



# Tools, techniques and infrastructure – bioinformatics in marine biotechnology

Second Marine Biotechnology ERA-NET Stakeholder Meeting

13<sup>th</sup> - 14<sup>th</sup> October 2016, Hotel Marivaux, Brussels, Belgium

Nils Peder Willassen

ELIXIR Norway

UiT The Arctic University of Norway



*European Life Sciences Infrastructure for Biological Information*  
[www.elixir-europe.org](http://www.elixir-europe.org)

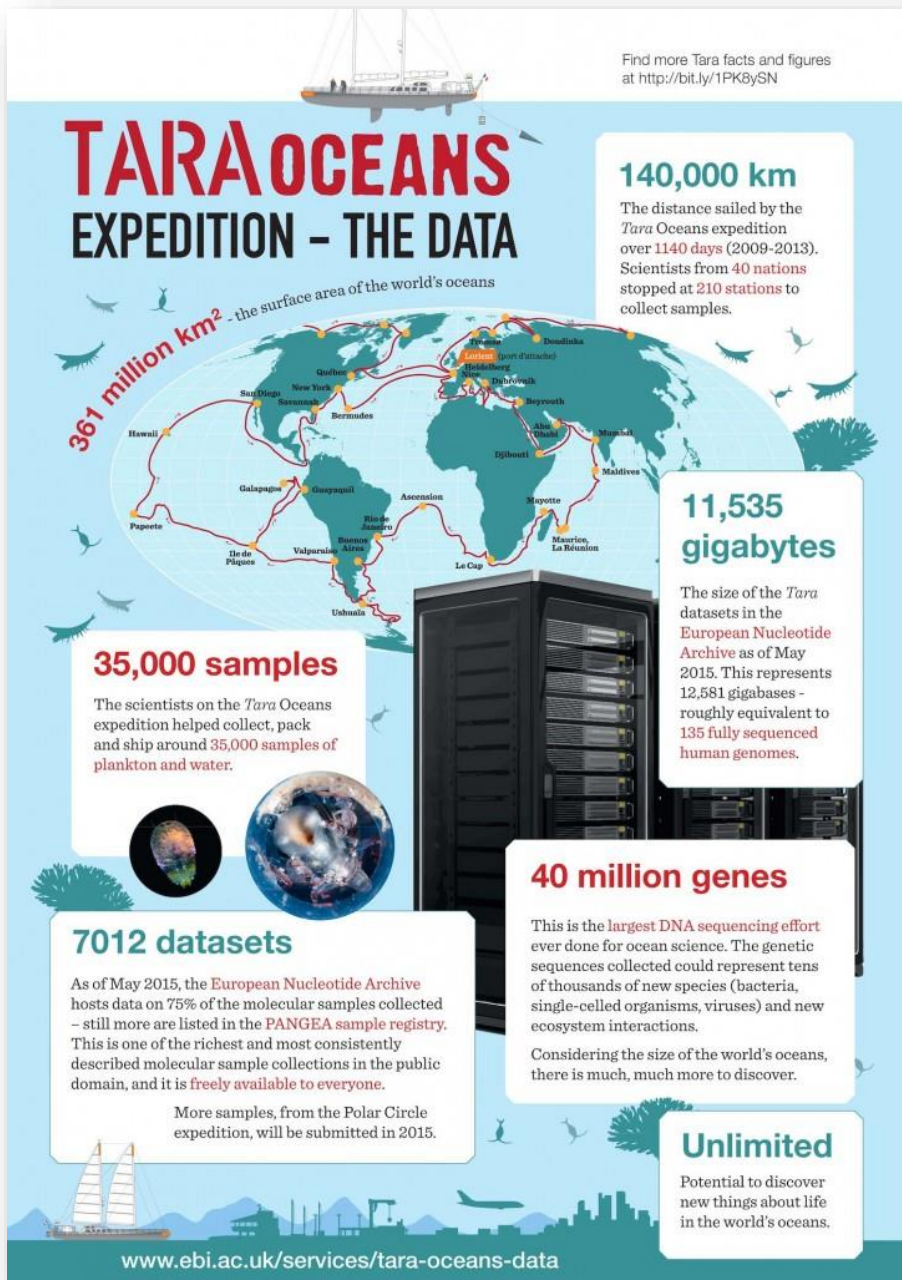


# What is Bioinformatics?

- The science of storing, retrieving and analysing large amount of biological information
- An interdisciplinary science involving biologists, biochemists, computer scientists and mathematicians
- At the heart of modern biology



