Workshop in Data Management for Life Science projects

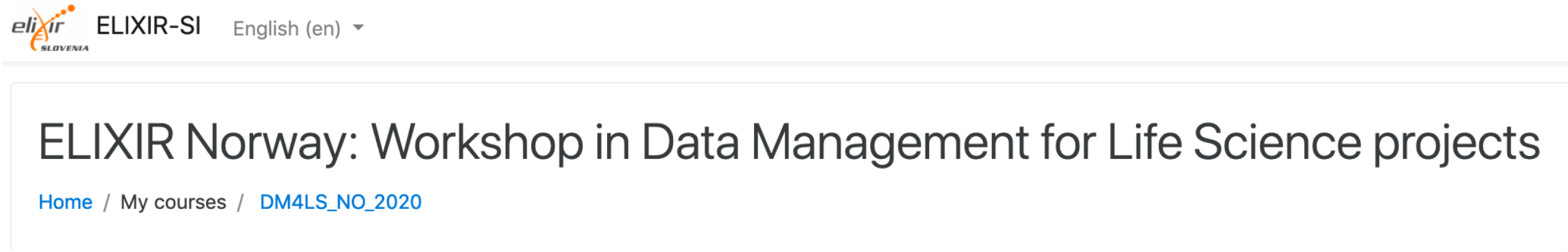
ELIXIR - The national bioinformatics infrastructure

Erik Hjerde
Korbinian Bösl

www.elixir-europe.org

Time schedule and course material

https://elixir.mf.uni-lj.si/course/view.php?id=51



Introduction to ELIXIR

A distributed European infrastructure for life-science information

Erik Hjerde

www.elixir-europe.org

www.elixir-europe.org

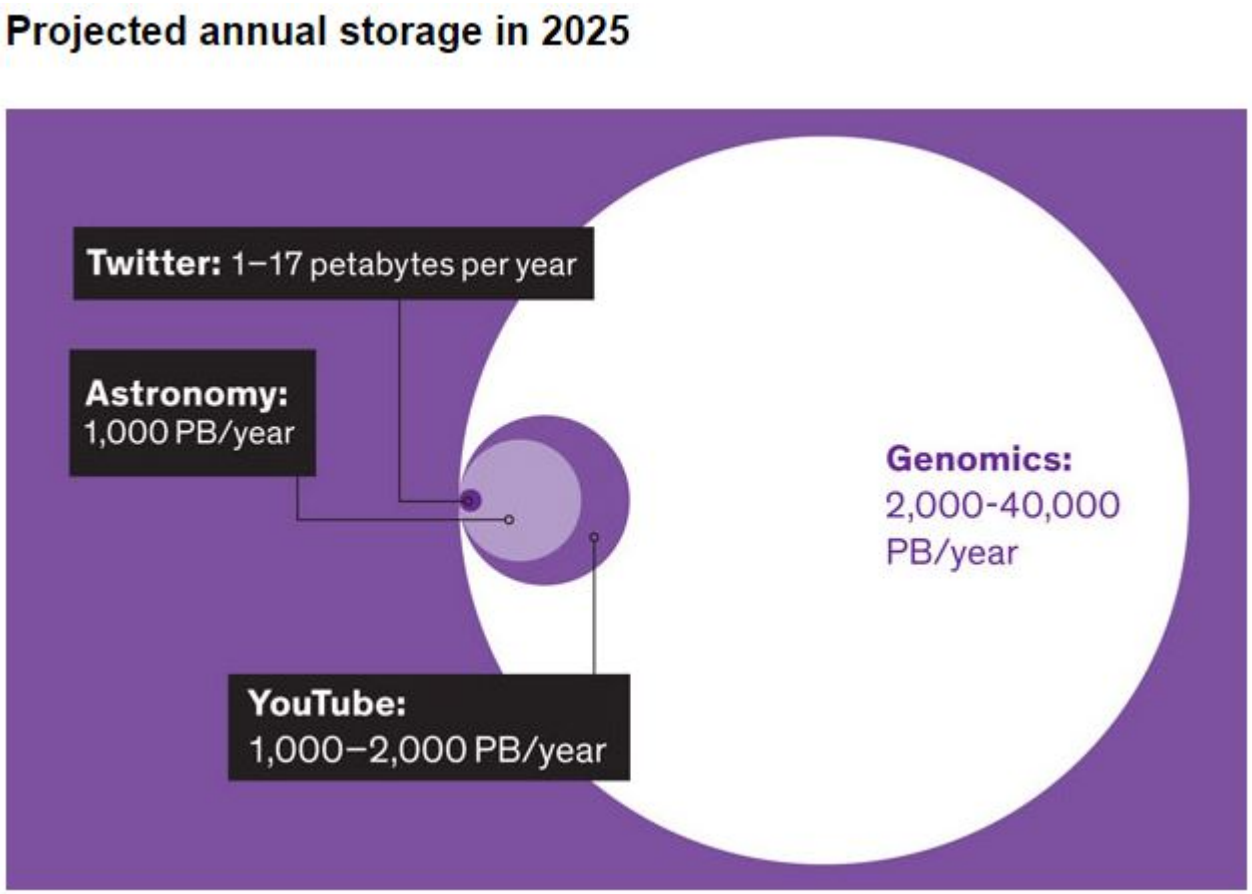
ELIXIR provides similar services for life-science data

Tools, databases, storage, sharing, compute, common standards, educational resources

