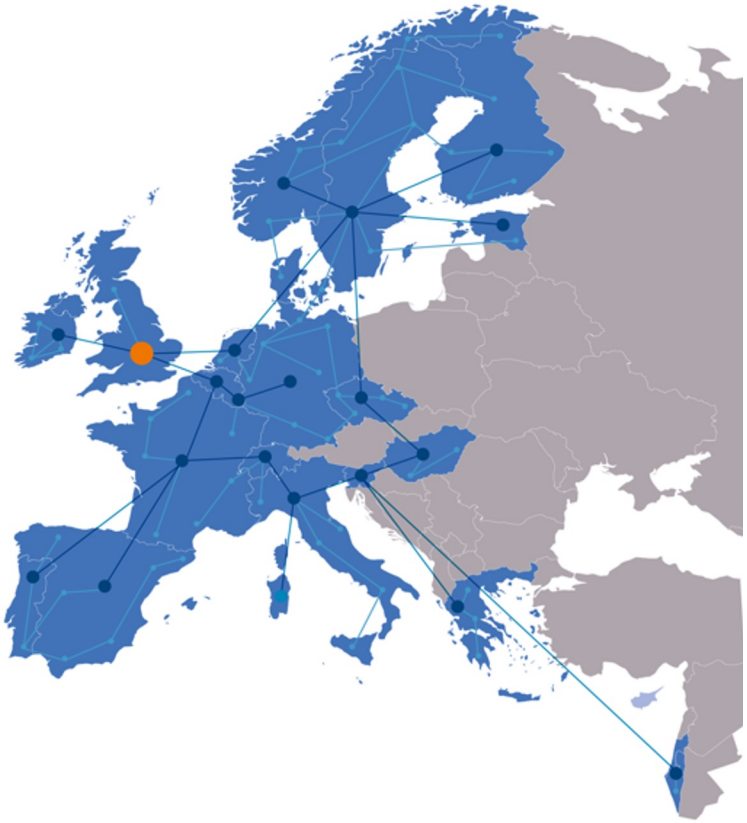
ELIXIR3: Empowering Biodiversity Research through Collaboration and Innovation

Michael Dondrup, Terje Klemetsen
University of Bergen | UiT the Arctic University

Introducing ELIXIR and ELIXIR Norway

Mission: to operate a sustainable European infrastructure for biological information, supporting life science research and its translation to society, the bio-industries, environment and medicine.



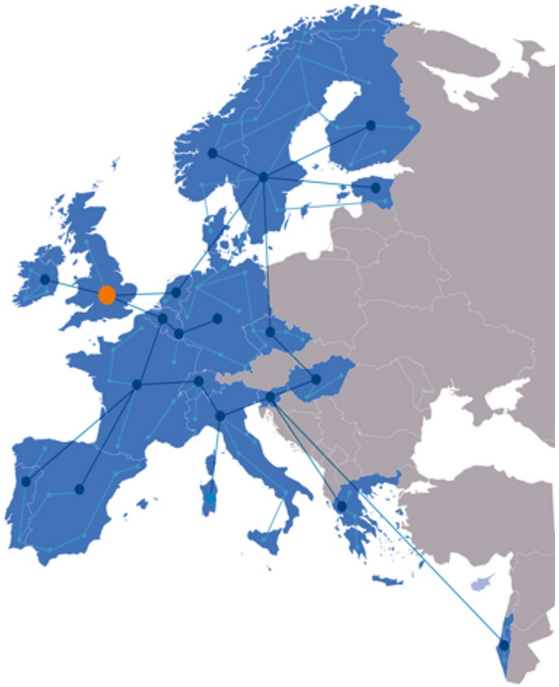
SINCE 2013

ELIXIR: ‘Data for Life’

An intergovernmental organisation that brings together life science resources such as databases, software tools, training materials, standards and compute resources from across Europe.