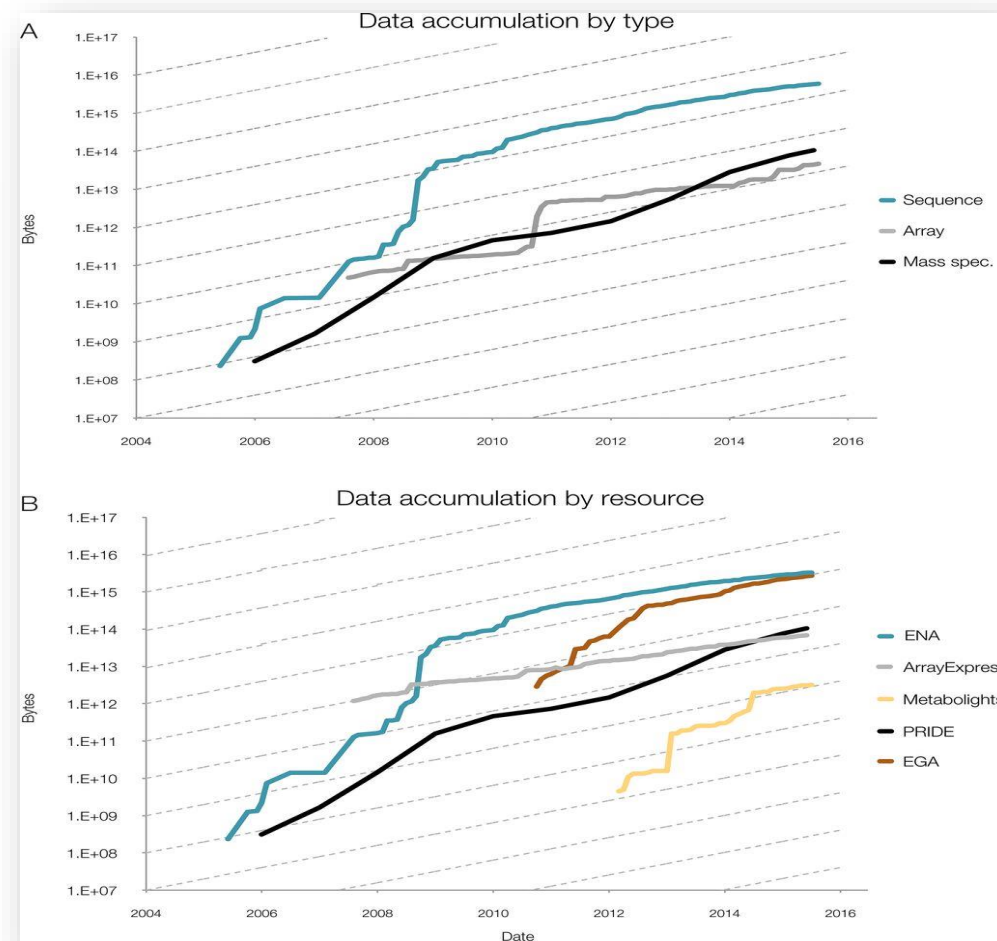
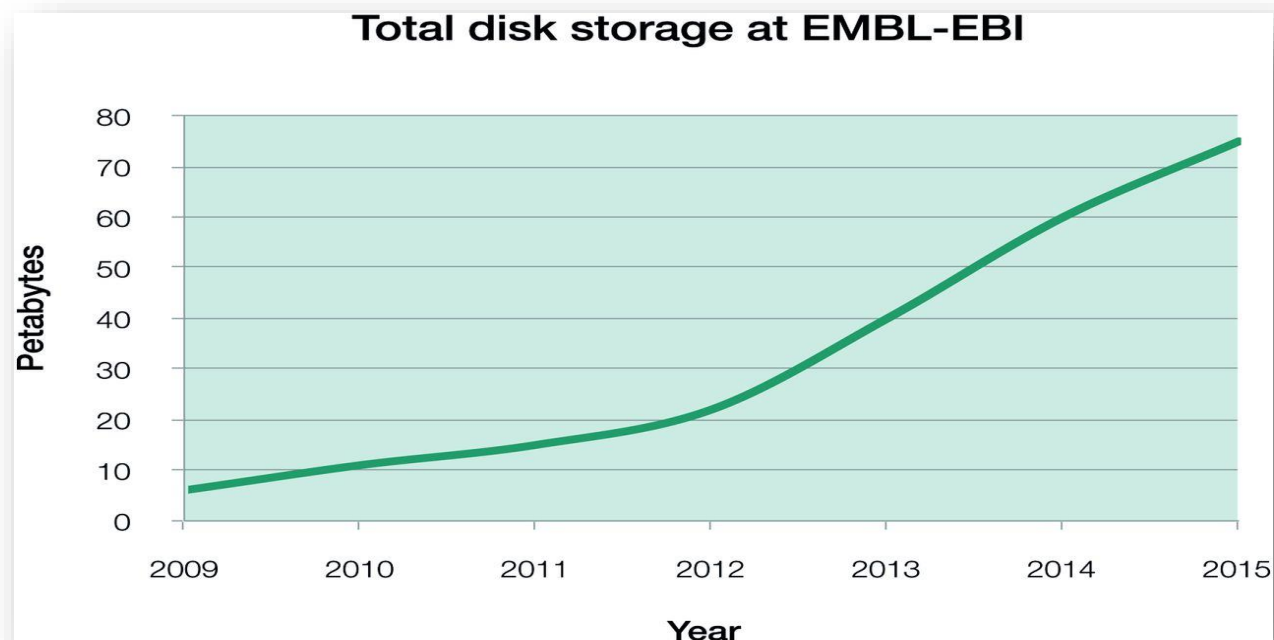
# The data challenge: Geographic spread



