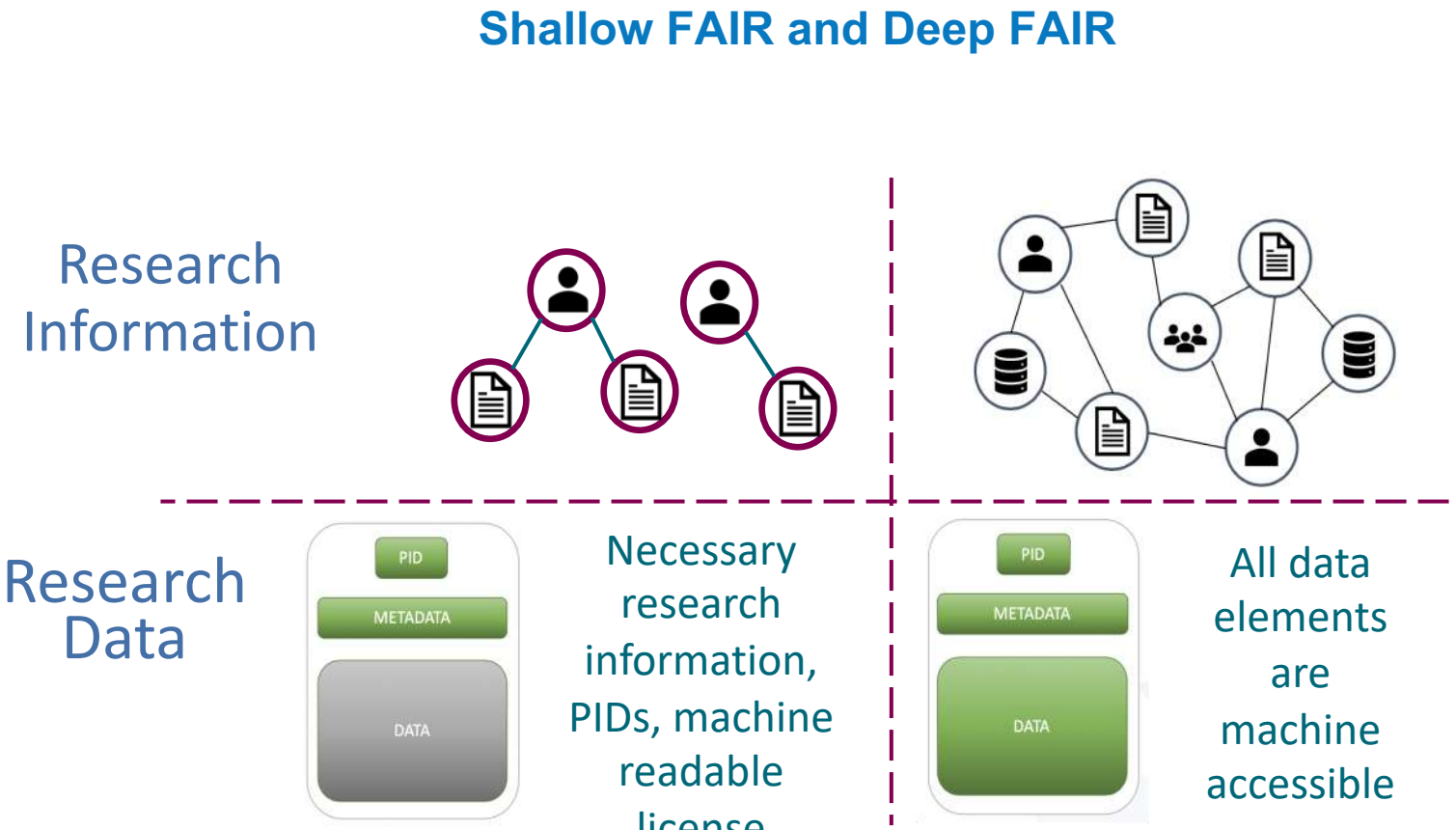
FAIR priciples can be applied to all research output(s)



Data and can be open, restictedly available or controlled. Terms of use (for research) have been described in the administrative metadata and machine readable lisence.

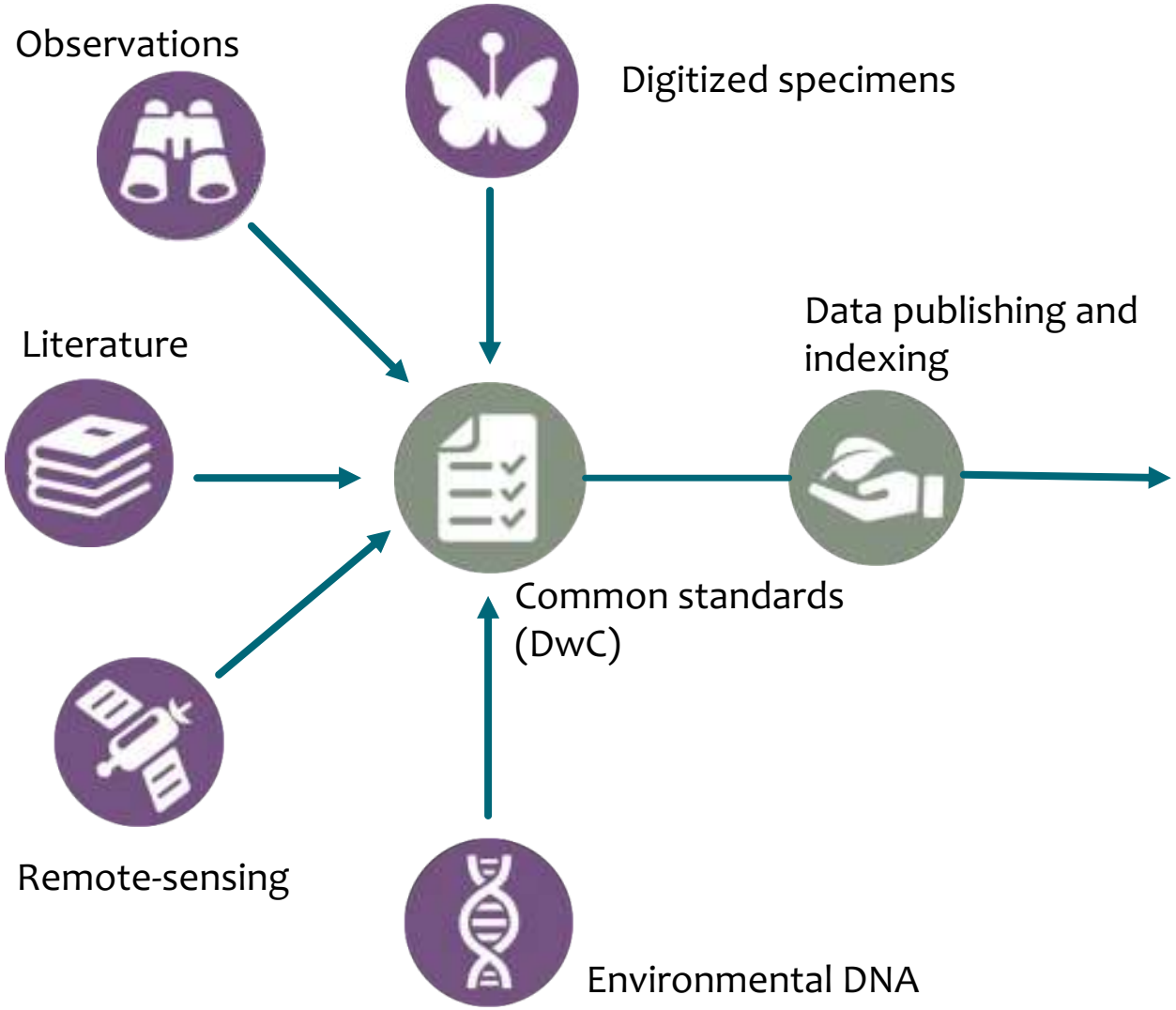
Biodiversity data

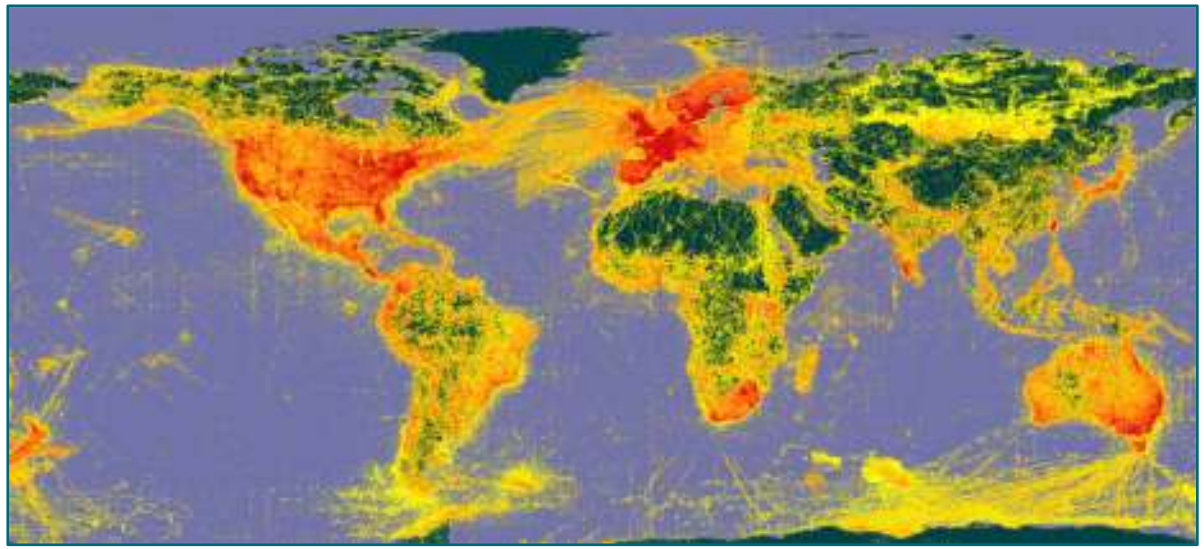
- Biodiversity data are shared and searched on **data-level**
- BD-data requires some level of **"deep FAIR"**
- **Meta-data** is used for describing the *fittness-for-purpose*, but **NOT** so much for *data discovery*
- **Data Quality** is not an absolute measure, but it is **dependent on the purpose** (or re-purpose of data use)



26

Biodiversity data →
A window on evidence about where species HAVE liveD, and when





Data discovery and use

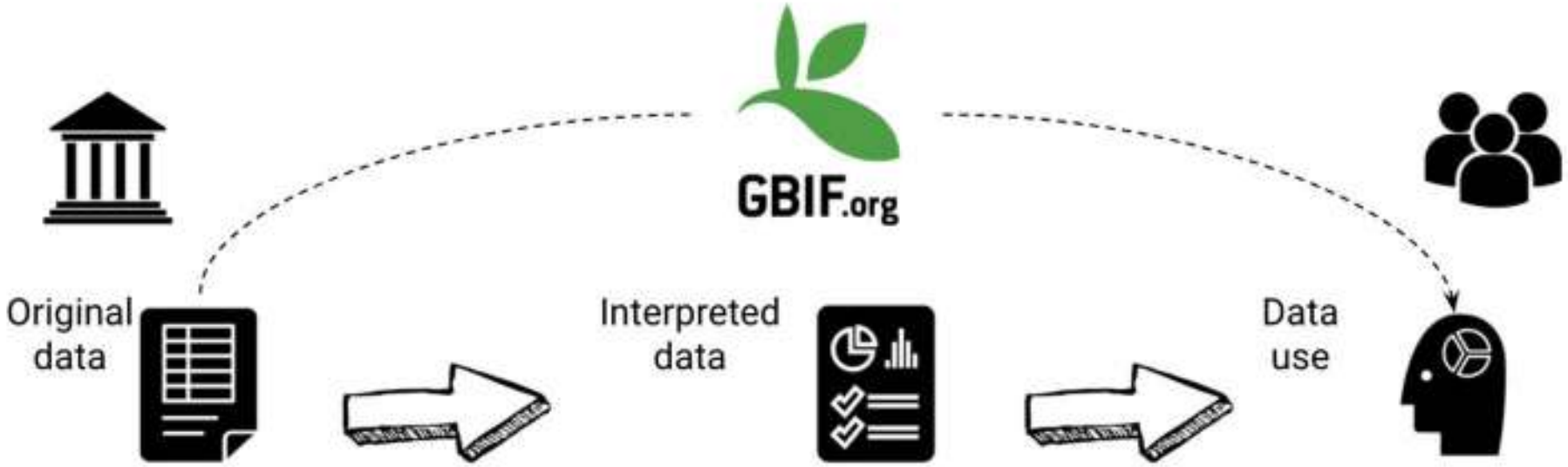
Role of gbif.org as data aggregator

GBIF is a **data aggregator** enabling sharing a multitude of BD data at **data level**

It uses DarwinCore and its extensions to describe data:

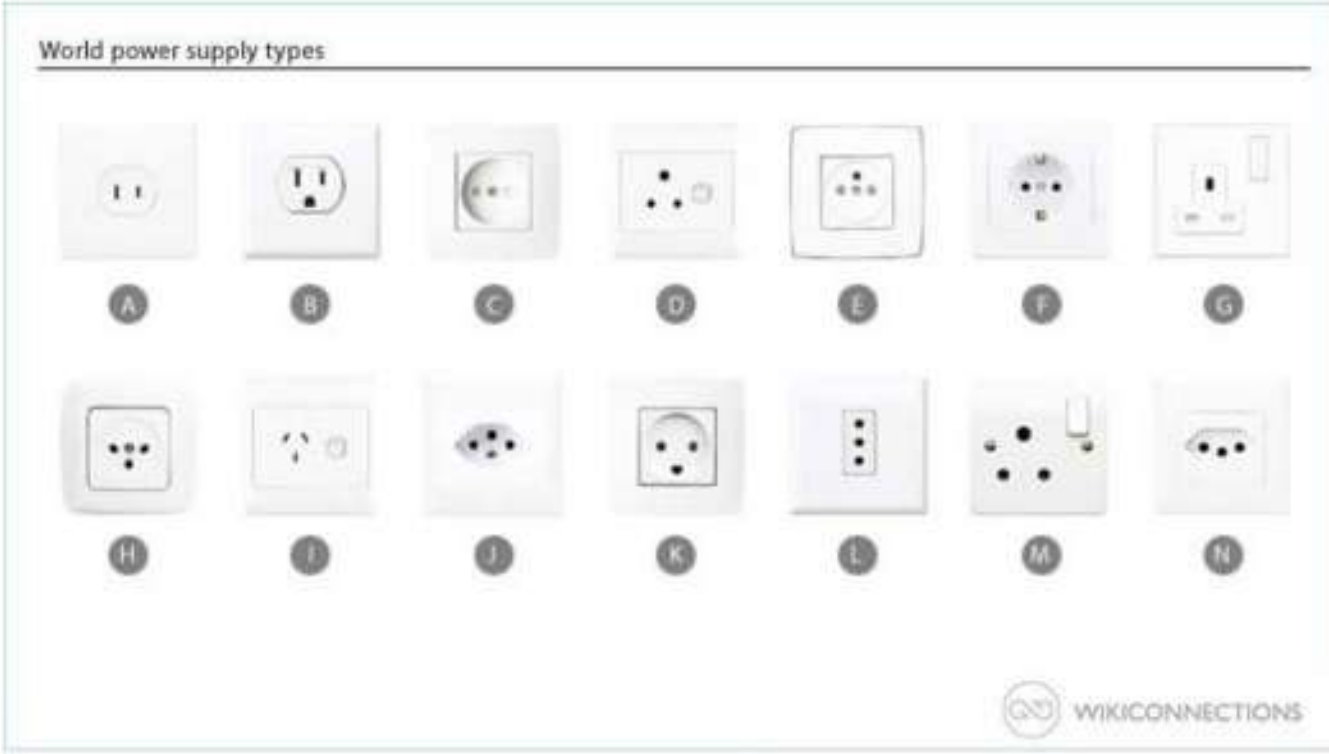
→ **EML (metadata)** is used express methods and environment variables
→ **Metadata is mainly NOT useful for BD data discovery**

Data quality assessment is possible against **standards.**
However, standards do not necessarily produce quality.