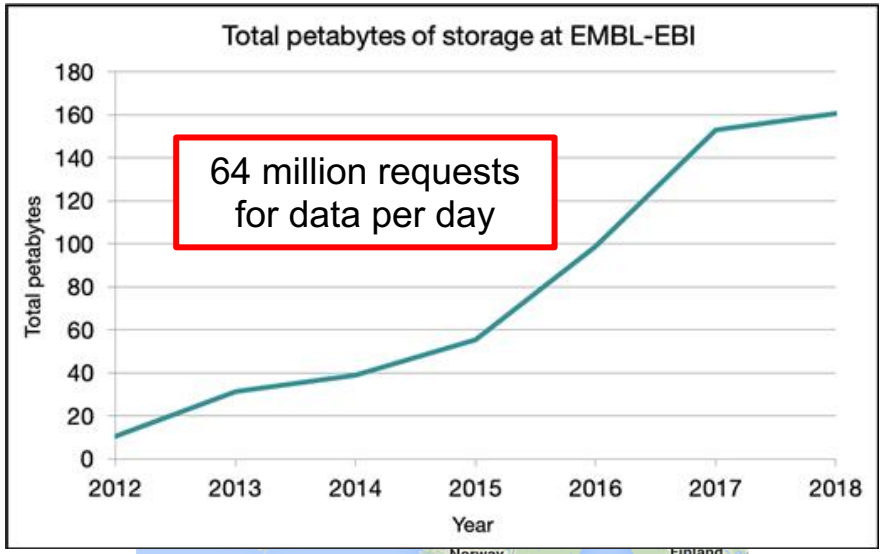
www.elixir-europe.org

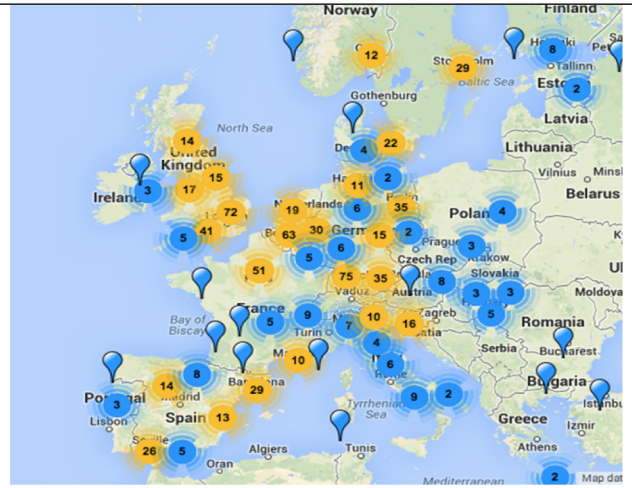
Production of life science data is enormous

Essential to standardise how we store data and metadata



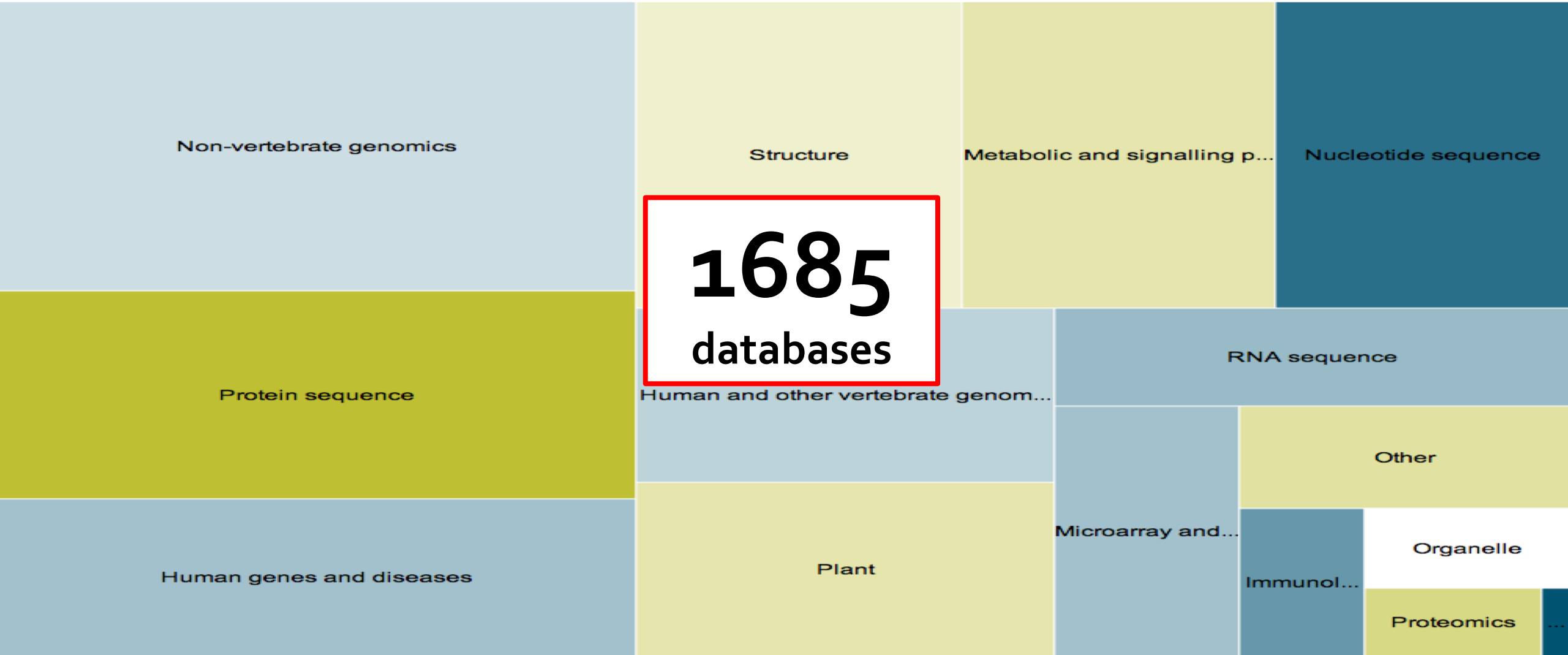
Source: "Big Data: Astronomical or Genomical?" PLoS Biology, 7 July 2015.