- Data production sites increasing across Europe
- Illumina seq. sales up 18 % in 2015
- New emerging sequencing technologies
- Number and size of projects increases



# Data growth in the life sciences






Data growth at EMBL-EBI

Source: Charles E. Cook et al. *Nucl. Acids Res.* 2016;44:D20-D26



# We generate data faster than we can deposit it

		Network file transfer rate
24 hours		100 Mb
	DNA sequencing ~100 GB	~5 hours
	Mass spectrometry ~4 TB	~4 days
	Microscopy ~4 TB	~4 days

# Data resources in life science

Non-vertebrate genomics

Structure

Metabolic and signalling p...

Nucleotide sequence

**1800**  
databases

Protein sequence

Human and other vertebrate genom...

RNA sequence

Other

Human genes and diseases

Plant

Microarray and...

Organelle

Immunol...

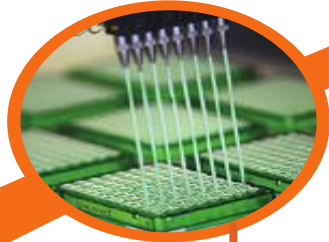
Proteomics

# ELIXIR

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



environment



bioindustries



agriculture

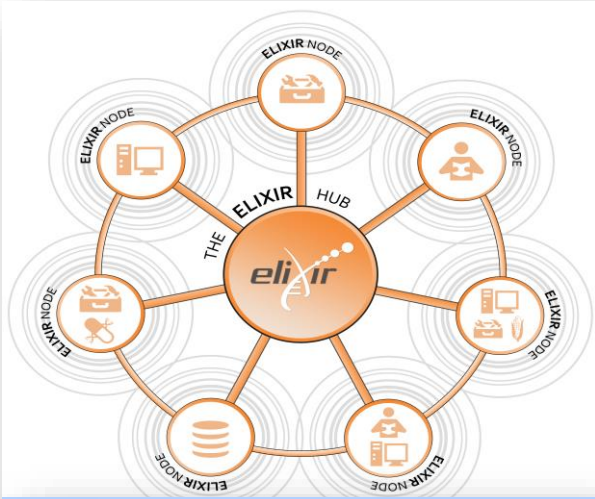


medicine

*ELIXIR underpins life science research – across academia and industry*



# *A distributed infrastructure to scale with the challenge*



- ELIXIR Nodes build local bioinformatics capacity throughout Europe
- ELIXIR Nodes build on national strengths and priorities
- 20 members and one observer
- Over 140 institutes involved in ELIXIR Nodes