Goal: coordinate life science resources from across Europe so they form a single infrastructure. This makes it easier for scientists to:

- Find and share data
- Exchange expertise
- Agree on best practices in scientific research



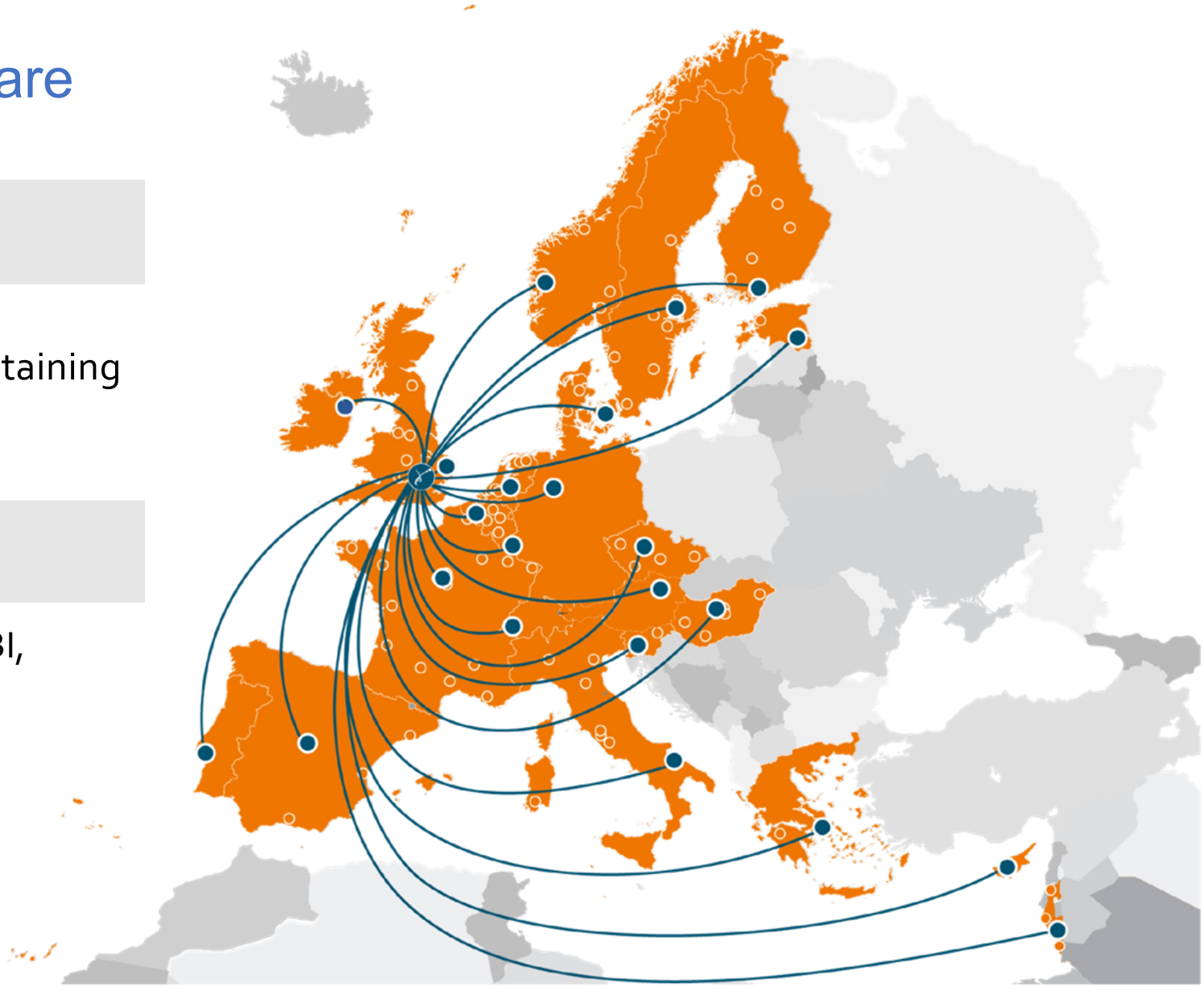
ELIXIR – who we are

24 Nodes

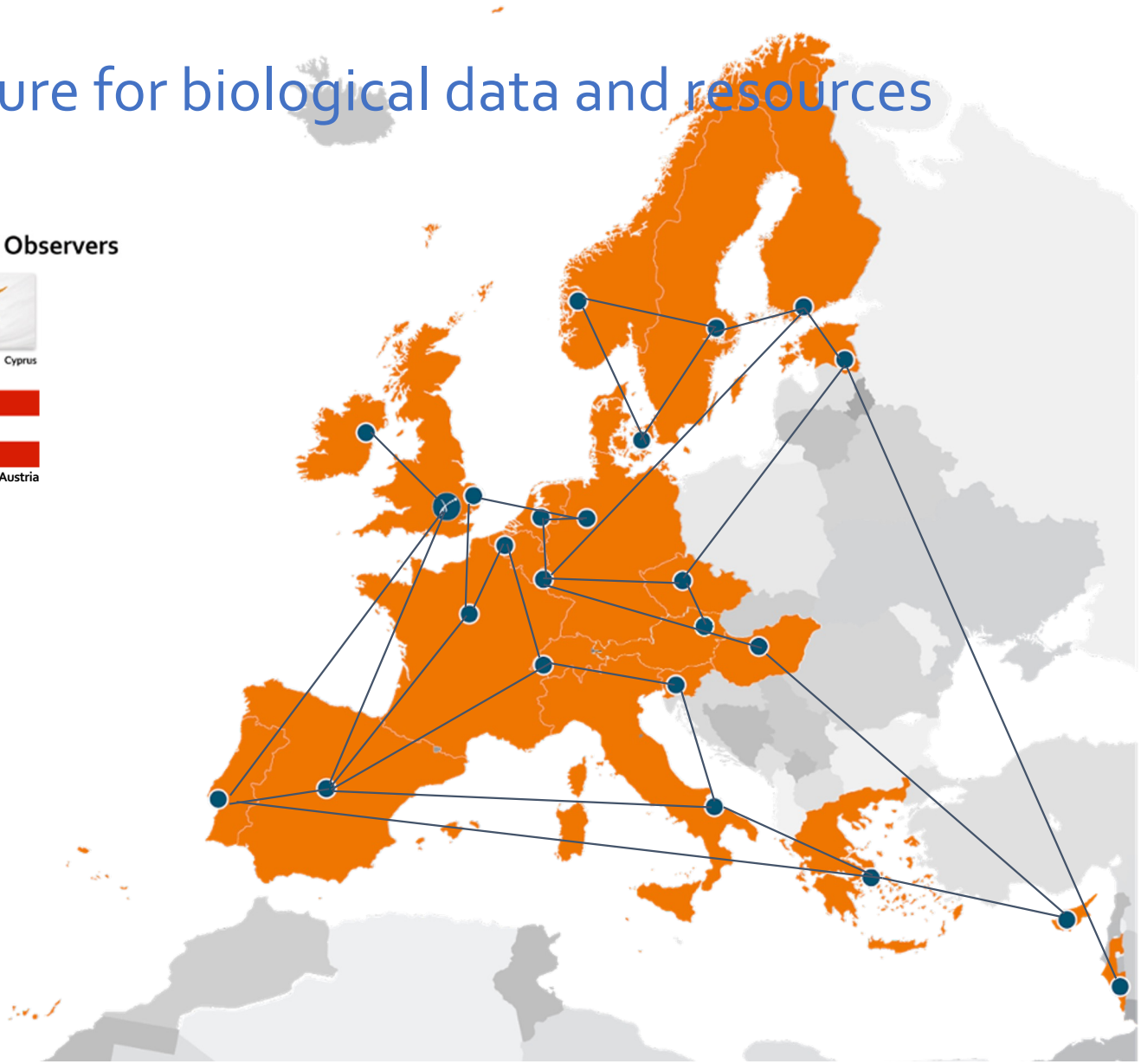
ELIXIR Nodes form national networks of excellence, containing a total of **245** institutes/universities

1 Hub

The Hub, located at EMBL-EBI, Cambridge, UK, provides the secretariat and coordinating activities for ELIXIR



A sustainable infrastructure for biological data and resources



ELIXIR Norway

SINCE 2014

The Norwegian Node of ELIXIR since 2014

Currently funded by:

The Research Council of Norway (2026)

The involved institutions: UiB, UiO, NMBU, NTNU and UiT

Services to the users:

- e-infrastructure, training and support, for analysis and management of life science data

Deliverables towards ELIXIR Europe:

- 12 Norwegian data resources and analysis tools

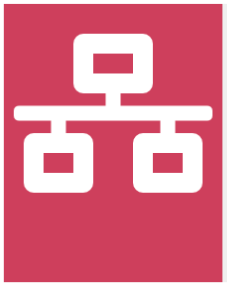
https://elixir.no/

Helpdesk

Short- and long-term support with bioinformatics analyses, programming and data management tasks. Contact the helpdesk.