Community standards define the structure of data

- **Community standards** like EML (Ecological Metadata Language) and DarwinCore (DwC) and their extensions are needed for **making data available in an understandable (FAIR) format**
- Standards are needed for **data Quality Assurance**
- **Data Quality** is not an absolute measure, but **dependent on the purpose** (or repurpose of data use)
- **Vocabularies are used with standards** to express the data quality **as rich as possible**



Data richness levels supported by Gbif

Dataset description,
taxonomic/geographic/temporal scope

Dataset metadata

List of taxa
regional or thematic (e.g. invasive, medicinal)

Species checklists

Species occurrences
dates, coordinates, basis of record

Occurrence-only data

Species occurrences and sampling events
dates, coordinates, sampling effort / protocol, abundance

Sampling-event data

DwC and extensions

DwC has provided a simple and effective framework for supporting the growth of species occurrence data
Used for expressing different views to the data
Biodiversity information standards (TDWG) community maintains DwC and extensions

GBIF has (currently 34) registered extensions

- DwC Occurrence
- DwC Taxon
- DwC Event
- DwC MeasurementOrFact (and extended measurementOrFact)
- DNA derived data
- Audubon multimedia description
- GBIF Relevé (for vegetation plot surveys)
- Taxon Description, Alternative Identifier
- Trait measurement score, Trait descriptor, Trait measurement trial

TDWG community task groups works with **new extensions:**

- Humbold Core for species inventories
- Ecological survey data exchange specification
- Plinian core (properties or traits related to taxa)
- Camera trap data
- iNaturalist data
- …

https://www.tdwg.org/community/

MeasurementOrFact

measurementID | measurementType | measurementValue | measurementAccuracy | measurementUnit | measurementDeterminedBy | measurementDeterminedDate | measurementMethod | measurementRemarks

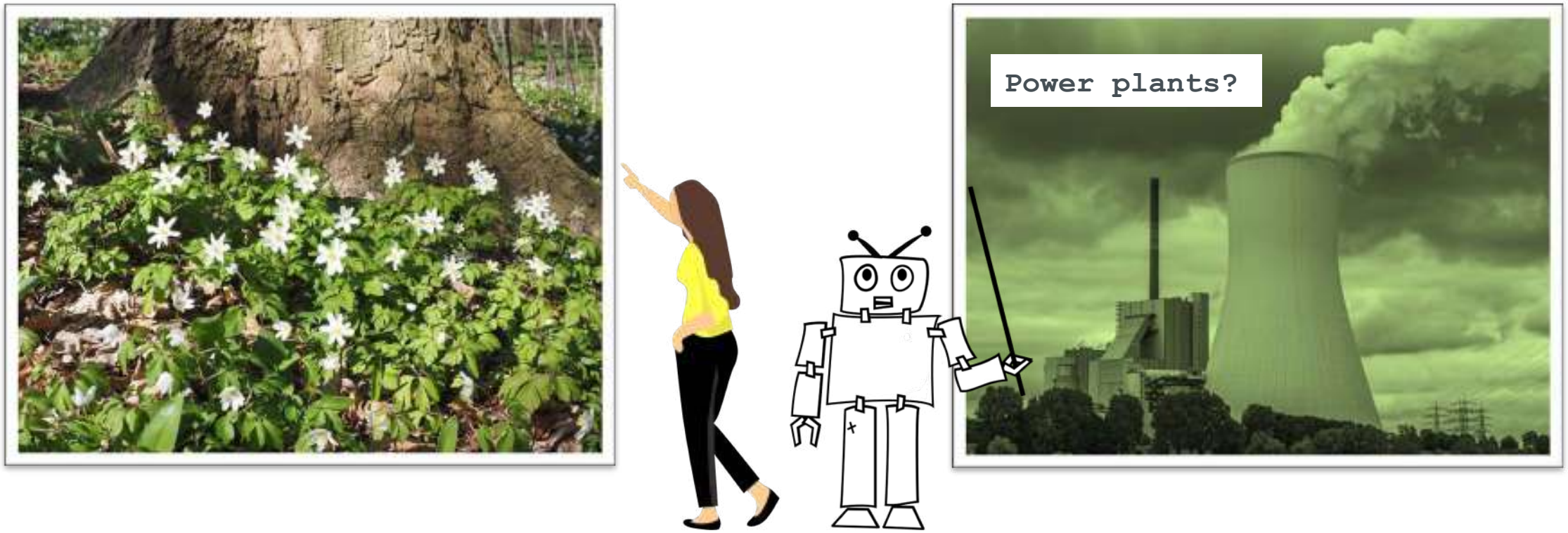
MeasurementOrFact	Class
Identifier	http://rs.tdwg.org/dwc/terms/MeasurementOrFact
Definition	A measurement of or fact about an rdfs:Resource (http://www.w3.org/2000/01/rdf-schema#Resource).
Comments	Resources can be thought of as identifiable records or instances of classes and may include, but need not be limited to dwc:Occurrence, dwc:Organism, dwc:MaterialSample, dwc:Event, dwc:Location, dwc:GeologicalContext, dwc:Identification, or dwc:Taxon.
Examples	The weight of an organism

measurementType	Property
Identifier	http://rs.tdwg.org/dwc/terms/measurementType
Definition	The nature of the measurement, fact, characteristic, or assertion.
Comments	Recommended best practice is to use a controlled vocabulary.
Examples	`tail length` , `temperature` , `trap line length` , `survey area` , `trap type`

measurementValue	Property
Identifier	http://rs.tdwg.org/dwc/terms/measurementValue
Definition	The value of the measurement, fact, characteristic, or assertion.
Comments	
Examples	`45` , `20` , `1` , `14.5` , `UV-light`

Human vs machine

What comes to your mind, when thinking about green plants?



Concepts represent the content on an abstract level.

Term is a label for the human user (string of letters).

Concepts & terms

Concept

http://tun.fi/MX.37879

Scientific term

Anemone nemorosa

Term in English

Wood anemone

Term in Finnish

valkovuokko

Term in Swedish

vitsippa

```
<ns0:scientificName>"http://tun.fi/MX.37879"/</ns0:scientificName>
      <skos:prefLabel xml:lang="sv">vitsippa</skos:prefLabel>
      <skos:prefLabel xml:lang="fi">valkovuokko</skos:prefLabel>
      <skos:prefLabel xml:lang="la">Anemone nemorosa</skos:prefLabel>
```

Concepts & terms

Term in Finnish

vuokot

Scientific term

Anemone

Term in Swedish

sippor

https://laji.fi/taxon/MX.37878

Anemone genus

Concept

Concept

http://tun.fi/MX.37879

Scientific term

Anemone nemorosa

Term in English

Wood anemone

Term in Finnish

valkovuokko

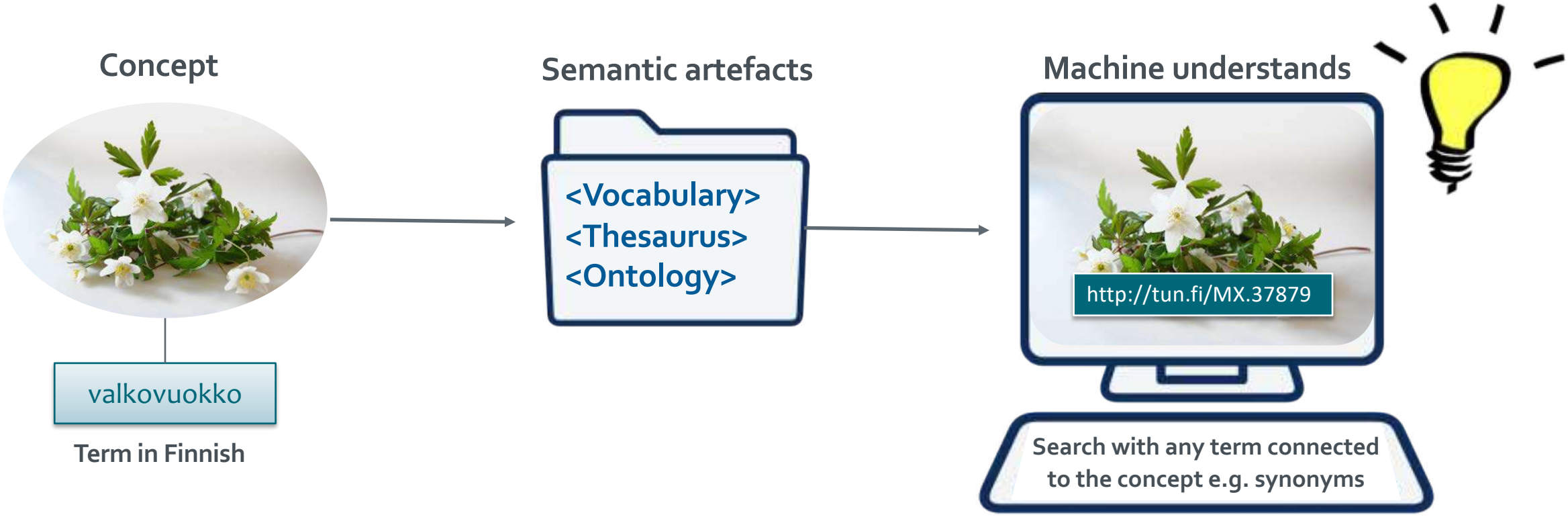
Term in Swedish

vitsippa

```
<ns0:scientificName>"http://tun.fi/MX.37879"/</ns0:scientificName>
      <skos:prefLabel xml:lang="sv">vitsippa</skos:prefLabel>
      <skos:prefLabel xml:lang="fi">valkovuokko</skos:prefLabel>
      <skos:prefLabel xml:lang="la">Anemone nemorosa</skos:prefLabel>
```

Semantic artefacts

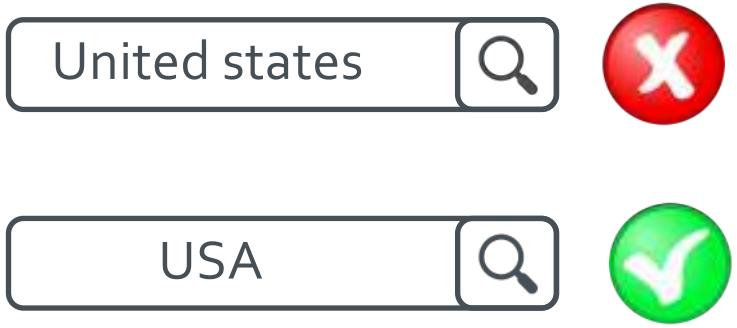
= machine actionable models such as controlled vocabularies or ontologies, which make terms understandable also for machines.



Low semantics: Lists & controlled vocabularies

List of words
from which the computer can retrieve information

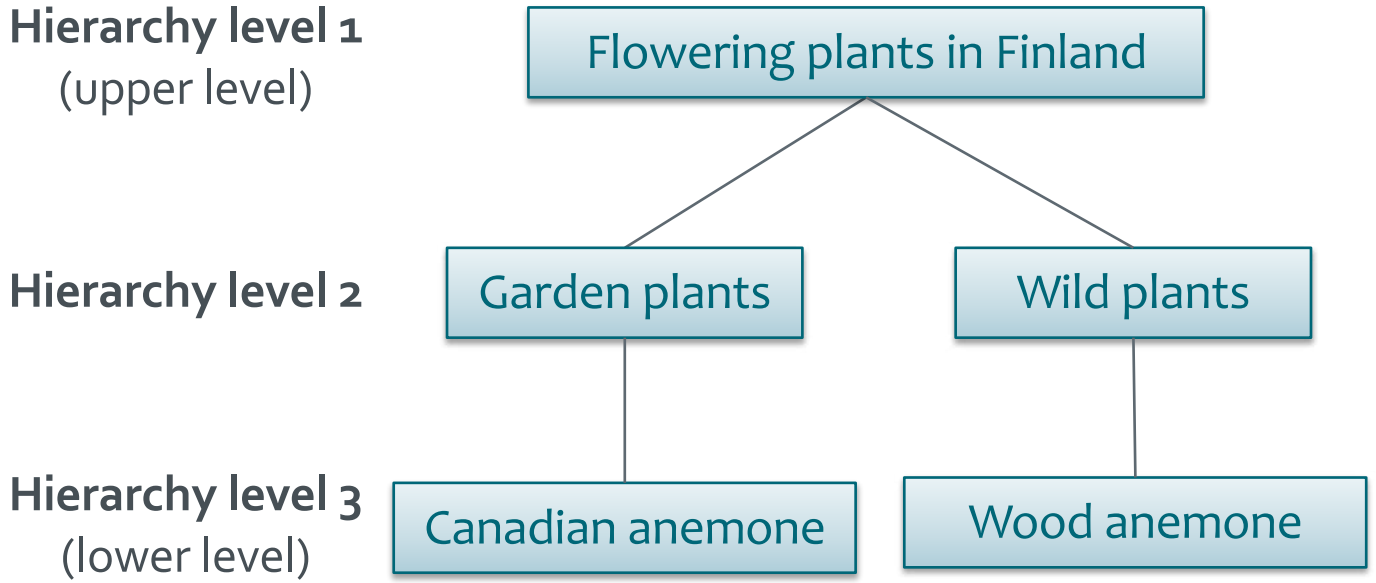
Search with computer:
You can only get a match when the letters match.