# ELIXIR-EXCELERATE

## Five technical platforms



*Data*



*Interoperability*



*Tools*



*Compute*



*Training*

## Four use cases



*Marine metagenomics*



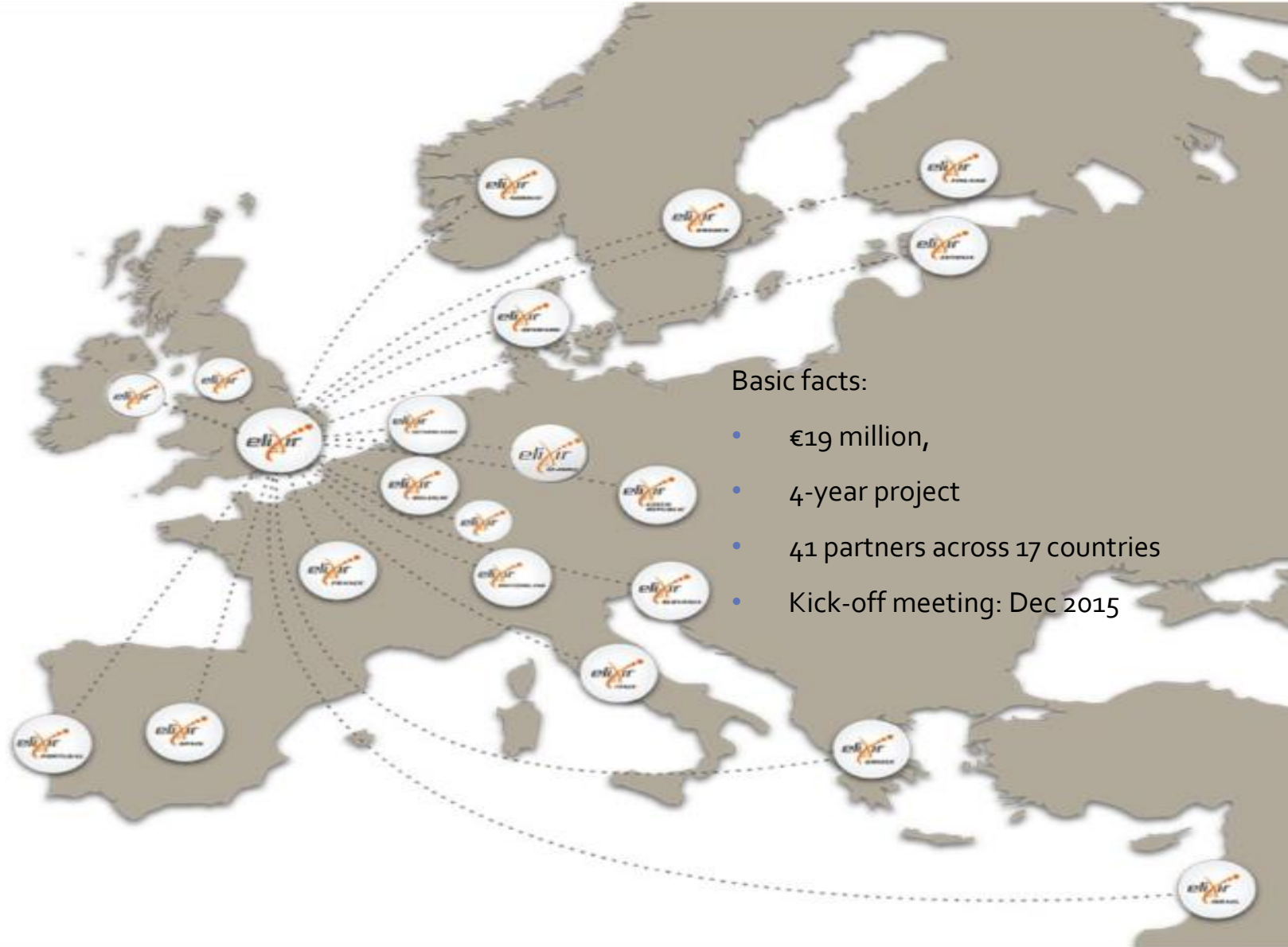
*Crop and forest plants*



*Human data*



*Rare diseases*



## Basic facts:

- €19 million,
- 4-year project
- 41 partners across 17 countries
- Kick-off meeting: Dec 2015