Services

Analysis and management of life science data within marine, health, genomics, proteomics, structural analysis and more.



NeLS e-infrastructure

NeLS, the Norwegian e-Infrastructure for Life Sciences, for analysis, sharing and storage of high throughput genomics data.

Training

Local and national training in bioinformatics analysis and use of our tools.

Services to Norwegian users

E-Infrastructure for Data analysis, storage and sharing

1. Non-sensitive data:
   - **NeLS - the Norwegian e-infrastructure for Life Sciences**

2. Sensitive data:
   - FEGA Norway

Federated European Genome-phenome Archive Norway

Support - through our nationally coordinated Helpdesk:

- **Analysis**
- **Scripting**
- **Workflow development**
- **FAIR Data management**
- **Data Brokering** (FHI (SARS-CoV2++), **EBP-Nor**)

`support@elixir.no`

RDMkit

Training:

- Using our tools and workflows
- Data management

DSW

DS Wizard ELIXIR-Norway

Data management support

Life cycle of Research Data

ELIXIR research data management resources

Guidance for data stewards, project managers and researchers

Overview of good data management practices

RDMkit

The Research Data Management Kit (RDMkit) guides you through the whole data management life cycle and includes advice specific to your domain, your role and your country.

Step-by-step instructions

FAIR cookbook

The FAIR Cookbook contains step-by-step recipes to accomplish specific data management tasks and to make your data FAIR (Findable, Accessible, Interoperable, Reusable).