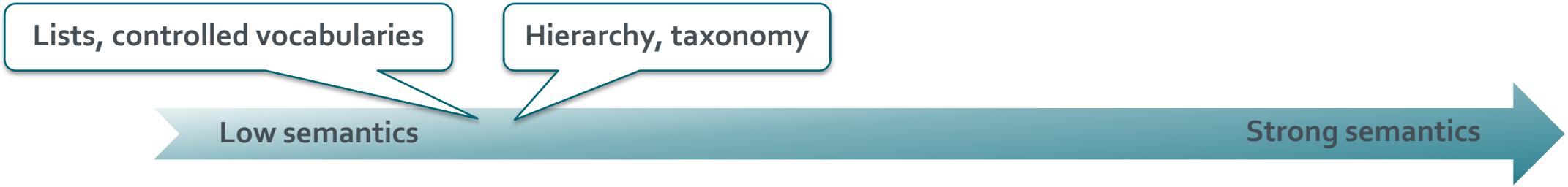


Lists, controlled vocabularies

Low semantics → Strong semantics

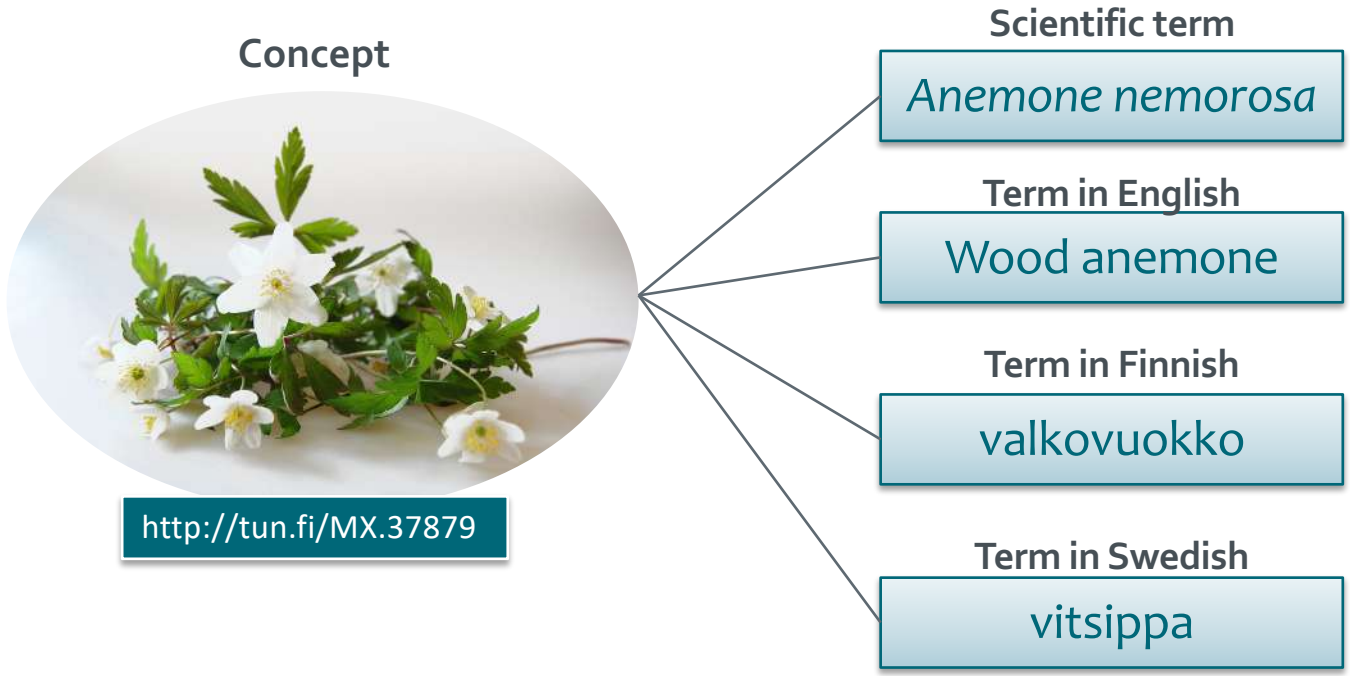
Low semantics: Hierarchies & taxonomies

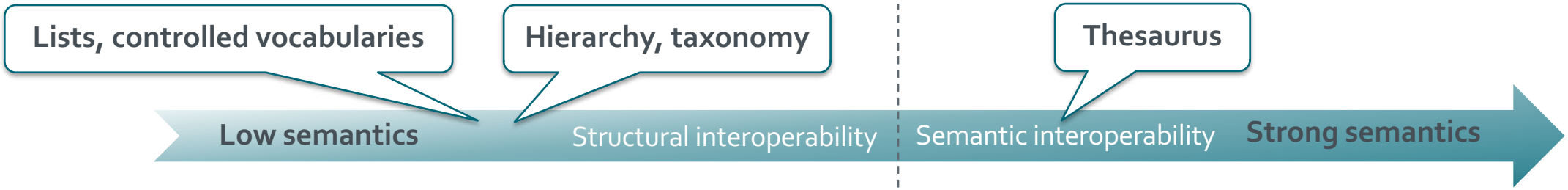




Stronger semantics: Thesaurus

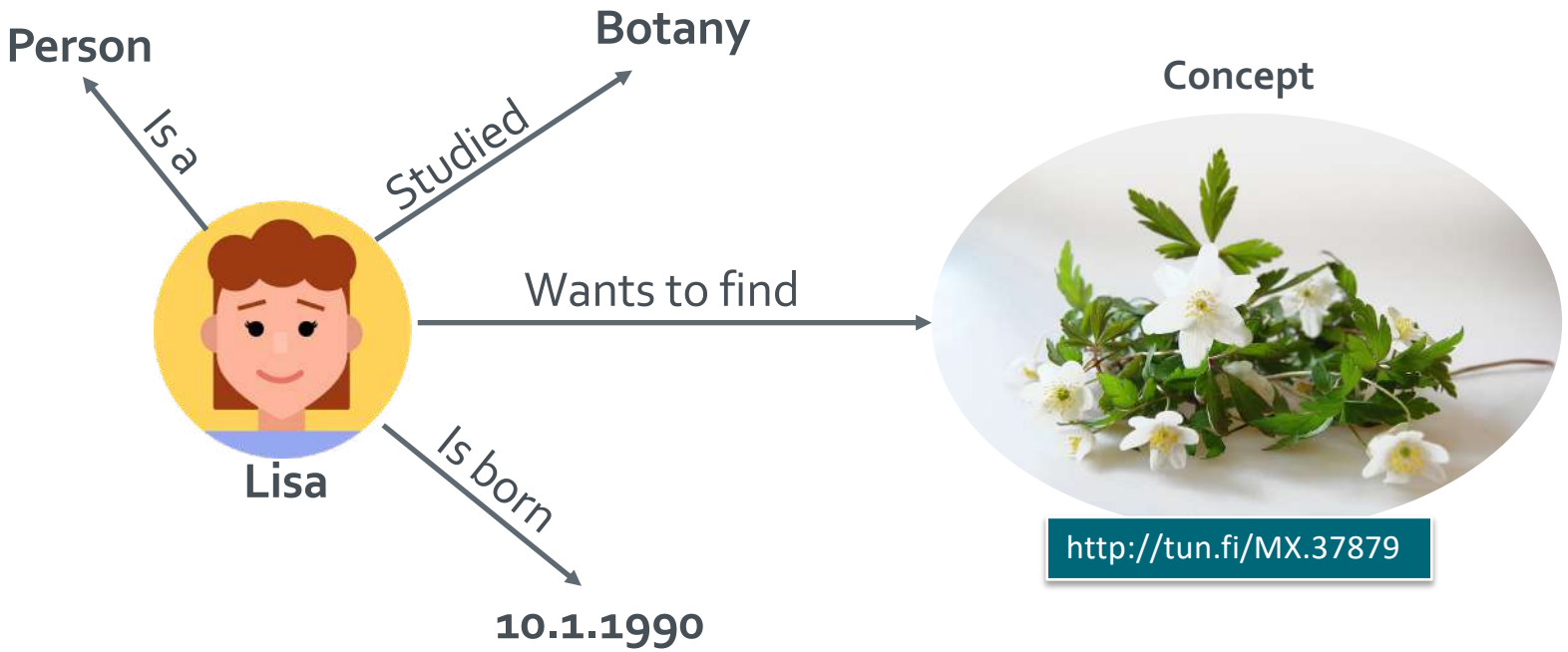
Relations between concepts are modelled

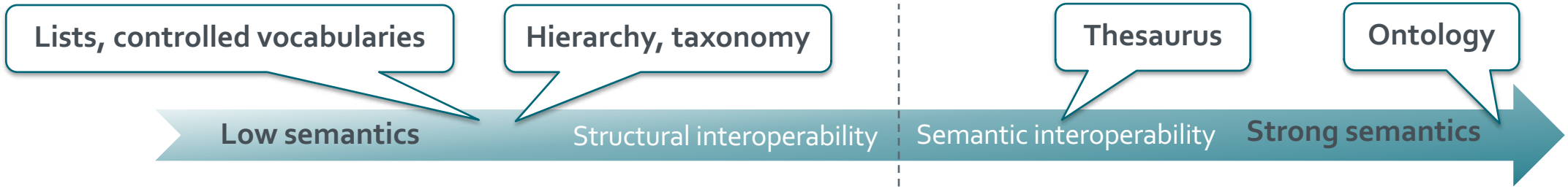




Strong semantics: Ontology

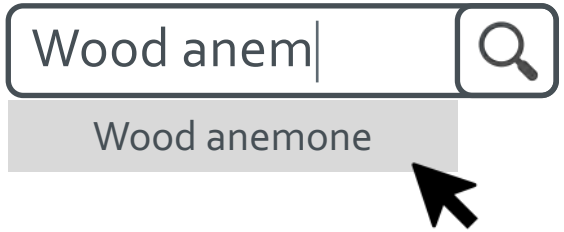
Relations between concepts are modelled





Semantic interoperability has many benefits!

✓Predictive text input



✓Intelligent searches

Others were also looking at

The common hepatica

Coltsfoot

Vocabularies important in BD data

- Locality information that are INSPIRE compatible
- Habitat types / land cover (from Copernicus https://land.copernicus.eu/pan-european/corine-land-cover)
- Taxon check-lists (GBIF taxon backbone DOI10.15468/39omei)
- EnvThes for variables (https://vocabs.lter-europe.net/envthes/en/)

Tools help creating readily harmonized data

Pros:

→Can be used to validate input data

→Help interpriting or capturing locality information

→Use the right check-list for taxons

→Help input of variables by offering suitable vocabularies

→Using local or regional reference data (for example habitat types)

→May draw metadata automatically from sensors or image files

Cons:

→Erronous validation can create biases

→Original coorinates and accuracy should not be dicharged, even if the tool offers interpritted and harmonised data

→Automated lists may force to give falce acurates ie. bias uncertain observations

→Original data is needed for mapping to other reference data

It is important to always keep raw data raw → harmonize only for interpritations and reporting


facebook.com/CSCfi


twitter.com/CSCfi


linkedin.com/company/csc---it-center-for-science


github.com/CSCfi

biodiversa+
European Biodiversity Partnership

The “Meetnetten” Flemish webtool – data architecture and data workflows

By Dimitri Brossens, INBO

Meetnetten.be

data information on a number of priority plant and animal species

dimitri.brosens@inbo.be

Vlaanderen
is wetenschap

vlaanderen.be/inbo

Meetnetten.be

- A suite of monitoring networks
- For collecting high-quality information on about 78 priority plant and animal species.
  - species on which Flanders has to report on
    - habitat and bird directives
  - species that are important for the Flemish nature policy
    - species protection plan available
- Cooperation between
  - Agency for Nature and Forest
  - Natuurpunt
  - Research Institute for Nature and Forest

Meetnetten.be

- Since 2016, operational since 2018
- Data collected by trained citizen scientists and professionals
- No casual observations
  - although important, they don’t do the trick for exhaustive monitoring and reporting
- Developed by Zostera

The Monitoring Networks

- Fixed sample locations
- Target species are counted based on standardized protocols.
- Data collection relies mainly on specialized volunteers, coordinated by the NGO Natuurpunt Studie.
- Fieldwork is planned and monitored with the web tool https://meetnetten.be, which is also used for entering the collected data.

The website

Welkom bij Meetnetten.be

De Vlaamse overheid wil door middel van meetnetten kwaliteitsvolle informatie verzamelen over een aantal prioritaire plant- en diersoorten. Dat zijn soorten waarover Vlaanderen moet rapporteren aan Europa in het kader van de Habitat- en Vogelrichtlijn, maar ook andere soorten die van belang zijn voor het Vlaamse natuurbeleid.

In een meetnet worden gegevens ingezameld volgens een strikt vastgelegde methodiek. De te bezoeken locaties liggen vast. Meetnetten.be is de webtool voor de planning en opvolging van dit veldwerk en voor de invoer van de veldgegevens op een laagdrempelige en efficiënte manier.

Hoe kan ik meedoen?

1. Registreer je hier.
2. Meld je aan voor een meetnet.
3. Kies één of meerdere telocaties in je buurt.

De meetnetcoördinator zal je contacteren voor verdere afspraken. De monitoring gebeurt in nauwe samenwerking met Natuurpunt, dat de vrijwilligers binnen dit project zal coördineren. Iedereen met een goede kennis van de soort in kwestie kan zich aanmelden voor deelname aan een meetnet. Voor sommige meetnetten is gespecialiseerd materiaal nodig. Dit materiaal en de nodige vergunningen worden ter beschikking gesteld van de medewerkers. Waar nodig voorzien we ook opleiding.

Tijdens het veldwerk moeten soms privé-terreinen betreden worden. Daarom werd er ook een afsprakenkader en gedragscode opgesteld.

In de loop van de komende jaren starten we regelmatig nieuwe meetnetten op en voegen ze toe aan deze website. Een overzicht van alle geplande meetnetten vind je in de blauwdrukken.

Met vragen kan je altijd terecht bij de coördinatoren of via info@meetnetten.be.