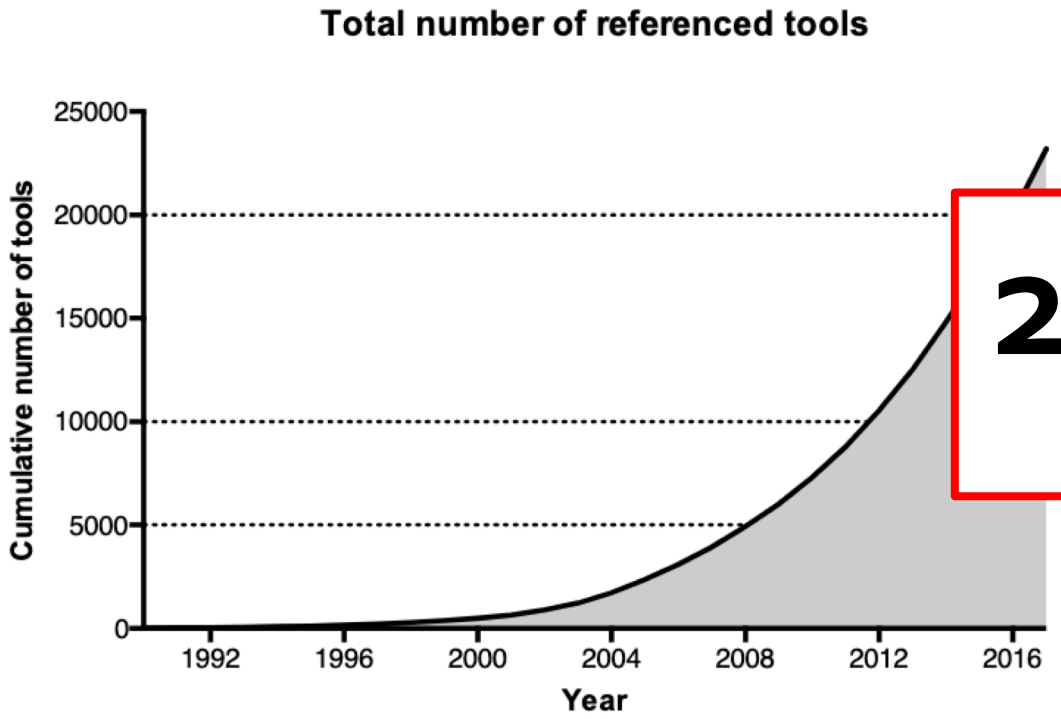
Life-science databases

Dispersed throughout the world, and serve many different research communities

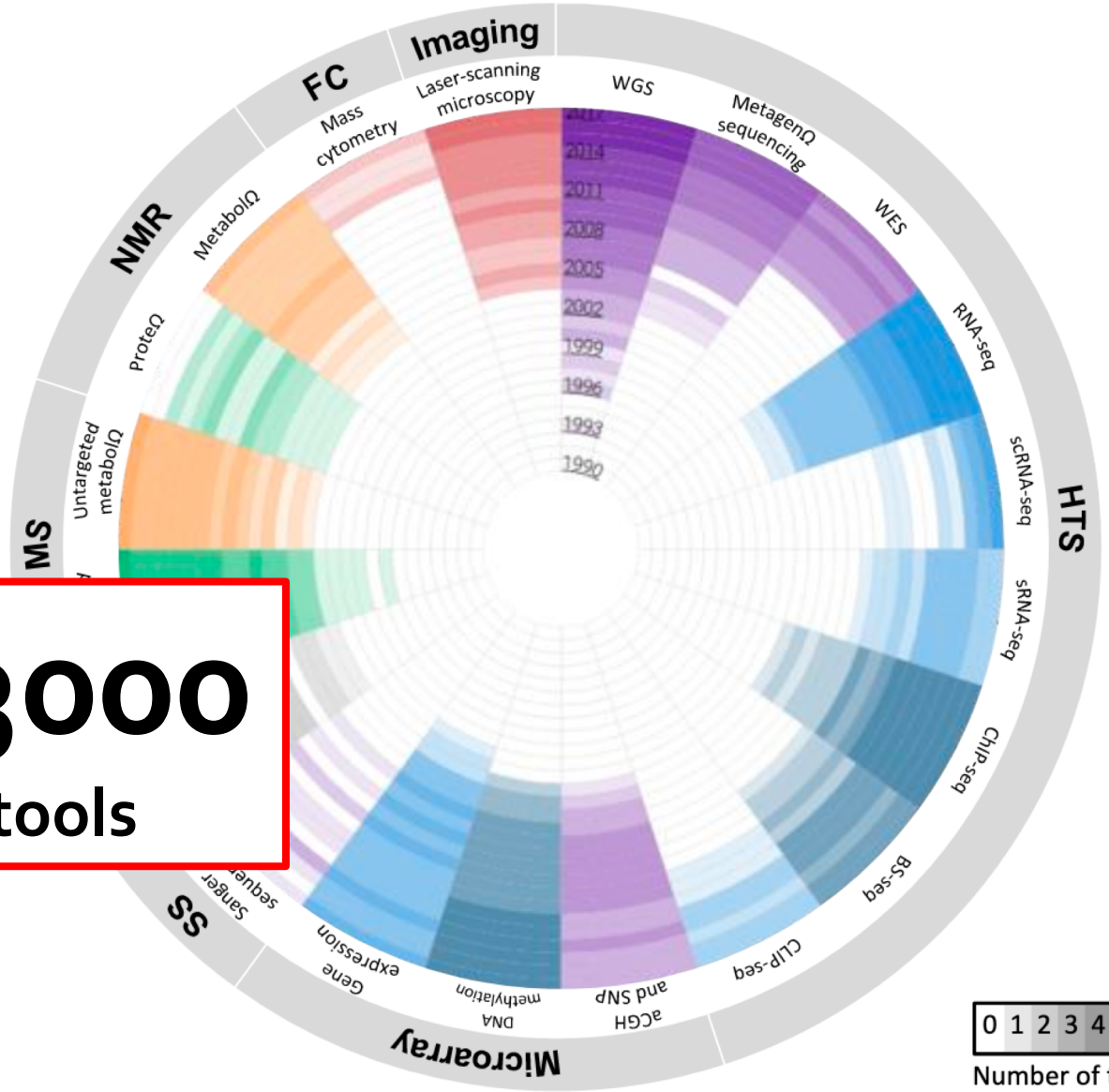


Bioinformatic tools

Crucial component of the infrastructure for life sciences



23000 tools



omicX

Clément Levin et al. 2018

elixir

Do we need ELIXIR????????

Data management planning is required for all grant applications

Important to capture metadata in all parts of the project

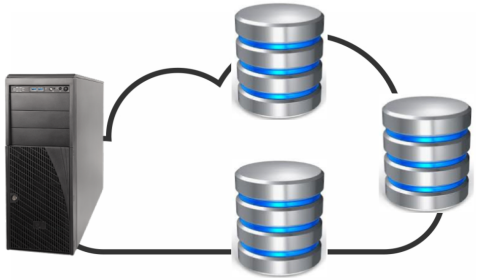
Important that the data is FAIR

Sample collection

Data production

Data storage/sharing

Data processing



Data achieving

Data management

Do we need ELIXIR????????

Data management planning is required for all grant applications

Important to capture metadata in all parts of the project

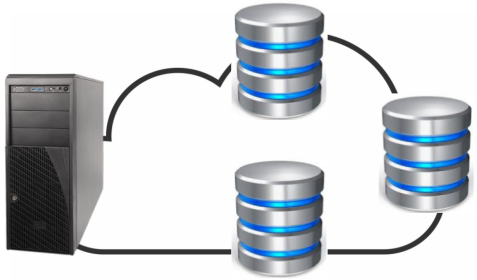
Important that the data is FAIR

Sample collection

Data production

Data storage/sharing

Data processing



Data achieving

Data management

ELIXIR - driver for the use of common standards and ontologies

Metadata standards – controlled vocabulary for

Structured comment name	Item	Description	Examples	Expected value	Value syntax	Preferred units / suffix
alt_elev	Geographic location (altitude/elevation)	Sample taken at given elevation above sea level, defined in meters(m) as a positive floating number with two decimals.	Ex 1: 3.06 Ex 2: 1.80-2.15	-	{float} or {range}	meters (m)