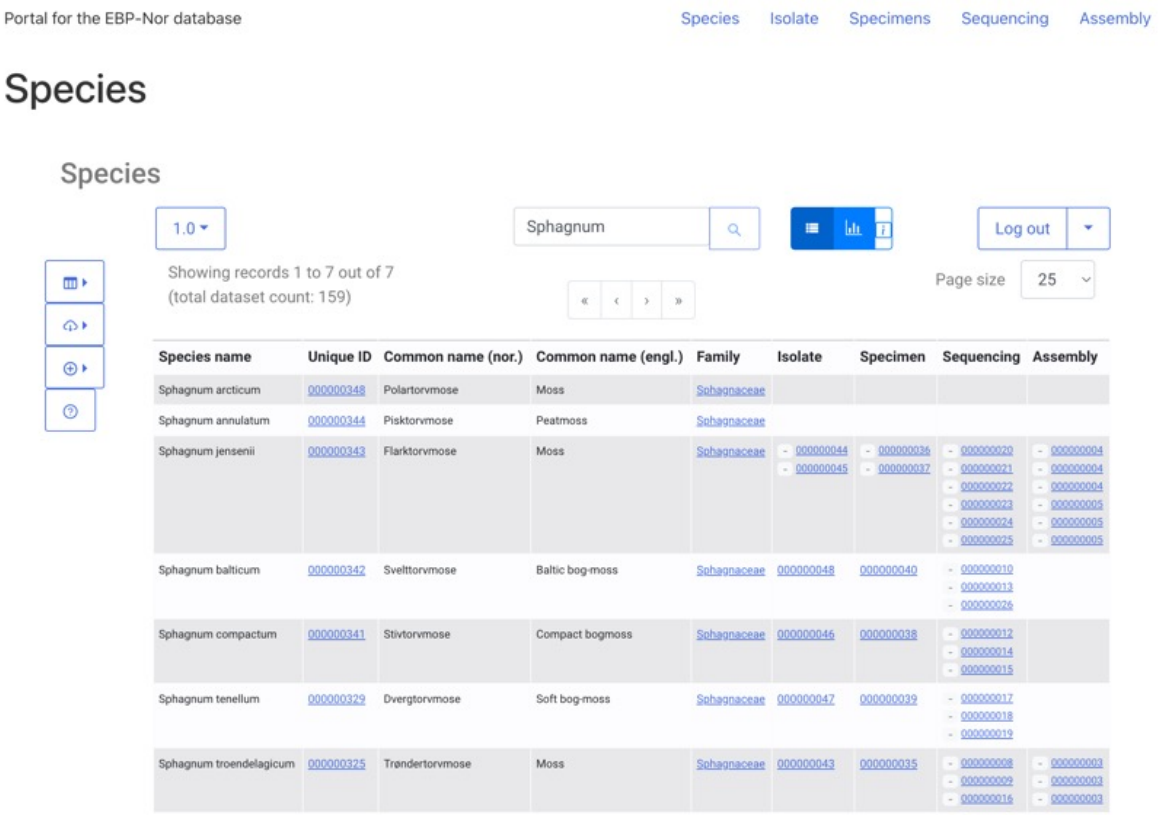
Data management plan wizard

DSW

The Data Stewardship Wizard (DSW) is an online tool that guides researchers and