Meer info? Zie http://www.natuurpunt.be/vaak-gestelde-vragen-soortenmeetnetten

Projecten

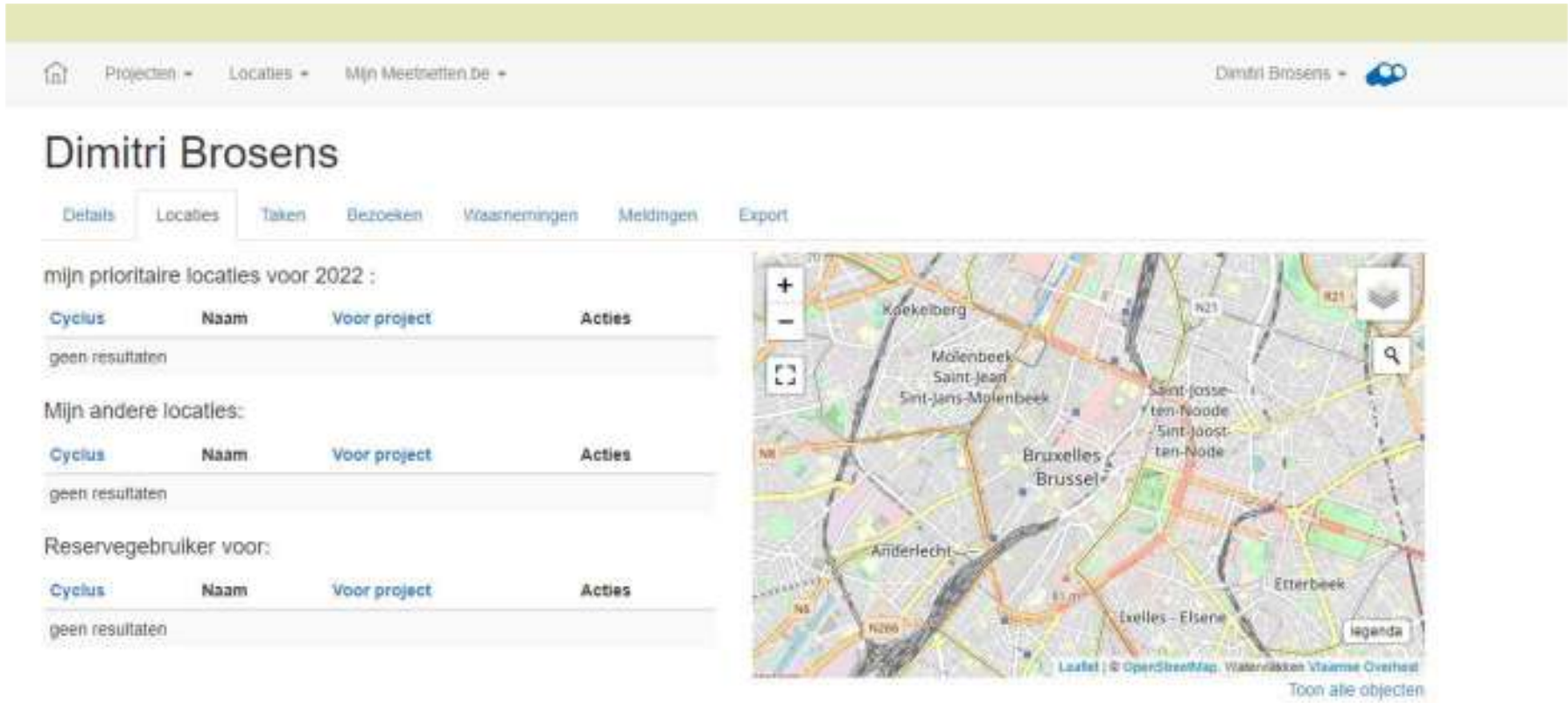
Amfibieën | Libellen | Planten | Dagvlinders | Vleermuizen | Andere zoogdieren | Mollusken | Vogels | Overige

- Boomkikker
- Heikikker
- Kamsalamander
- Knoflookpad
- Poelkikker
- Rugstreeppad
- Vroedmeesterpad

Meetnetten.be

NATUURPUNT | INSTITUUT NATUUR- EN BOSONDERZOEK | AGENTSCHAP NATUUR EN BOS

The dashboard



Meetnetten.be

NATUURPUNT | INSTITUUT NATUUR- EN BOSONDERZOEK | AGENTSCHAP NATUUR EN BOS

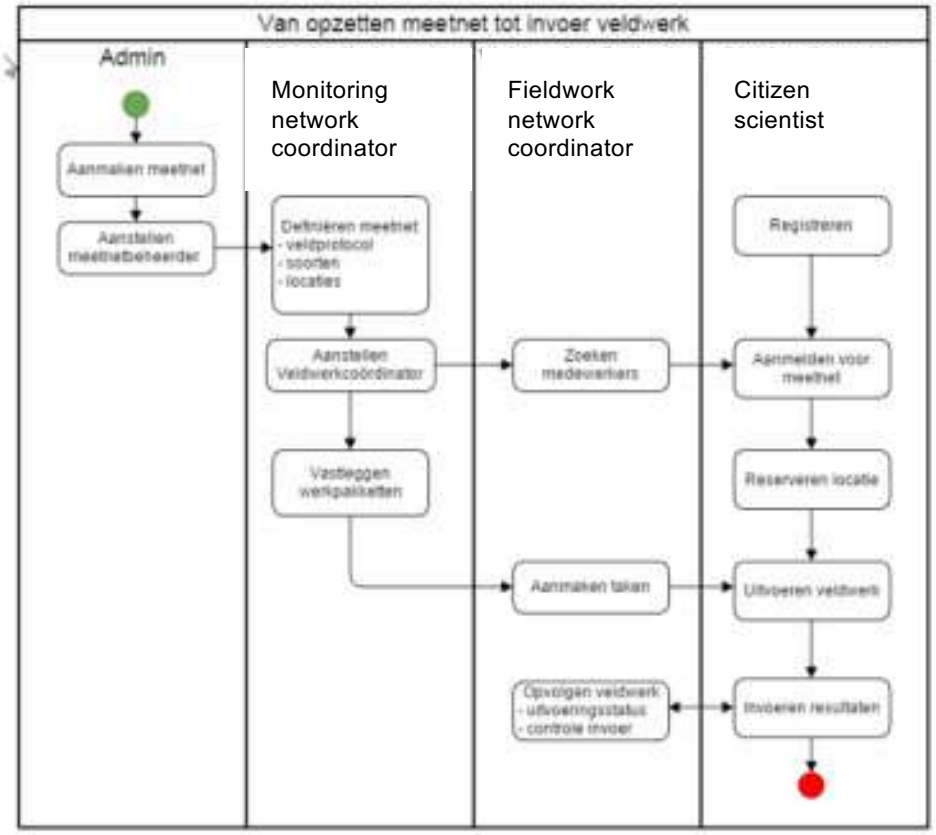
The Goal

- Longtime trend for the number of individuals (population size) for Flanders
- Distribution of species in Flanders (+ casual observations)
- Create valid complex models based on a standardized field protocol

Based on

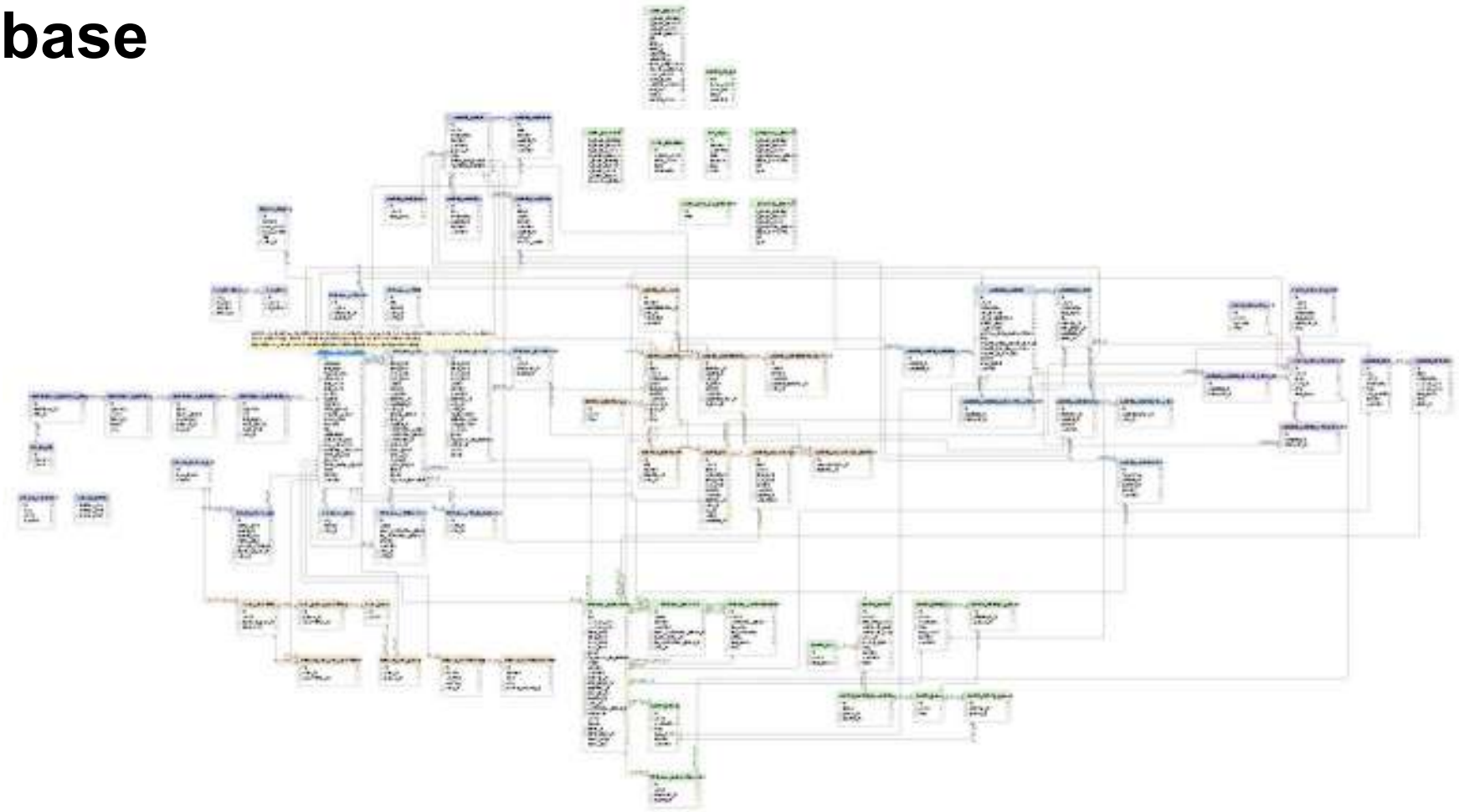
- Defined count period
- Defined count frequence
- Defined time
- Defined locations

How it works



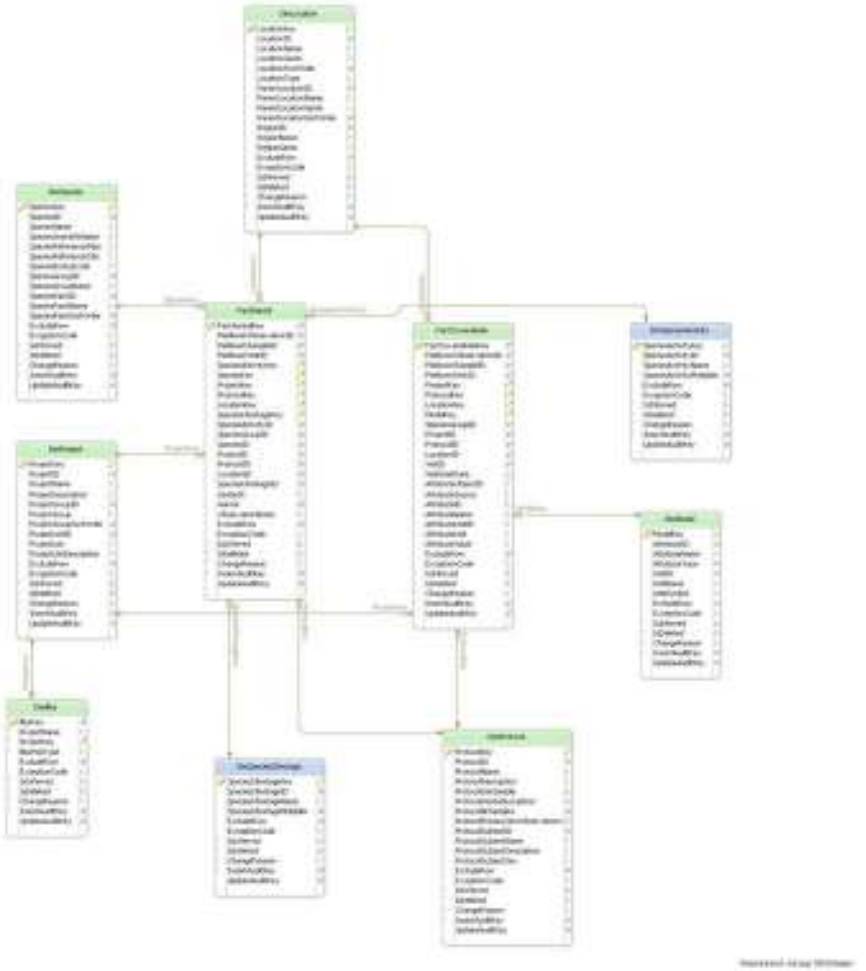
- from building a monitoring network to data import

The Database



From database to dissemination and publication

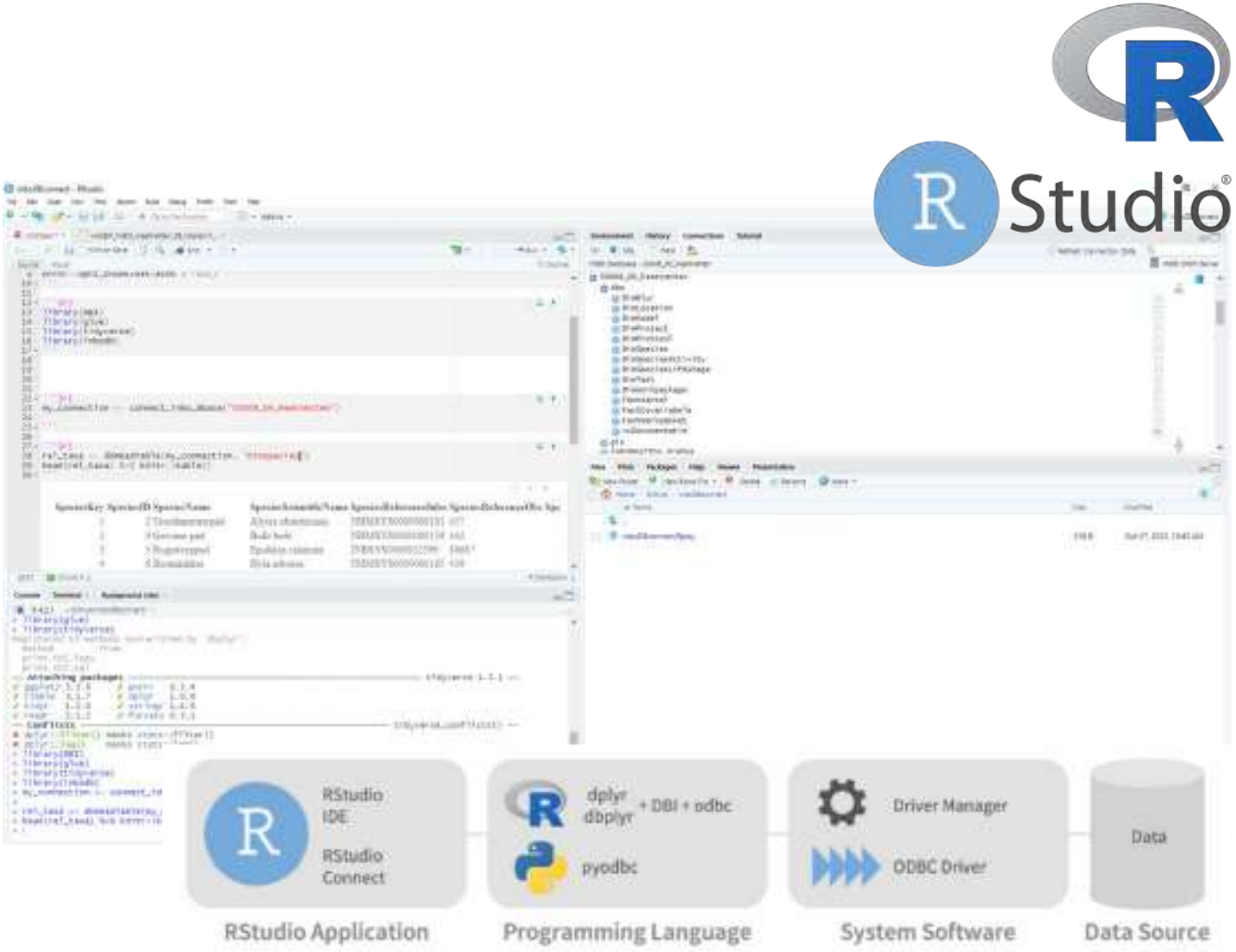
- Dissemination in the **Research Institute**
  - Data Warehouse
    - Star Scheme(easy to query)
  - Rstudio - R
  - ODBC - Access
- Dissemination to our partner **ANB**
  - Mapservice (High resolution)
- Datapublication to **GBIF**
  - IPT / Darwin Core (Lower resolution)