collection_date	Collection date	The time of sampling, either as an instance (single point in time) or interval. In case no exact time is available, the date/time can be right truncated.	Ex 1: 2008-01-23T19:23:10+00:00 Ex 2: 2011-11-10 Ex 3: 2001-12 Ex 7: 2015 Ex 4: 2003--2006 Ex 5: 2010-01--2011-03 Ex 6: 2011-05-28--2011-08-10	date and time, range	{timestamp}	-
depth	Depth	Please refer to the definitions of depth in the environmental packages. Water: Sample taken at given depth below sea level, defined in meters(m) as a positive floating number or as a range, both with two decimals.	Ex 1: 355.20 Ex 2: 2.00-5.00	-		meters (m)
env_biome	Environment (biome)	In environmental biome level are the major classes of ecologically similar communities of plants, animals, and other organisms. Biomes are defined based on factors such as plant structures, leaf types, plant spacing, and other factors like climate. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO (v1.53) terms listed under environmental biome can be found from the link:(http://www.environmentontology.org/Browse-EnvO)	Ex 1: coral reef Ex 2: tropical	EnvO	{free text}	-
env_biome_ENVO	Environment (biome_id)	Corresponding ENVO identifier related to the term name of Environment (biome).	Ex 1: ENVO:00000150 Ex 2: ENVO:01000204	EnvO	{accession}	-

Not collected	->	missing
250 M	->	250
Not applicable	->	NA
Superficial	->	missing
-1 m	->	1
-2 m	->	2
-2901.0	->	2901
0 m.	->	0
1912 ft	->	582.80
40 mm from surface	->	0.04
0.75 m above seafloor	->	missing
700meters	->	700
Intracellular	->	missing
Surface water of 0 meter	->	0
Zero	->	0
Below surface	->	Missing