data stewards through their data management planning.

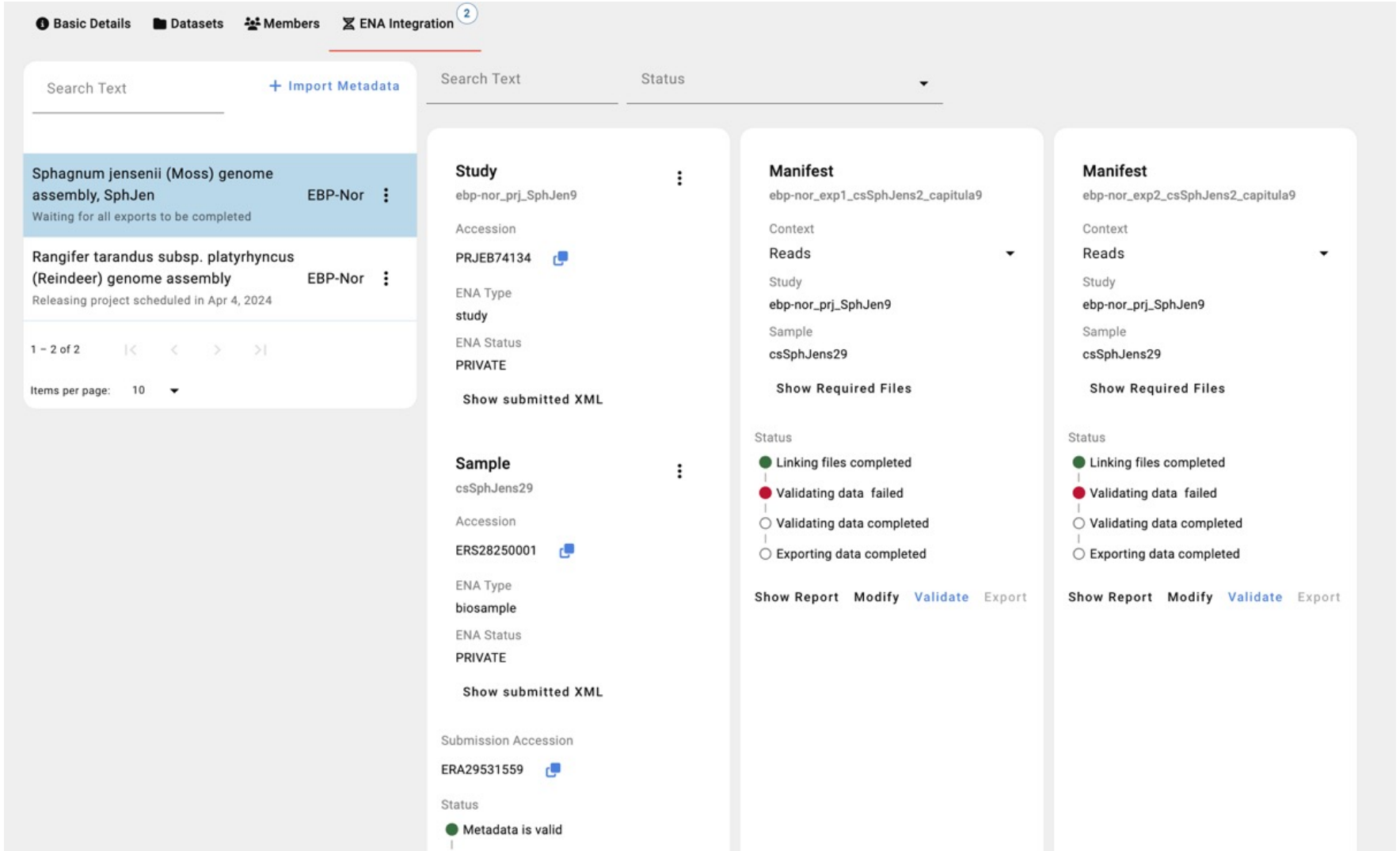
EBP-Nor Databases for Meta-data (aka. Brokering Part I)