From database to data dissemination in INBO

- R & R studio
- R Packages

```{r}
library(DBI)
library(glue)
library(tidyverse)
library(inbodb)
```



From database to data dissemination INBO

- Access to the database in R

```{r}
my_meetnetten <- connect_inbo_dbase("S0008_00_Meetnetten")
```

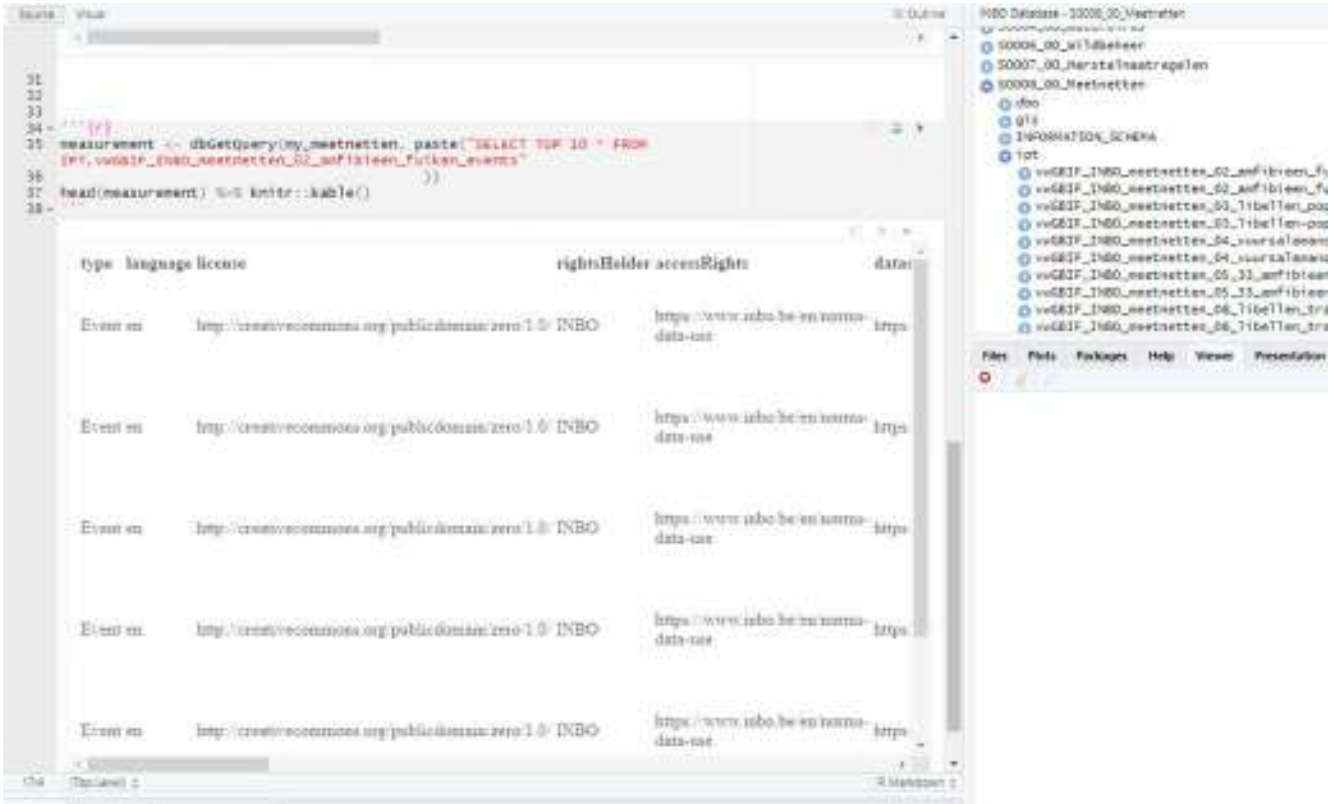
```{r}
rel_taxa <- dbReadTable(my_meetnetten, "DimSpecies")
head(rel_taxa) %>% knitr::kable()
```

SpeciesKey	SpeciesID	SpeciesName	SpeciesScientificName	SpeciesReferenceInbo	SpeciesR
1	2	Vroedmeesterpad	Alytes obstetricans	NHMSYS0000080161	457
2	4	Gewone pad	Bufo bufo	NHMSYS0000080159	442
3	5	Rugstreeppad	Epidalea calamita	INBSYS0000012399	80687
4	6	Boomkikker	Hyla arborea	NHMSYS0000080165	439
5	7	Alpenwatersalamander	Ichthyosaura alpestris	INBSYS0000012389	438
6	8	Vinpootsalamander	Lissotriton helveticus	INBSYS0000012390	456

From database to dissemination in INBO

- RUN SQL queries in R

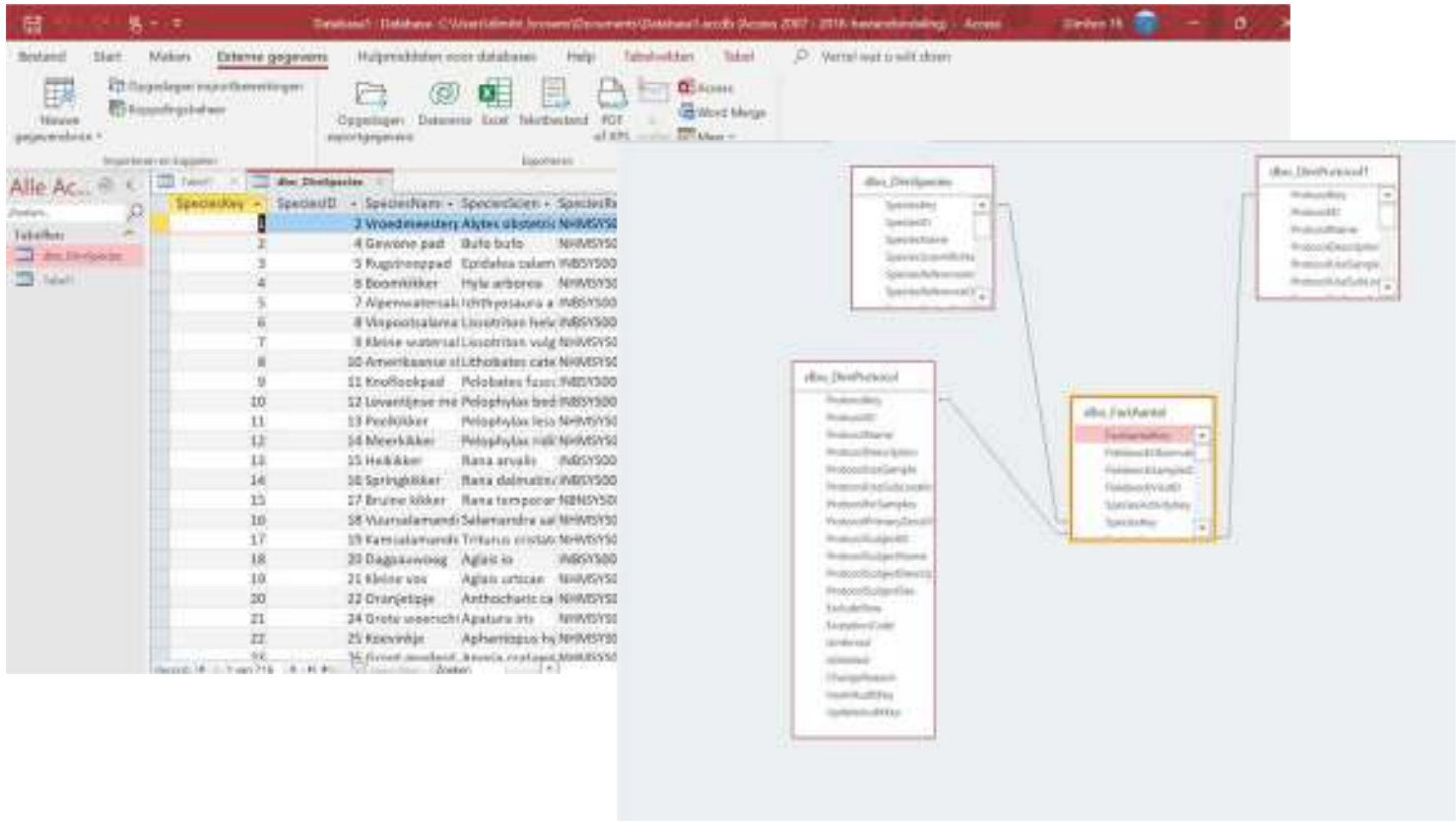
```{r}
measurement <- dbGetQuery(my_meetnetten, paste("SELECT TOP 10 * FROM IPT.vwGBIF_INBO_meetnetten_02_amfibieen_fuiken_events"
                                               ))
head(measurement) %>% knitr::kable()
```