ELIXIR - driver for the use of common standards and ontologies

Ontology Lookup Service (OLS) is a resource for biomedical ontologies

Structured comment name	Item	Description	Examples	Expected value	Value syntax	Preferred units / suffix
alt_elev	Geographic location (altitude/elevation)	Sample taken at given elevation above sea level, defined in meters(m) as a positive floating number with two decimals.	Ex 1: 3.06 Ex 2: 1.80-2.15	-	{float} or {range}	meters (m)
collection_date	Collection date	The time of sampling, either as an instance (single	Ex 1: 2008-01-	date and time, range	{timestamp}	-
depth	D					
env_biome	E					
env_biome_ENVO	E					

Ontology Lookup Service

Home | Ontologies | Documentation | About

OLS > eNanoMapper Ontology ENM > ENVO:00000447

marine biome

Search ENM

http://purl.obolibrary.org/obo/ENVO_00000447

An aquatic biome that comprises systems of open-ocean and unprotected coastal habitats, characterized by exposure to wave action, tidal fluctuation, and ocean currents as well as systems that largely resemble these. Water in the marine biome is generally within the salinity range of seawater: 30 to 38 ppt. [MA:ma ISBN-10:0618455043 ORCID:0000-0002-4366-3088 https://en.wikipedia.org/wiki/Ocean]

Tree view | Term history

- entity
  - material entity
    - biome
      - aquatic biome
        - marine biome

Graph view
Reset tree
Show all siblings

Term info

database cross reference
- SPIRE:Marine

has obo namespace
ENVO

has related synonym
marine realm

id
ENVO:00000447

The ENVO ontology describes the environment of the sampling

elixir

ELIXIR - driver for the use of common standards and ontologies

Ontology Lookup Service (OLS) is a resource for biomedical ontologies

Structured comment name	Item	Description	Examples	Expected value	Value syntax	Preferred units / suffix
alt_elev	Geographic location (altitude/elevation)	Sample taken at given elevation above sea level, defined in meters(m) as a positive floating number with two decimals.	Ex 1: 3.06 Ex 2: 1.80-2.15	-	{float} or {range}	meters (m)
collection_date	Collection date	The time of sampling, either as an instance (single	Ex 1: 2008-01-	date and time, range	{timestamp}	-
depth	D					
env_biome	E					
env_biome_ENVO	E					

Ontology Lookup Service

Home | Ontologies | Documentation | About

OLS > Gazetteer GAZ > GAZ:00002699

Kingdom of Norway

Search GAZ

http://purl.obolibrary.org/obo/GAZ_00002699

A country and constitutional monarchy in Northern Europe that occupies the western portion of the Scandinavian Peninsula. It is bordered by Sweden, Finland, and Russia. The Kingdom of Norway also includes the Arctic island territories of Svalbard and Jan Mayen. Norwegian sovereignty over Svalbard is based upon the Svalbard Treaty, but that treaty does not apply to Jan Mayen. Bouvet Island in the South Atlantic Ocean and Peter I Island and Queen Maud Land in Antarctica are external dependencies, but those three entities do not form part of the kingdom. [url:http://en.wikipedia.org/wiki/Norway]

Synonyms: Kongeriket Norge {language: Norwegian}, Norway, Kongeriket Noreg {language: Norwegian}

Tree view | Term history

- geographic location
  - Kingdom of Norway
    - Bouvet Islands
    - Dronning Maud Land
    - Jan Mayen
    - Metropolitan Norway
      - Lake Polden

Graph view | Reset tree | Show all siblings

Term info

database cross reference
- ISO3166-1:NO
- ISO3166-2:NO
- ISO3166-1:578
- ISO3166-1:NOR