- (Semi)-automatic import from sampling sheets
- + manual curation
- Federated with NeLS/StoreBioinfo
- Used to provide annotation for upload to European Nucleotide Archive (ENA)



databases.ebpnor.org

Raw data from NeLS/StoreBioinfo to ENA (Brokering Part II)

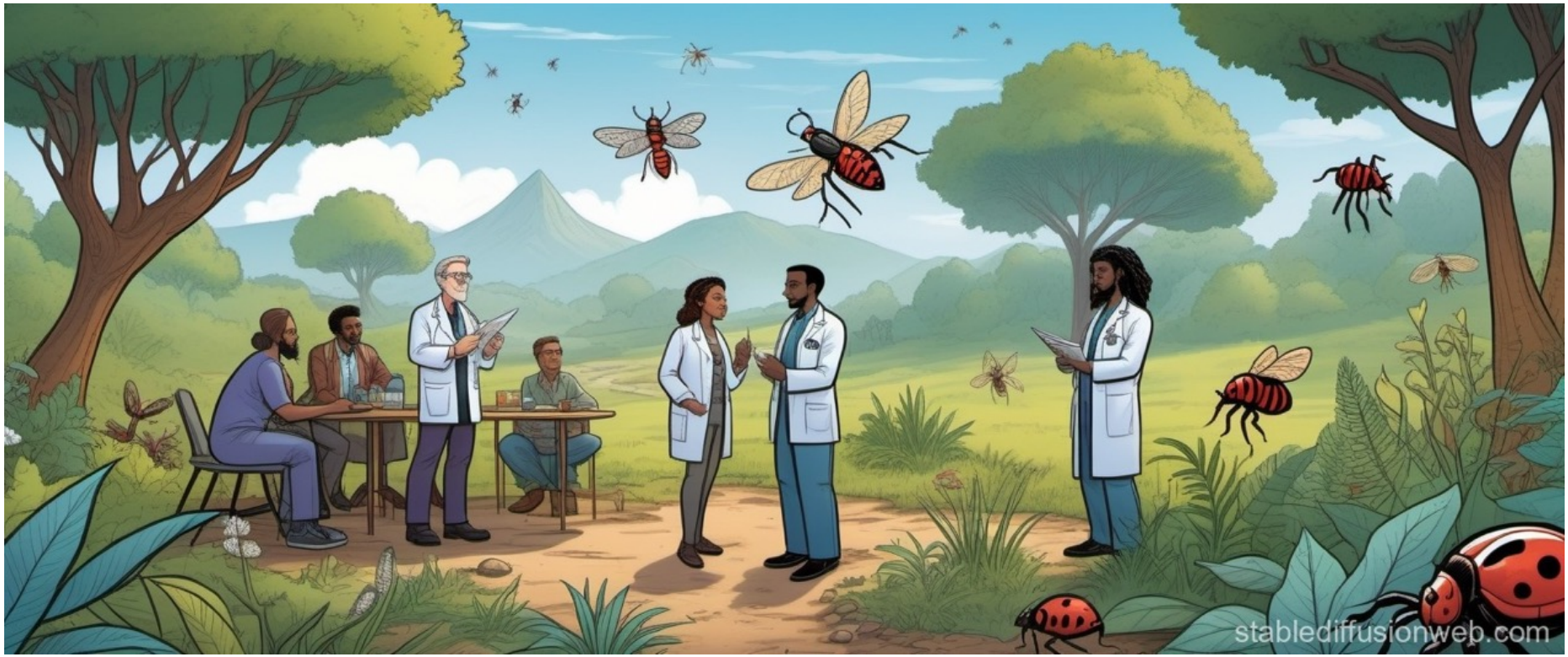


The Biodiversity Working Group

- Working group representing the key players in biodiversity
- Established through Elixir WP6
- Monthly meetings
- Topics: Data collection, management best practices
- Establishing collaboration and information sharing
- Getting a better understanding of the needs of the community

NIVA Norsk institutt for vannforskning
NorBOL Norwegian Barcode of Life
NINA
Living Norway Ecological Data Network
EARTH BIOGENOME PROJECT
DiSSCo Distributed System of Scientific Collections
NHMO DNA Bank
GBIF

2024 Survey of the Norwegian Biodiversity Community



Nettskjema

Driving Progress in Biodiversity Research: Your Thoughts on Enhancing Molecular Biodiversity Data Management

Updated: 11 March 2024 at 15:54

Are you working with biodiversity data? Help the community improve by identifying the current needs by taking our 10-minute questionnaire. Your contribution is valuable, and your privacy is protected - no personal information is collected, only role and affiliation which can be anonymized to your preference.

- Designed in WP6
- Distributed via the Biodiversity Working Group in March ’24
- 58 questions covering the whole RD life-cycle
- Respondents from 16 different Norwegian institutions
- There are some key results

Association with initiatives is very diverse

- Majority of respondents is associated with large national or transnational initiative
- Also, a large group `other`

(2) Are you affiliated with one or more large national, transnational initiatives?

Number of submissions: **58**

Submissions	Count	% of submissions	
EBP	16	27.6%	27.6%
ERGA	6	10.3%	10.3%
BIOSCAN / IBOL / NorBOL	12	20.7%	20.7%
ELIXIR	15	25.9%	25.9%
DiSSCo	3	5.2%	5.2%
GBIF	8	13.8%	13.8%
eLTER	1	1.7%	1.7%
Other	26	44.8%	44.8%

Most are both data provider and consumer of public data

(10) Do you use pre-existing, published datasets?

Number of submissions: **58**

Submissions	Count	% of submissions	
Yes	52	89.7%	89.7%
No	4	6.9%	6.9%
Don't know	2	3.4%	3.4%

(20) Are the datasets in your projects intended for re-use by others?

Number of submissions: **58**

Submissions	Count	% of submissions	
Yes, always	23	39.7%	39.7%
Most of the time	29	50%	50%
Seldom	5	8.6%	8.6%
No, never	0	0%	0%
Don't know	1	1.7%	1.7%

Over 90% publish molecular data.

- Strong support for using controlled vocabularies
- Support of meta-data standards and standard operating procedures
- Majority does validation checks on meta-data
- Good awareness of the FAIR principles

But there are some challenges with existing data tools…

- Wide-spread use of spreadsheet programs
- Lack of meta-data in published data
- Taxonomic classification is sometimes challenging
- Need for more taxonomic expertise

(13) What do you find most challenging when using published sequence databases?

Number of submissions: **57**

Submissions	Count	% of submissions	
Finding the right one	18	31.6%	31.6%
Understanding the format	8	14%	14%
Understanding the annotation	11	19.3%	19.3%
Lack of metadata	31	54.4%	54.4%
I don't find it challenging	13	22.8%	22.8%
Not relevant	6	10.5%	10.5%