[www.elixir-europe.org](http://www.elixir-europe.org)



[@ELIXIREurope](https://twitter.com/ELIXIREurope)



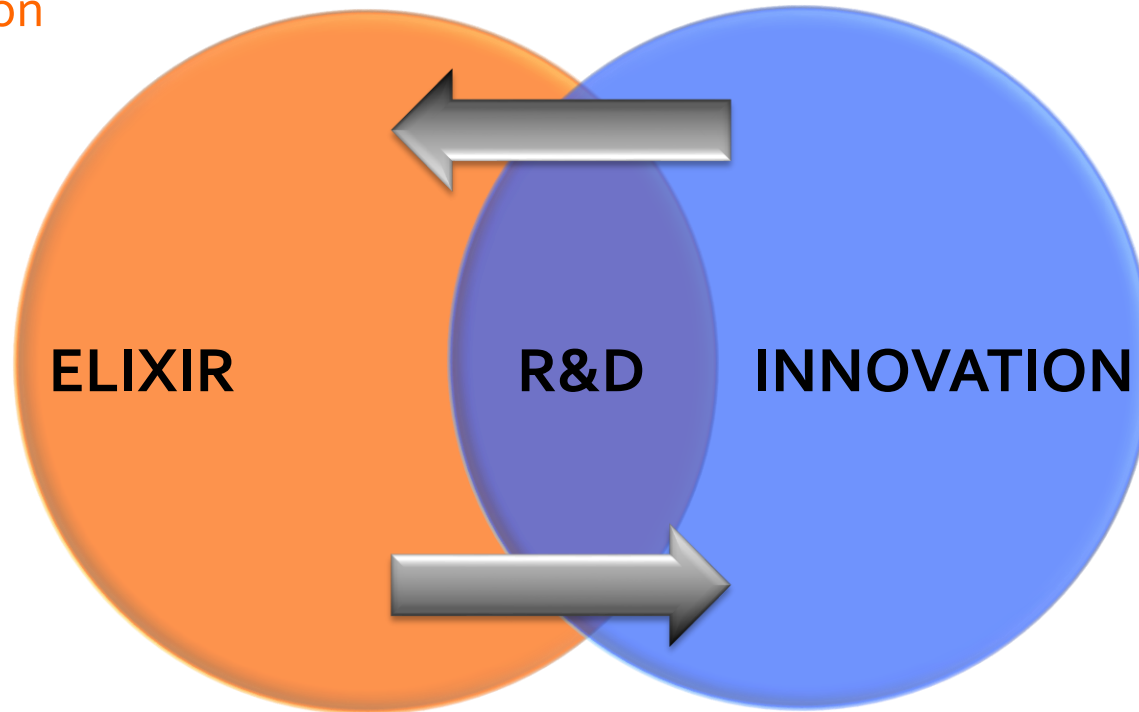
# *ELIXIR in the innovation chain*

EMBL EBI  
198 000 direct active users during 2015, accessing data 88 million times.

*ELIXIR provides FAIR data services to enable R&D and innovation*

*Data generated on research projects is captured by ELIXIR and made available again*

Over 8,000 patents in 2014 quoted bioinformatics repositories



*Innovation in life sciences creates new technologies and products e.g. bioactives*

*Upstream*

*Downstream*



# Bioinformatics in the marine domain

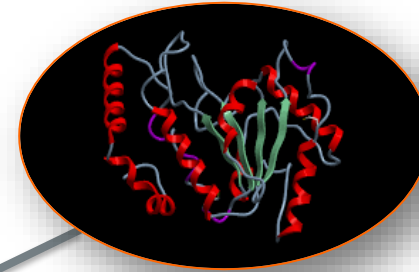
Marker-assisted selection of farmed fish