ABBREVIATION
Norway

The GAZ ontology describes the geographical location of the sampling

We need common standards to describe data

Common data formats

Common ontology to describe eg. a gene

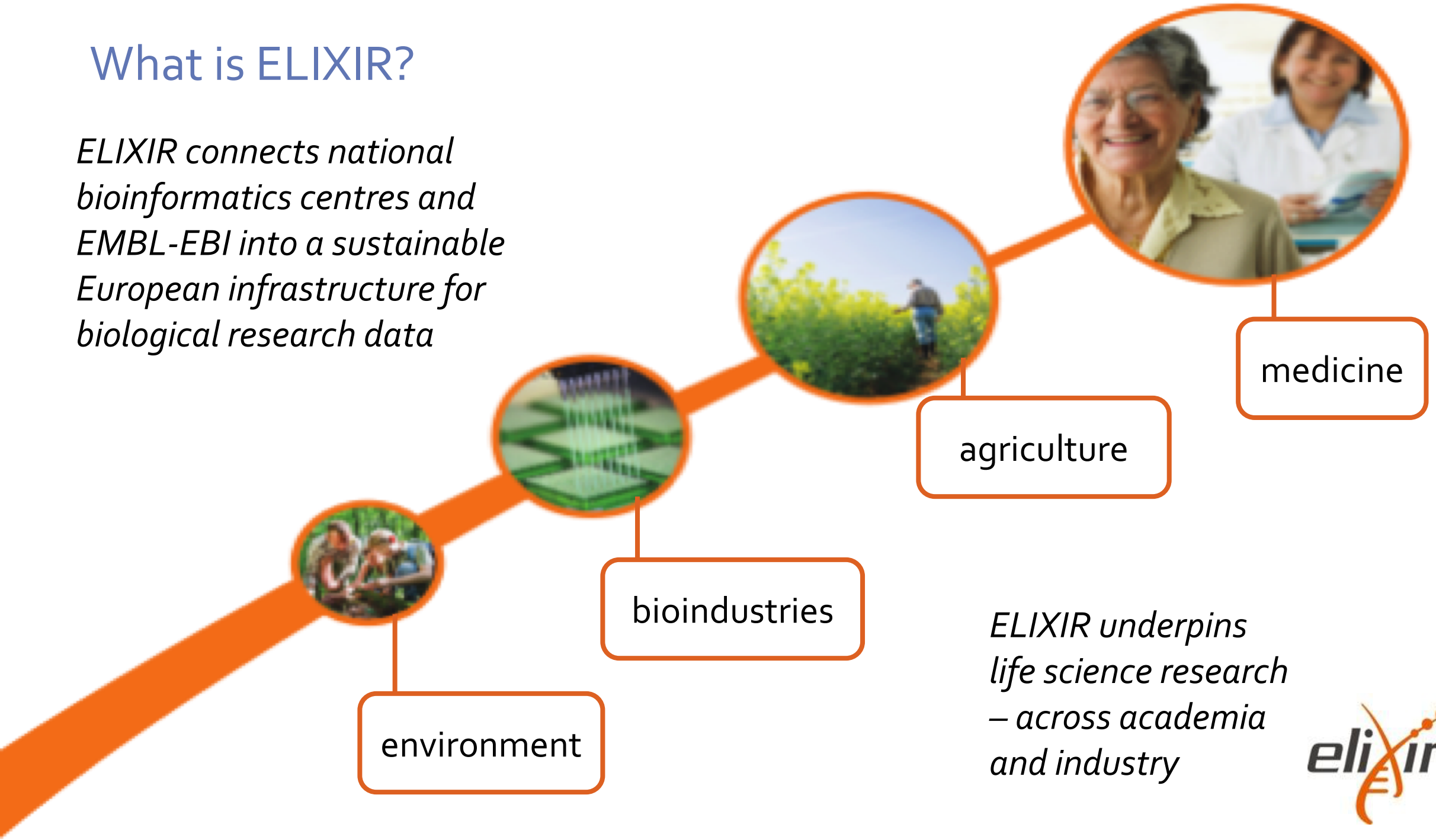
Common standards to describe metadata

Database registry

Tool registry

What is ELIXIR?

ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data



ELIXIR underpins life science research – across academia and industry

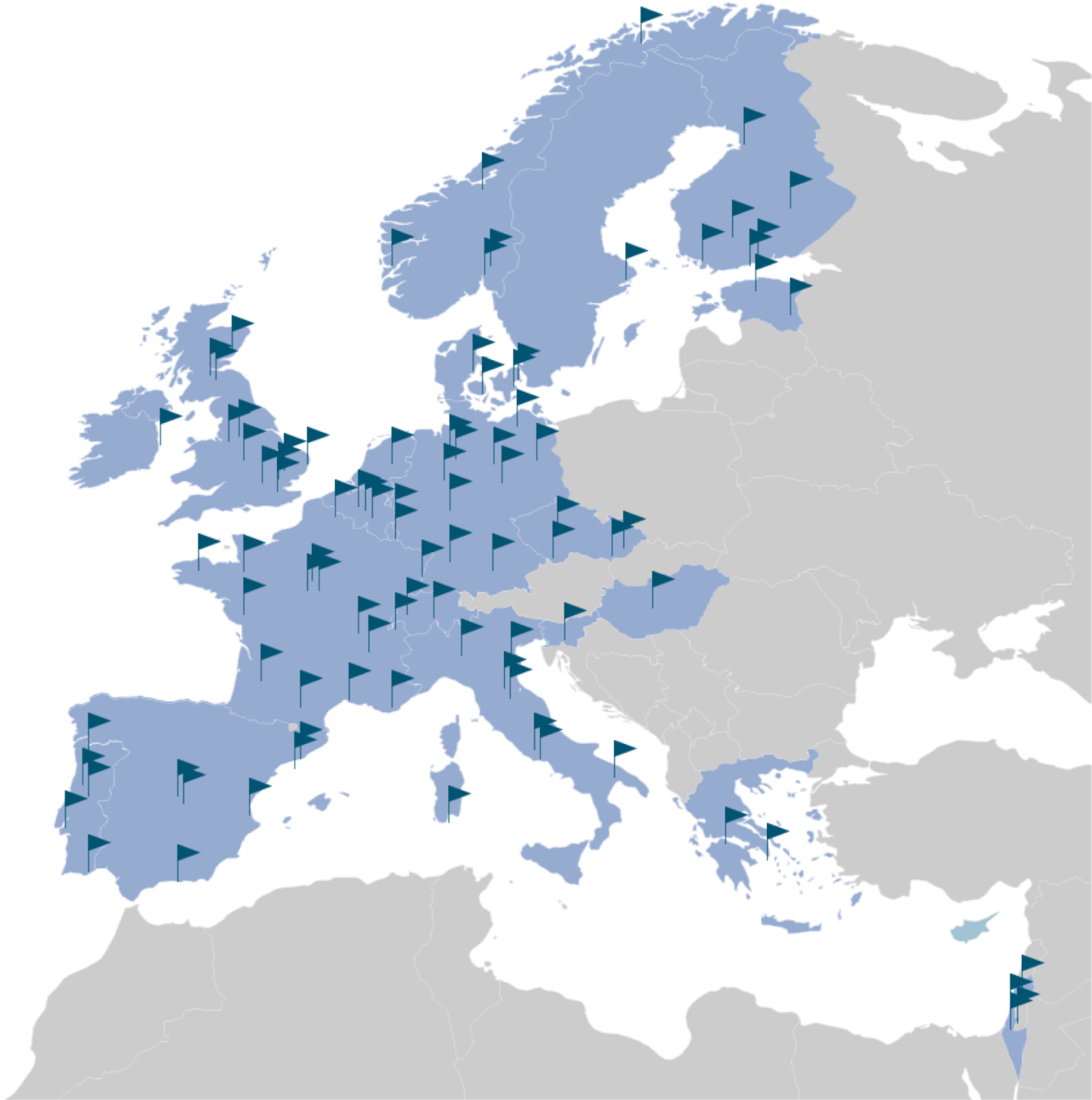
ELIXIR members

ELIXIR Members

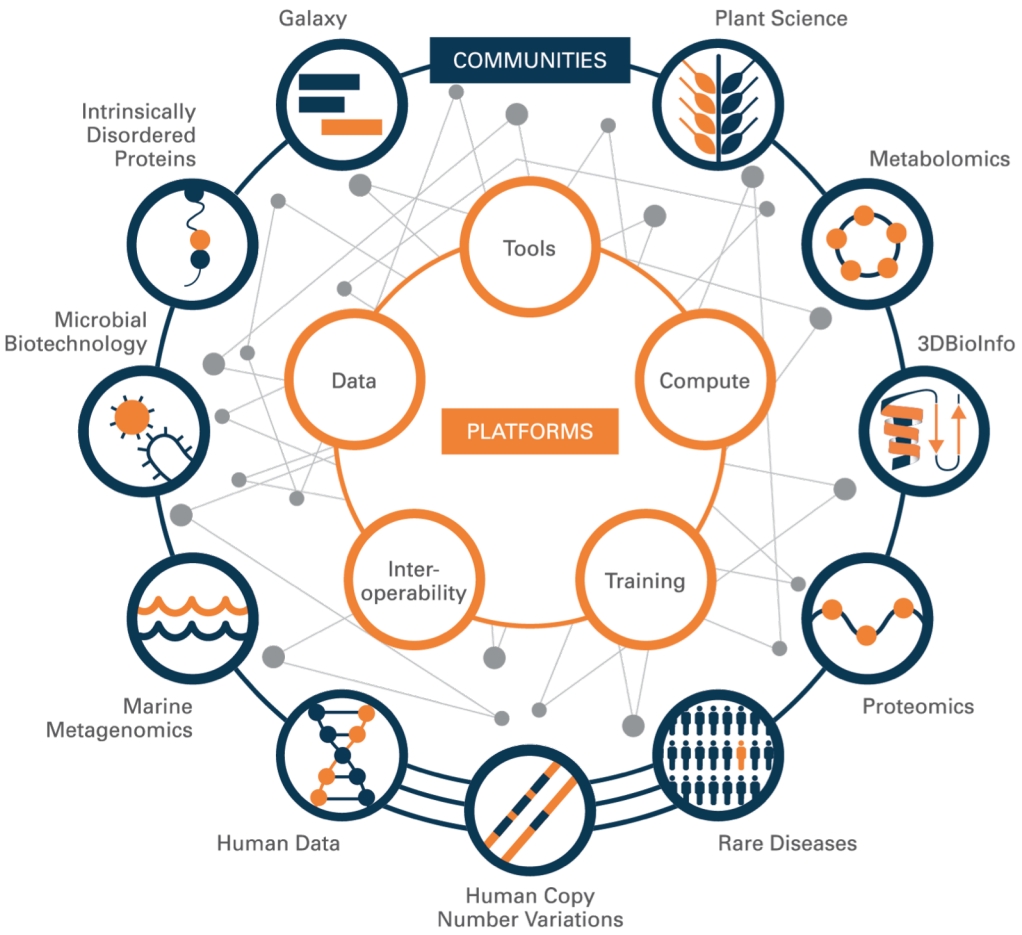
Belgium
Czech Republic
Denmark
EMBL
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Luxembourg
Israel
Netherlands
Norway
Slovenia
Portugal
Spain
Sweden
Switzerland
United Kingdom

ELIXIR Observers

Cyprus



ELIXIR structure



Five technical platforms :

- Compute
- Data
- Tools
- Interoperability
- Training

Complemented by Communities:

- Marine metagenomics
- Plants sciences
- Proteomics, Metabolomics
- Galaxy
- Human Data Communities (Rare diseases, Federated human data, Human copy number variation)
- Intrinsically Disordered Proteins
- Microbial biotechnology

Do we need ELIXIR for data management????????