From database to dissemination in INBO

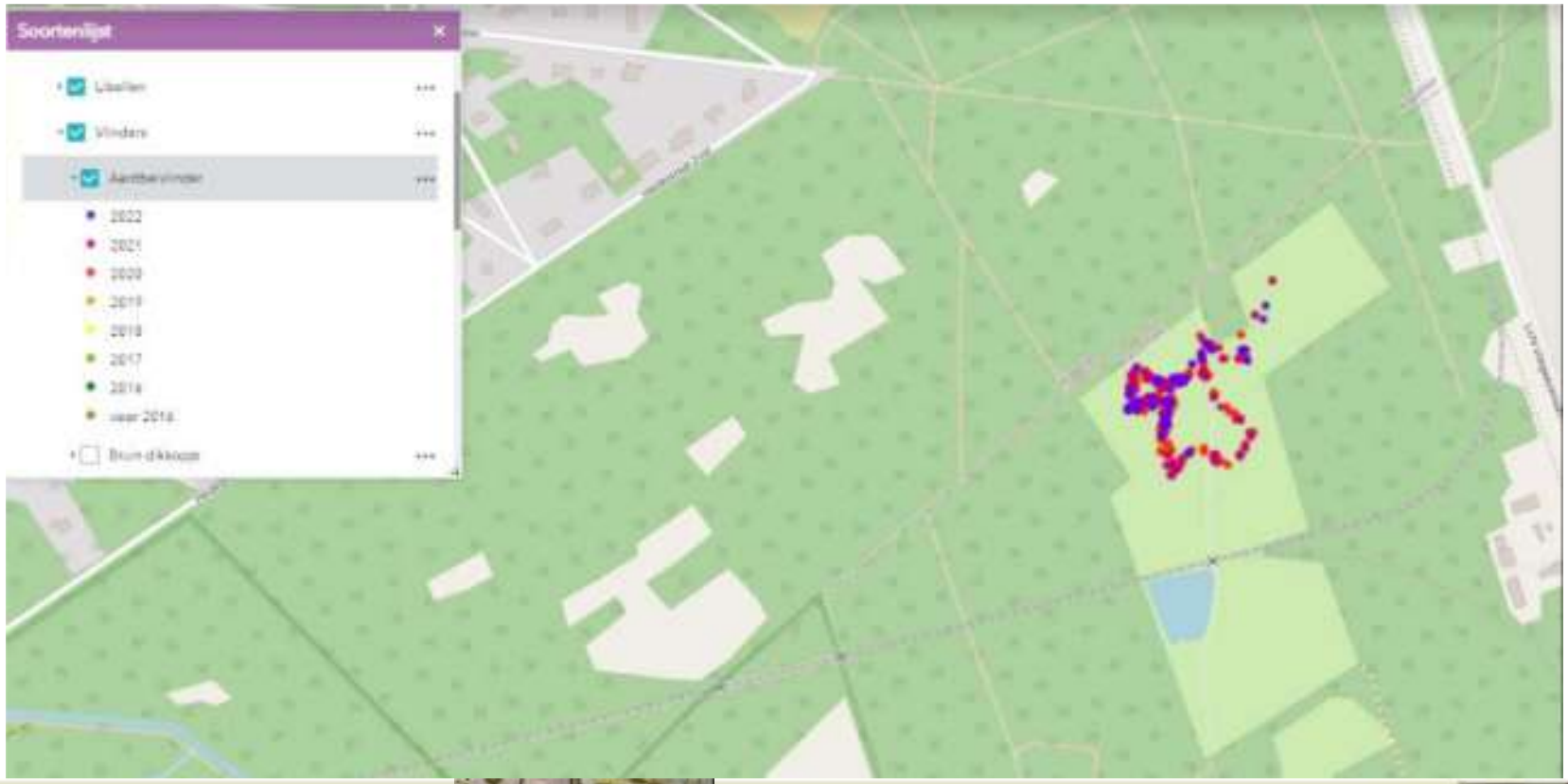
ACCESS

- ODBC connection



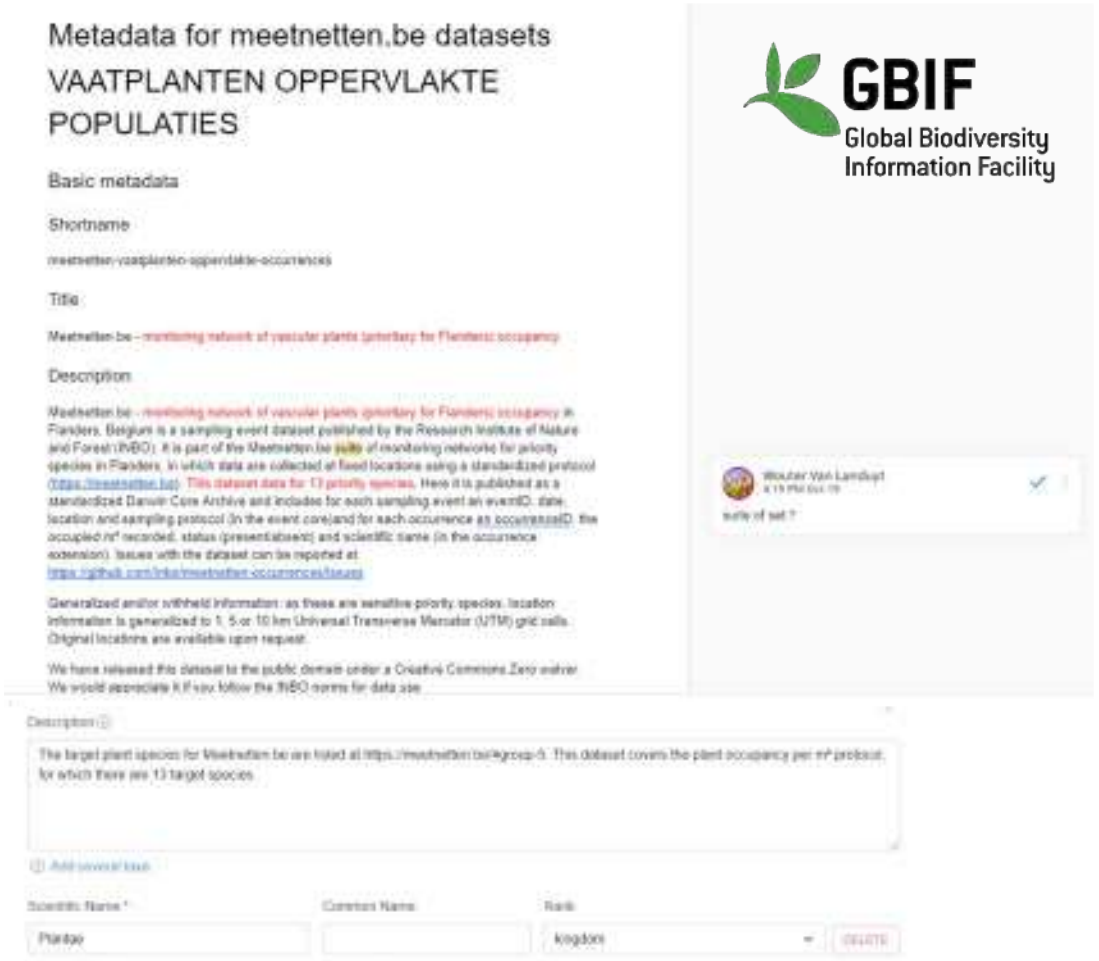
From database
to dissemination for our partner ANB

- mapservice



From database to data publication on GBIF

- Metadata completion
  - Data publication team
  - Check by Monitoring Network Coordinator
  - Copy Metadata in IPT - tool
  - Eml.xml (https://eml.ecoinformatics.org/)



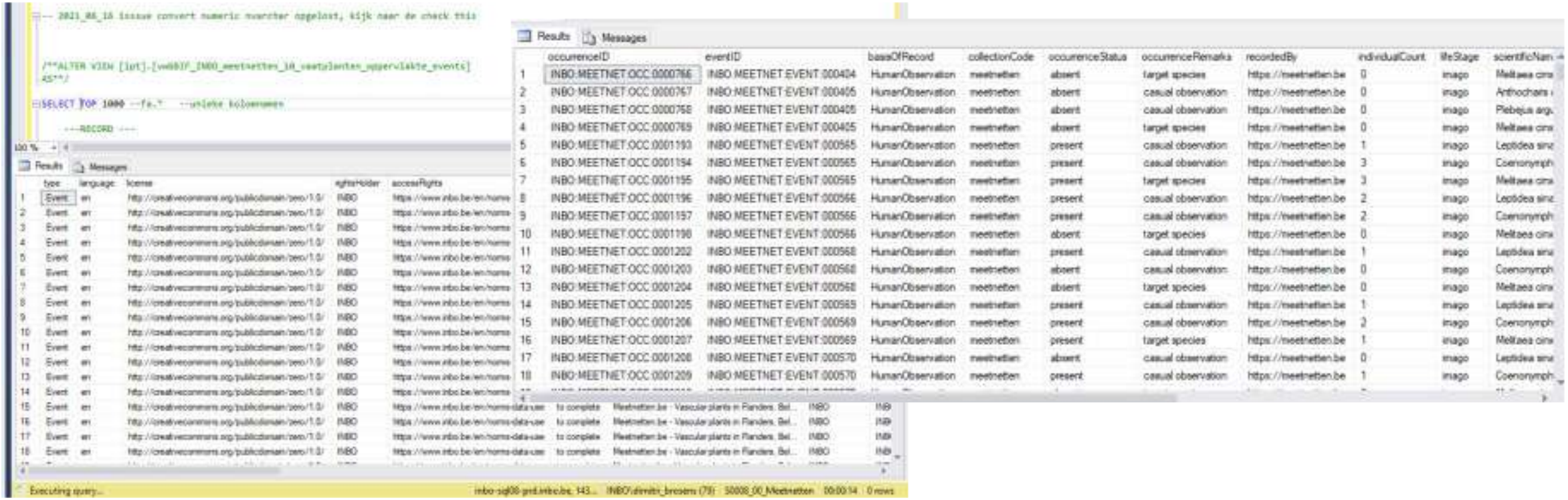
Meetnetten.be
NATUURPUNT | INSTITUUT NATUUR- EN BOSONDERZOEK | AGENTSCHAP NATUUR EN BOS

From database to data publication on GBIF

- Mapping of data to Darwin Core
  - On Staging version
  - Create DwC - GBIF - IPT SQL view

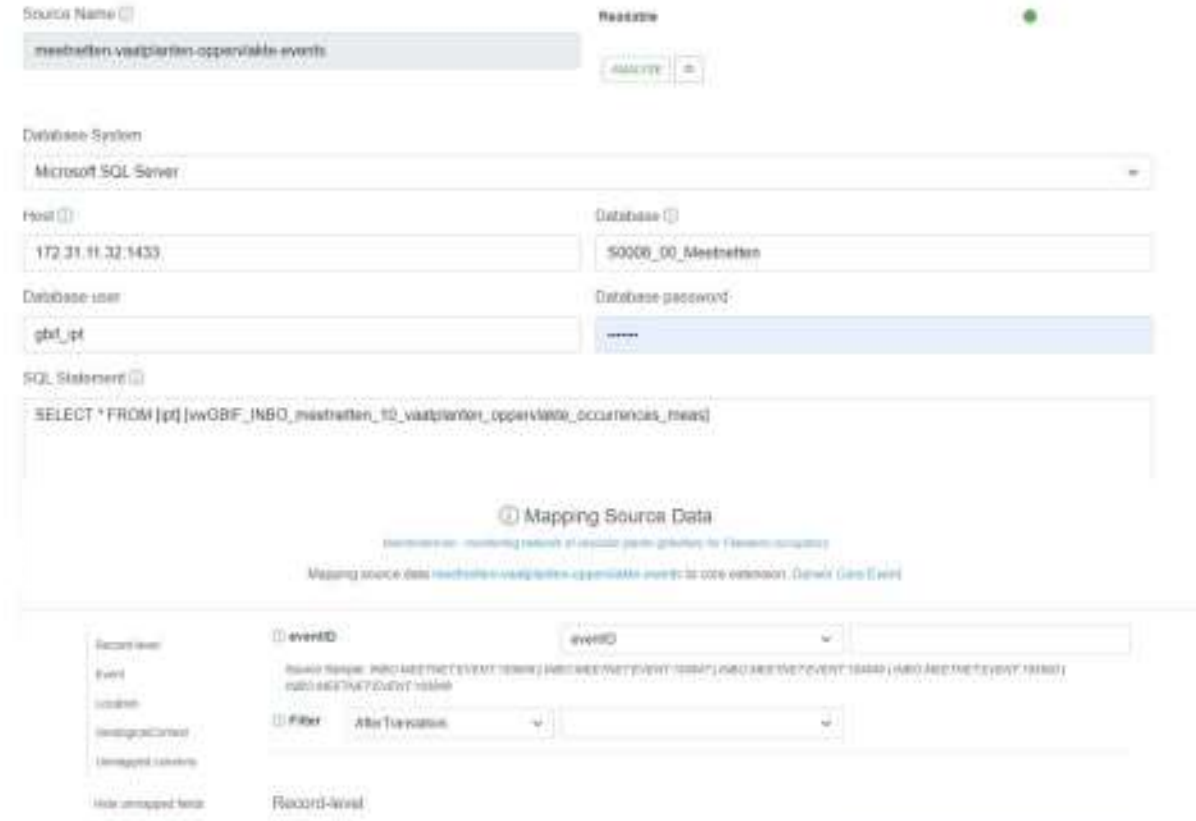
From database to data publication on GBIF

- Mapping of data to Darwin Core
  - Event Core
  - Occurrence Extension
  - Measurement or Fact Extension



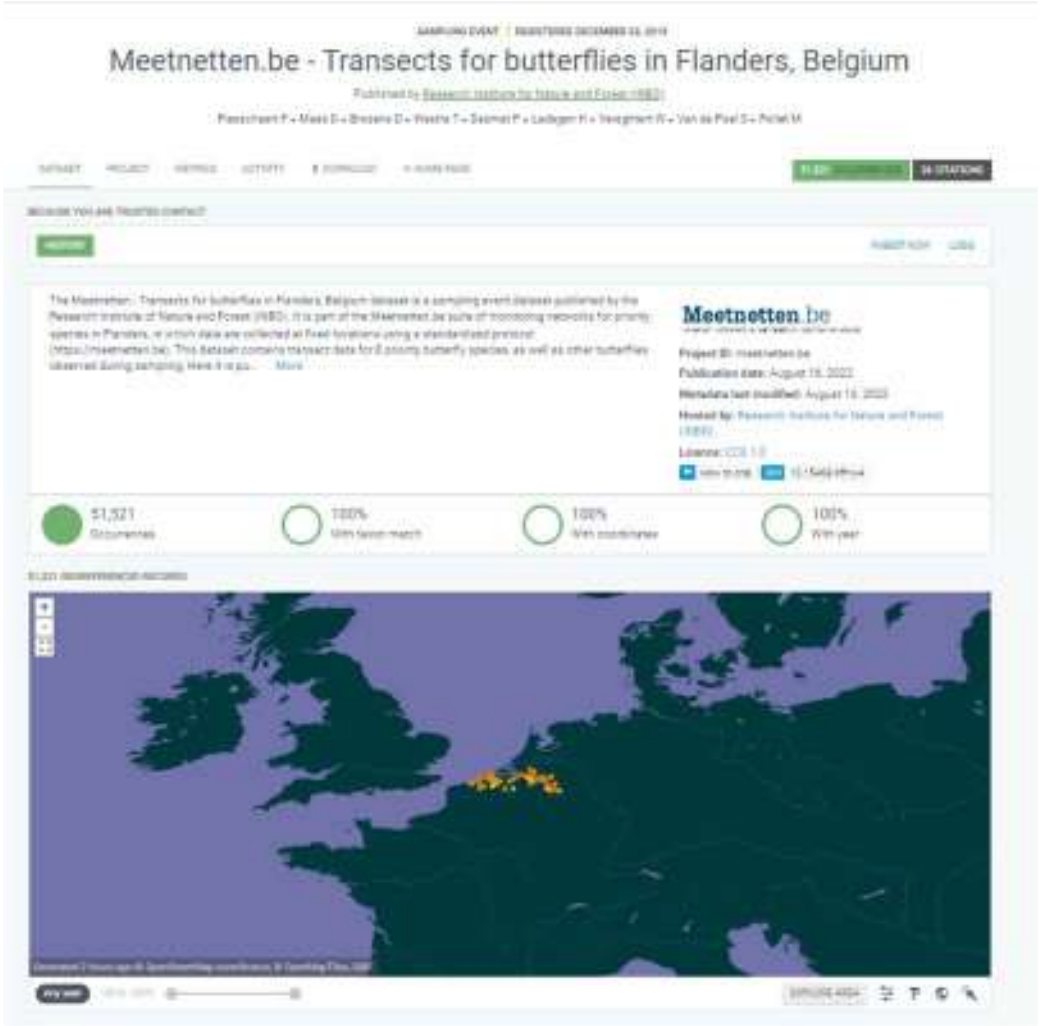
From database to data publication on GBIF

- Connect data in IPT
- DwC mapping (automatic)



From database to data publication on GBIF

- Publish on GBIF
  - Make data public
  - Register to GBIF
  - Publish resource



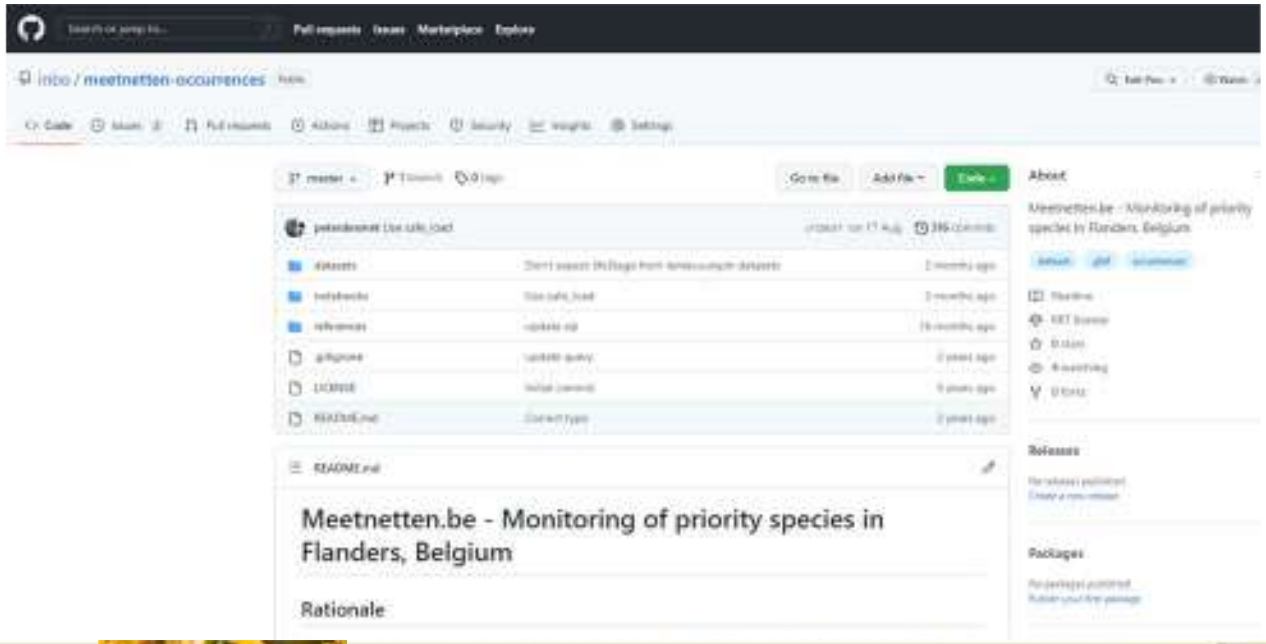
From database to data publication on GBIF

- Find all monitoring data on GBIF
  - https://www.gbif.org/occurrence/search?collection_code=meetnetten
  - Almost 100000 records available

Vlaanderen is wetenschap

From database to data publication on GBIF

- All documentation (data publication) on https://github.com/inbo/meetnetten-occurrences



Meetnetten.be
NATUURPUNT | INSTITUUT NATUUR- EN BOSONDERZOEK | AGENTSCHAP NATUUR EN BOS

From database to data publication on GBIF

Method steps

1. Researchers from INBO and Natuurpunt Studie define and document the appropriate sampling protocol for the target species.
2. Fieldwork is planned and coordinated by Natuurpunt Studie, using https://meetnetten.be.
3. Data are collected in the field by specialized volunteers, using the predefined sampling protocol.
4. Volunteers enter the collected data in https://meetnetten.be.
5. A custom SQL view is created in the meetnetten.be database to map the original data to Darwin Core as an event core and occurrence extension.
6. The Darwin Core views are connected to the INBO IPT and documented with metadata.
7. The dataset is published and registered with GBIF.