Vaccine



Bioprospecting – hunting for novel enzymes and compounds



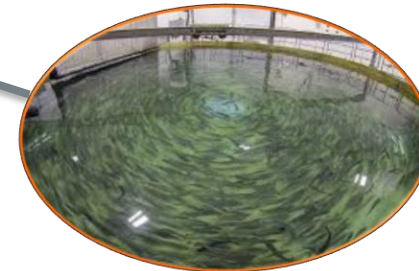
Monitoring pollution



Function and ecology of the oceans



Monitoring and optimizing RAS biofilters



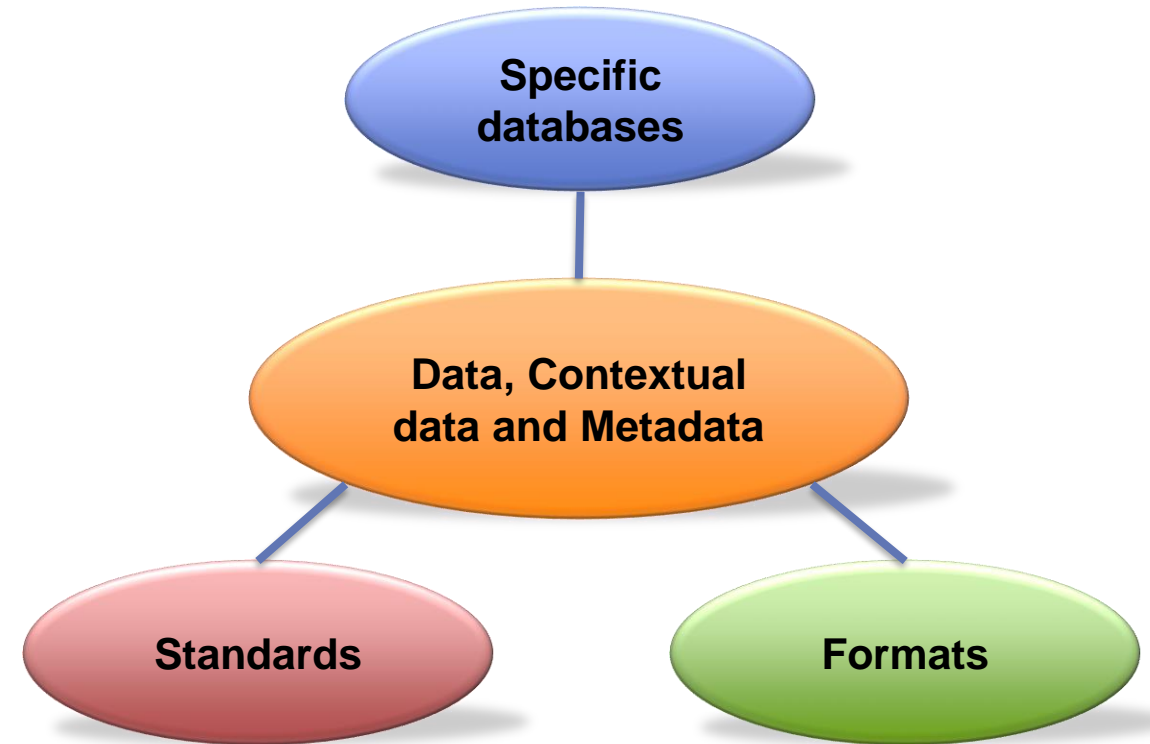
# Enhance utilization of the marine biological data

- Need meeting points
  - Stakeholders meetings, networks etc.
  - ELIXIR SME and Innovation forum
  - Collaboration (e.g. ESFRI cluster projects - EMBRIC, CORBEL)
- Training
  - Provide training to researchers and industry end users



# Enhance utilization of the marine biological data

- Domain specific data resources
  - Specialised marine databases
  - Reliable and sustainable
- Standardization
  - Data (e.g. raw sequence data)
  - Contextual data (samples and sampling)
  - Metadata (provenance and tracking)



# Enhance utilization of the marine biological data

- Build domain specific tools and pipelines for processing, analysis and visualization
  - Compute, storage and archiving
  - Access to domain specific databases
- Data integration
  - Integration of different datatypes
  - Combining data residing in different sources





# Thank you !



**Contact:**

Nils Peder Willassen: [nils-peder.willassen@uit.no](mailto:nils-peder.willassen@uit.no)

Mail-list: [marine-community-bounces@elixir-europe.org](mailto:marine-community-bounces@elixir-europe.org)

Web-site: <https://www.elixir-europe.org/excelerate/marine>

[www.elixir-europe.org](http://www.elixir-europe.org)