Data management planning is required for all grant applications

Important to capture metadata in all parts of the project

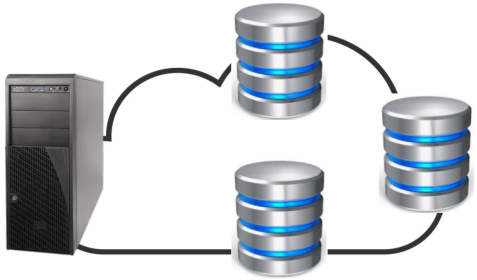
Important that the data is FAIR

Sample collection

Data production

Data storage/sharing

Data processing



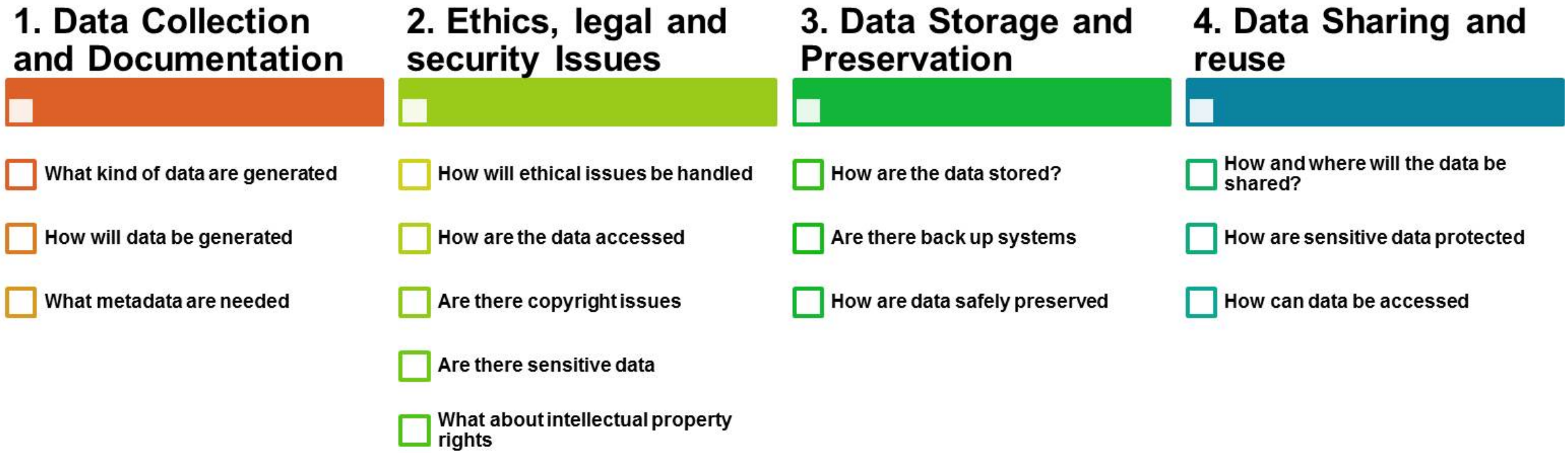
Data achieving

Data management

What is a data management plan (DMP)?

A data management plan is a document that describes how to handle research data from start to finish

Funders now request a data management plan as part of a grant application, or as one of the deliverables of a project

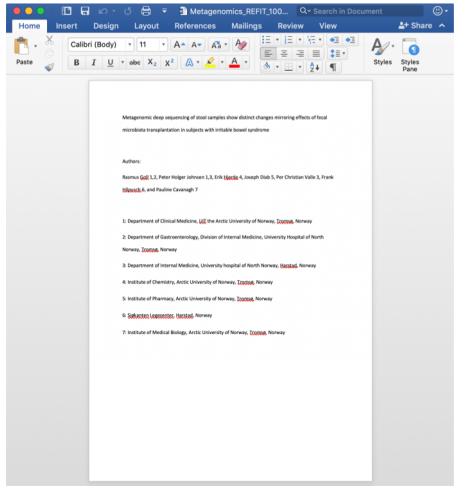


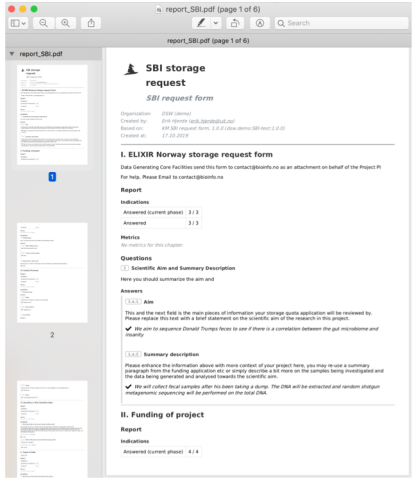
https://www.uzh.ch/blog/hbz/2018/11/15/data-management-plan-in-a-nutshell/?lang=en

What is a data management plan (DMP)?

A DMP is a document often created before a project begins

Mandatory creation of a DMP can be viewed as annoying administrative task





ELIXIR provide a tool that contain pre-made community, institutional or grant specific web form questionnaires

What ELIXIR Norway offers

HELPDESK
Norwegian Bioinformatics Platform

contact@bioinfo.no

Service and collaboration and support

Access to data, tools, compute & storage

Data management planning

Provide connection to other international ELIXIR nodes

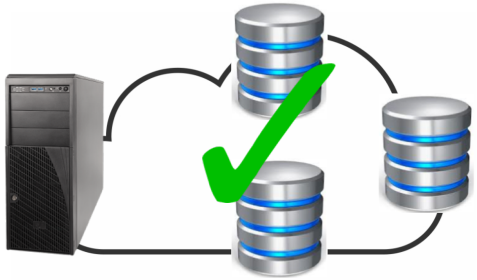
Training, courses and Hands-on workshops

Sample collection

Data production

Data storage/sharing

Data processing



Data archieving

Data management

DSW

elixir

ELIXIR Norway survey

HELPDESK
Norwegian Bioinformatics Platform

contact@bioinfo.no

We would like to use this opportunity to kindly ask you to complete a survey about ELIXIR and NeLS, which we will use to improve our services.

https://bit.ly/ELIXIRsurvey

The participants have the chance to win a gift card of 1000 NOK.