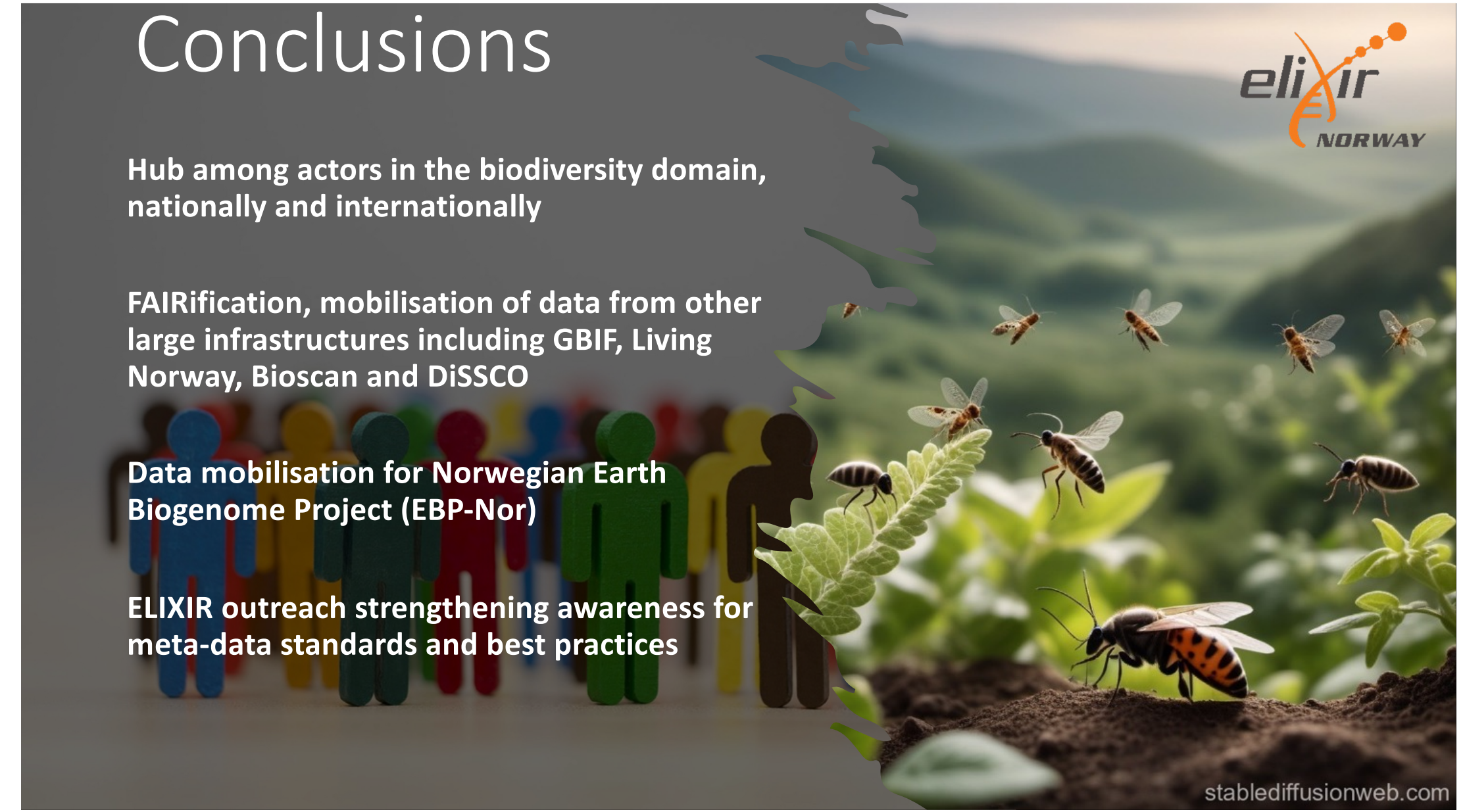
Conclusions

Hub among actors in the biodiversity domain, nationally and internationally

FAIRification, mobilisation of data from other large infrastructures including GBIF, Living Norway, Bioscan and DiSSCO

Data mobilisation for Norwegian Earth Biogenome Project (EBP-Nor)

ELIXIR outreach strengthening awareness for meta-data standards and best practices



elixir.no

contact@elixir.no

@elixirnorway

- Joint effort across 5 institutions
- More than 50 people involved:
  - Software developers
  - Bioinformaticians
  - Coordinators
  - Researchers

UNIVERSITY OF BERGEN

UiT The Arctic University of Norway

UiO : University of Oslo

NTNU
Norwegian University of Science and Technology

Norwegian University of Life Sciences