Meetnetten =

Toon Westra, Geert De Knijf, Hannes Ledegen, Luc De Bruynn, Dirk Maes, Thierry Onkelinx, Frederic Piesschaert, Wouter Vanreusel, Bernard Van Elegem, Marc Pollet, Paul Quataert, Dimitri Brosens…..

Vlaanderen verbeelding werkt

INSTITUUT NATUUR- EN BOSONDERZOEK

AGENTSCHAP NATUUR & BOS

natuurpunt

NATURA 2000

Life

ZOSTERA

Meetnetten.be

NATUURPUNT | INSTITUUT NATUUR- EN BOSONDERZOEK | AGENTSCHAP NATUUR EN BOS

Vlaanderen is wetenschap

biodiversa+
European Biodiversity Partnership

German NFDI4Biodiversity – data architecture and data workflows

By Barbara Ebert, managing coordinator of NFDI4Biodiversity

NFDI 4
BIODIVERSITY

www.nfdi4biodiversity.org
@NFDI4Biodiv
#NFDI4Biodiv

Mobilising biodiversity data in Germany

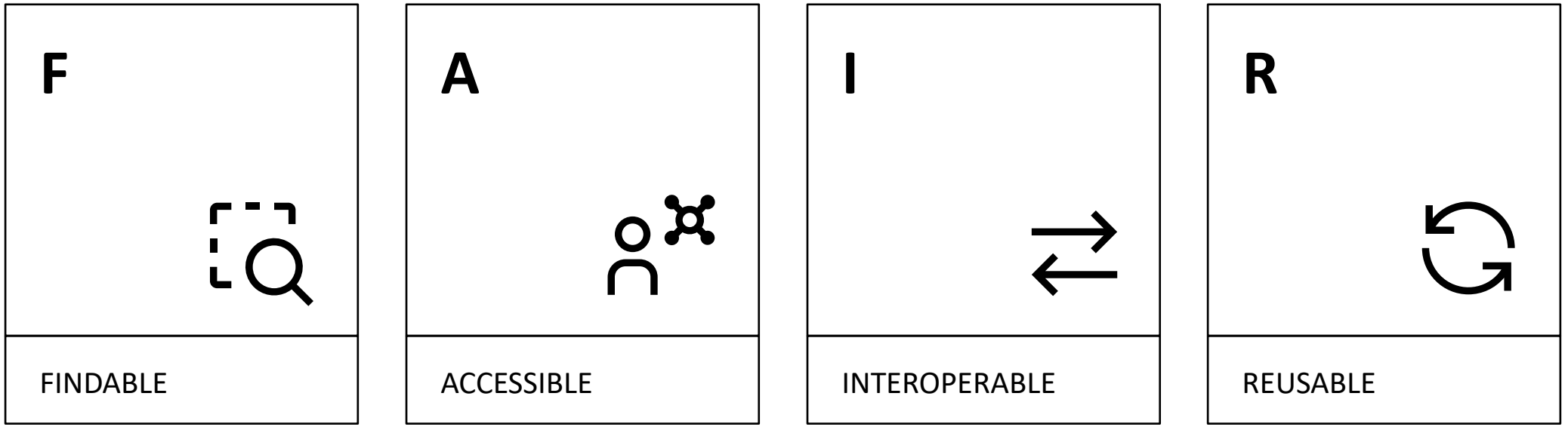
Workflows and service architecture

04.11.2022

NFDI4Biodiversity in a nutshell

- A practitioner network („community of practice“) with 50 partners and a 5-year work programme until 2025
- Funded by the Joint Science Minister Conference of the Federal Government and the Federal States in Germany as part of the National Research Data Infrastructure
- Transdisciplinary – science and public service
- Dedicated to mobilise data and roll out new services for the wider biodiversity community

FAIR Data principles



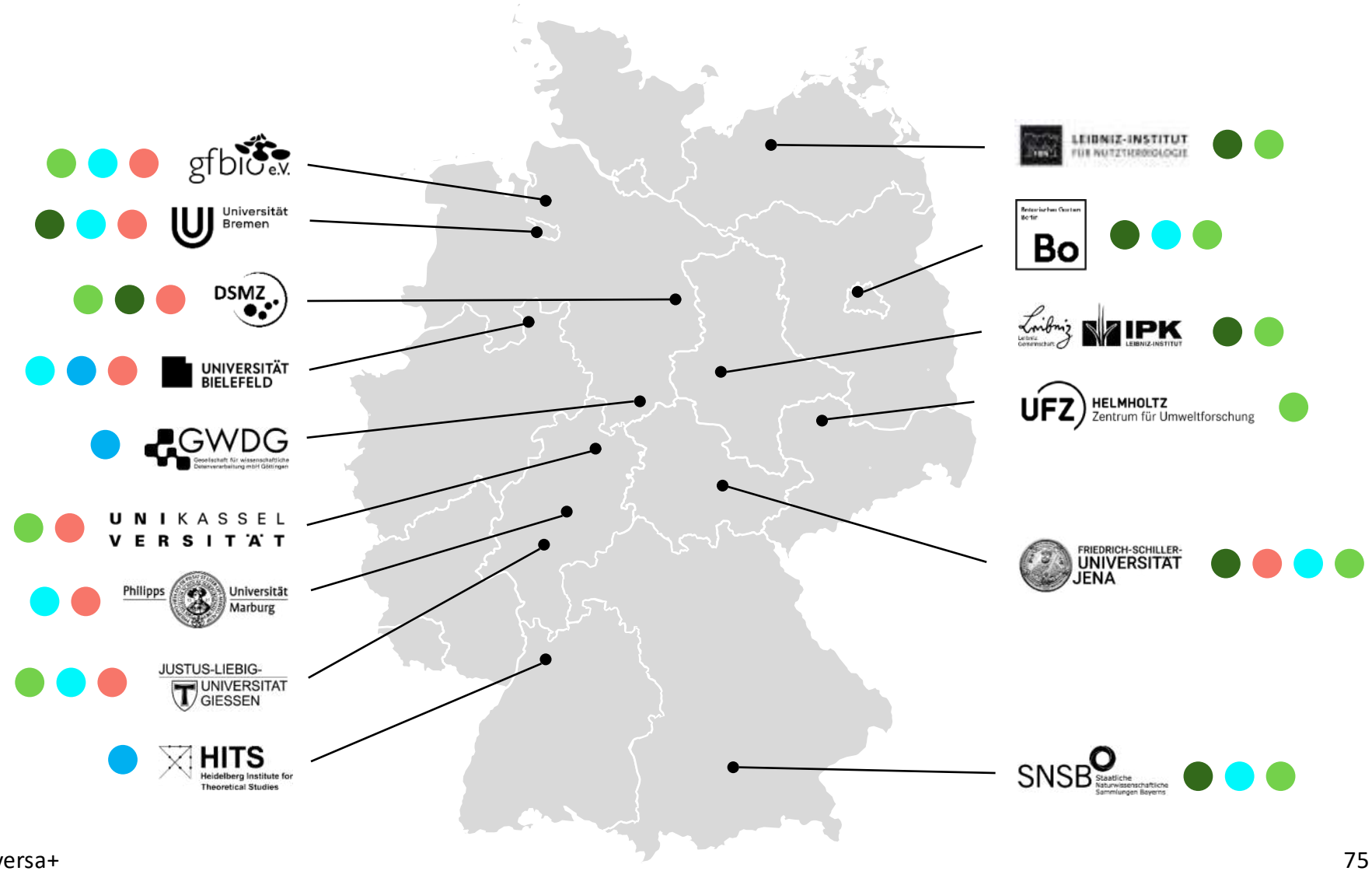
The Project

Funded through the National Research Data Infrastructure NFDI		
Up to **13.6** **Mio EUR** Funding for five years DFG Project No. 442032008	**Consortium of 50 partners from Academia, Data Centers, State Agencies and Societies/Citizen Science**	**> 120** **Active experts & staff**

The Consortium: 15 Co-Applicants

NFDI 4 BIODIVERSITY

- Data Center
- Infrastructure Provider
- Computer Science
- Biology / Env. Sciences
- Teaching / Training



The Consortium >30 Participants

NFDI 4 BIODIVERSITY

Academia

AWI
BIOfid
de.NBI
DNAquaNet
HIFMB
IfL
IGB
InfAI
IÖR
MPI-BGC
SUB
Uni Göttingen
Uni Leipzig

Collection Datacenters

BGBM-DI
MfN
SGN
SMNS
ZFMKZMT

State Agencies

BSH
HLNUG
Kühn Institut
LAU
LfULG
NP Bayerischer Wald
NP Hunsrück
Staatl. Archive Bayern
Thünen Institut

Societies & Citizen science

AraGes Spinnen
BUND
DDA-Vögel
GdO-Libellen
Gfl-Fische
GfÖ
Naturgucker
NetPhyD-Pflanzen

Use Case Projects

01 eLTER	02 Plants - IPK	03 AMMOD Hub	04 DE Barcode of Life	05 Marine Life	06 Land use/cover
07 DNAquaNet	08 Naturgucker	09 Insekten Sachsen	10 FAIRagro Consortium	11 MultiBase Users	12 NP Bayer. Wald
13 MultiBase Users (2)	14 AlgaTerra	15 GdO - Libellen	16 Ara-Ges - Spinnen	17 NetPhyD	18 GfI - Fische
19 Viz Tools	20 NP Hunsrück	21 MultiBase, LA (3)	22 iDiv PlantHub	23 Living Atlas	24 Staatl. Archive Bayerns
25 Bird monitoring data	26 Waterlink-R	...			

Community engagement

Use Cases at the Core

What is a use case?		
A distinct project within NFDI4Biodiversity	↘	Project team, goals, work plan etc. (up to 3 years)
Representing real data needs in biodiversity, e.g.	↘	Making specific collections of data „fit for sharing“ > data mobilization
	↘	Implementing standards (e.g. for structured data, reference lists etc.)
	↘	Implementing state of the art software for data handling and representation
	↘	Opening tried-and-tested software or architectural frameworks for wider use in the community

Blueprints (1)

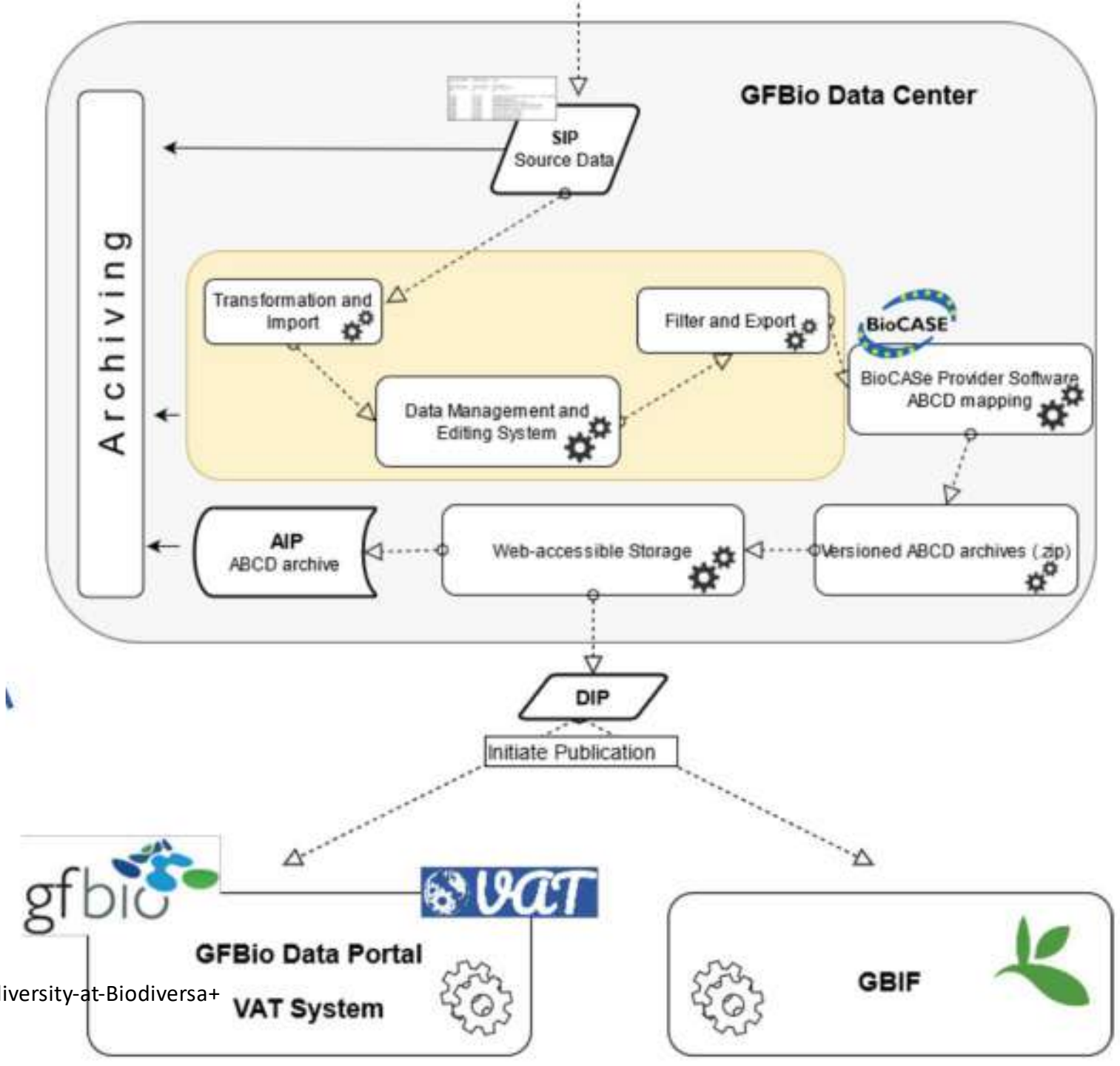
Harmonised data pipelines for occurrence data from collection data centers

GFBio achievements

- Portfolio for research data management from planning to archiving and publication
- Templates optimised for domain-specific data types
- Common (metadata) language
- Unified data submission system
- Data delivery from Data Centers to joint portal
- Integrated search and visualisation of data
- Outreach, support and training

Plan Prepare a custom Data Management Plan (DMP).	**Submit** Submit your data to GFBio.
Publish Make your data citable.	**Train** Train your data management skills.
Search Search the GFBio data pool.	**Visualize & Analyze** Dynamically integrate, analyze and visualize GFBio datasets.
Archive Deposit data and specimens in dedicated long-term archives.	**Annotate & Connect** Use the GFBio Terminology Service to describe your data and share terminologies with other researchers.

GFBio Portal


GFBio Data Center
SIP
Source Data
Archiving
Transformation and Import
Filter and Export
BioCASE
BioCASe Provider Software
ABCD mapping
Data Management and Editing System
AIP
ABCD archive
Web-accessible Storage
Versioned ABCD archives (.zip)
DIP
Initiate Publication
gfbio
GFBio Data Portal
VAT
VAT System
GBIF

Available as precedent for other interested natural history collections

Extensive documentation in the GFBio Public Wiki, for example:

- Publication of occurrence data via Biocase Data Pipelines – for each Data Center https://gfbio.biowikifarm.net/wiki/Publication_of_Type_1_Data_via_BioCASe_Data_Pipelines_at_GFBio_Data_Centers
- Eligibility Criteria for GFBio portal data providers https://gfbio.biowikifarm.net/wiki/Eligibility_criteria_for_GFBio_portal_data_providers

GFBio Data Centers

- Act as data archives for the community
  - Joint submission system
- Deliver data to a joint portal
  www.gfbio.org
  - Searchable
  - Combinable in a visualisation, transformation and analysis tool

Data Centers specialized on Nucleotide, Plant and Environmental Data

e!DAL Plant Genomics & Phenomics Research Data Repository

ENA European Nucleotide Archive

PANGAEA Data Publisher for Earth & Environmental Science

Data Centers at Natural Science Collections

Bo Botanischer Garten Berlin

DSMZ Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

Museum für Naturkunde Berlin

SENCKENBERG world of biodiversity

Naturkunde Museum Stuttgart

SNSB Staatliche Naturwissenschaftliche Sammlungen Bayerns

Forschungsmuseum Koenig

Blueprints (2)

In progress: Harmonised data pipelines for metabarcoding studies

Metabarcoding studies and data

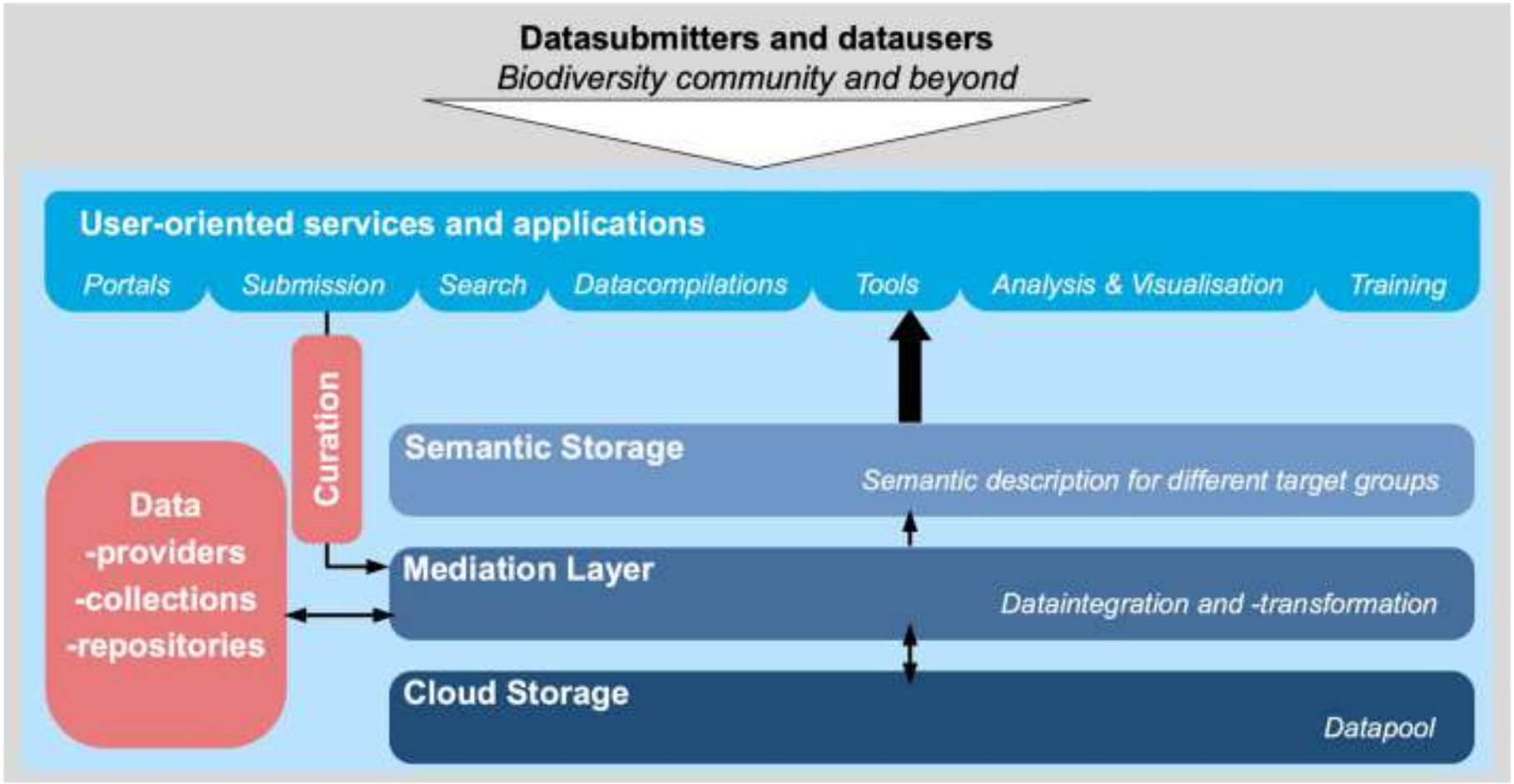
NFDI 4 BIODIVERSITY

01 eLTER	02 Plants - IPK	03 AMMOD Hub	04 DE Barcode of Life	05 Marine Life	06 Land use/cover
07 DNAquaNet	08 Naturgucker	09 Insekten Sachsen	10 FAIRagro Consortium	11 MultiBase Users	12 NP Bayer. Wald
13 MultiBase Users (2)	14 AlgaTerra	15 GdO - Libellen	16 Ara-Ges - Spinnen	17 NetPhyD	18 GfI - Fische
19 Viz Tools	20 NP Hunsrück	21 MultiBase, LA (3)	22 iDiv PlantHub	23 Living Atlas	24 Staatl. Archive Bayerns
25 Bird monitoring data	26 Waterlink-R	...			

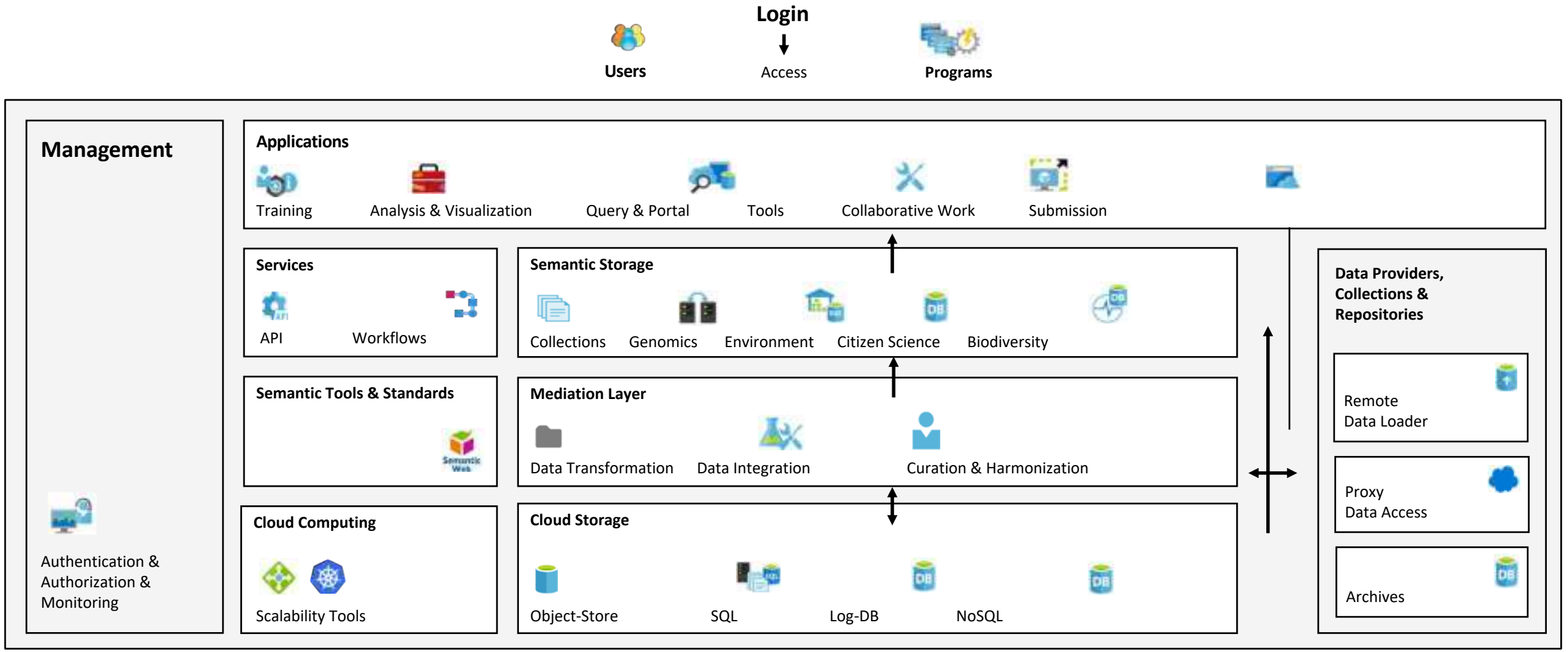
Vision

Cloud-based research data commons

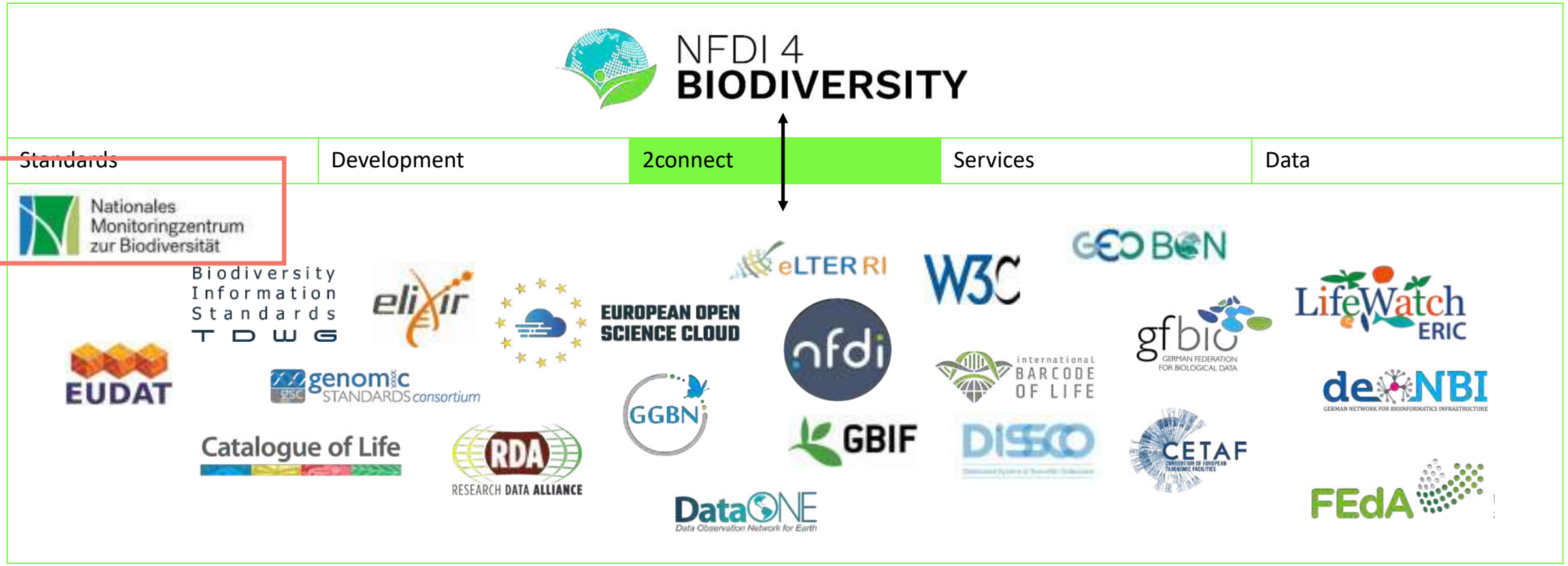
Moving services to the academic cloud



Research Data Commons – technical view



National and International Networking



Thank you!

NFDI 4 BIODIVERSITY

Funded by the DFG, project number 442032008

Q&A

biodiversa+
European Biodiversity Partnership

Introduction to the sub-group session

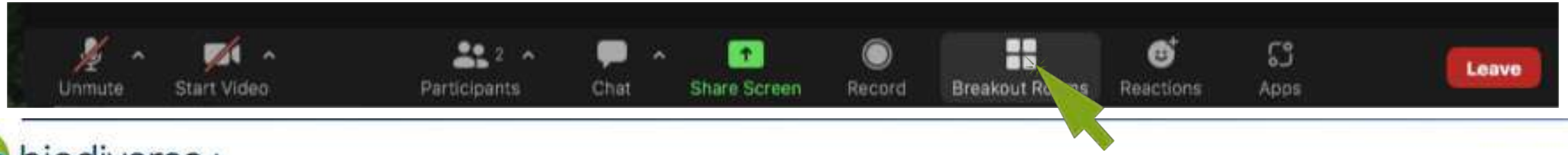
By Cécile Mandon, Biodiversa+ officer, FRB

We invite you to choose the group that most interest you

Objective: agree on action/ activities to be implemented by Biodiversa+ to address the focus of your sub-group

How will I join my sub-group?

You will have to select the sub-group that you would like to join



Let’s take a break!

10 min

biodiversa+
European Biodiversity Partnership

Recommendations from the sub-groups

By the rapporteurs

REC

The plenary sessions of this meeting will be recorded and shared on the Biodiversa+ website and Youtube channel

biodiversa+
European Biodiversity Partnership

Conclusions of the workshop

By Alberto Basset, MUR and Hilde Eggermont, BelSPO, Biodiversa+ Chair and Coordinator

biodiversa+
European Biodiversity Partnership

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Thank you!

www.biodiversa.org

contact@biodiversa.org

BiodiversaPlus