



ELIXIR - A distributed infrastructure for life-science information

Nils Peder Willassen

Centre for bioinformatics

UiT The Arctic University of Norway

ELIXIR Norway

Marine Metagenomics Workshop, Nov 26-30, 2018, Tromsø, Norway





Does bioinformatics matter?

Marine Metagenomics Workshop, Nov 26-30, 2018, Tromsø, Norway



Tracking the source of infectious disease



- Methicillin-resistant MRSA (*Staphylococcus aureus*) infection is a global problem.
- Transmission can be tracked using small variations (SNPs) in genome sequences.
- Tracking technology can easily identify the source of new outbreaks within wards and throughout the world.

Genome-wide analysis of crop plants

- Population growth and climate change: major challenges to food security.
- Traditional routes to crop improvement: too slow to keep up with demand.
- Plant genomes: identify which species are most tolerant to drought, salt and pests while still providing optimum nutrition.



Matching the treatment to the cancer

- One in 10 women in the EU will develop breast cancer before the age of 80.
- Can identify patterns of genes that are active in different tumours, allows earlier diagnose and treatment of cancers



Genome-assisted breeding



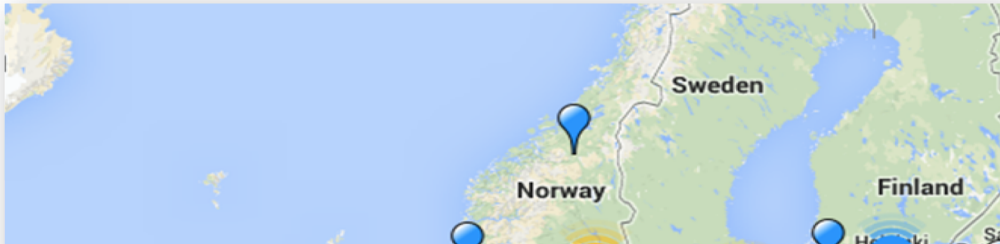
- New technologies for genetic enhancement of aquaculture species, includes marker-assisted selection (MAS), genome selection (GS), and genome editing (GE) are now in use.
- MAS allowed production of Infectious Pancreatic Necrosis (IPN)-resistant salmon, leading to a 75% reduction in the number of IPN outbreaks in the salmon farming industry.





Challenges

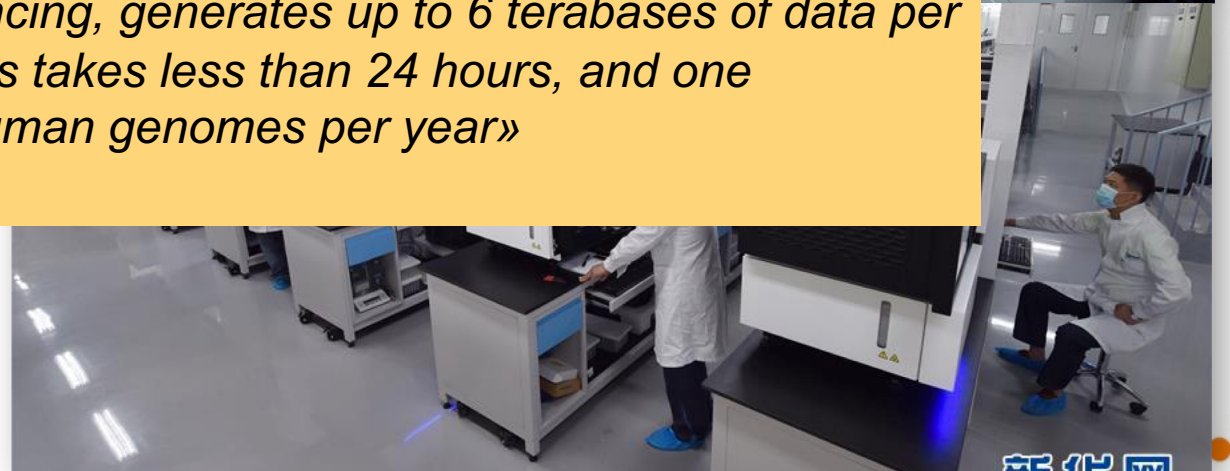
The data challenge: Geographic spread





Oct. 27.2018

MGI Introduces Groundbreaking Ultra-High-Throughput Sequencer, MGISEQ-T7

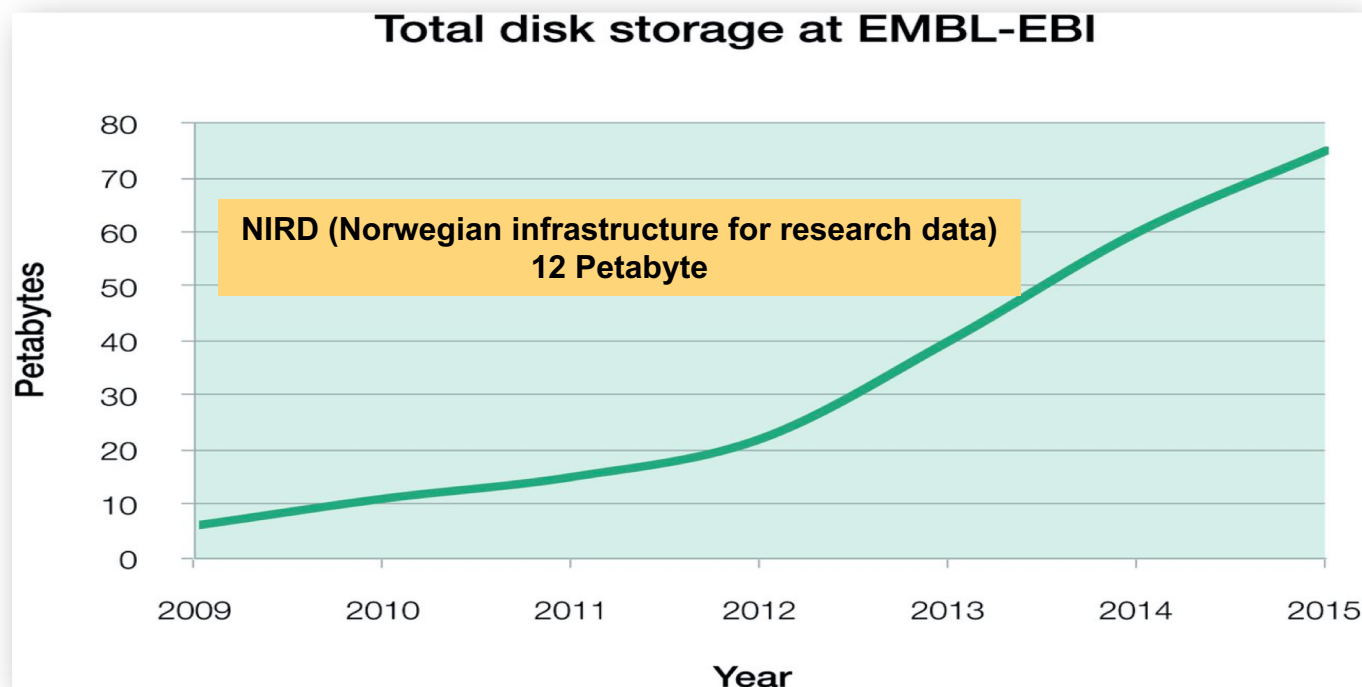
«The platform, which supports whole-genome sequencing, deep exome sequencing, epigenetic sequencing, and panel sequencing, generates up to 6 terabases of data per day. A run with 150-base paired-end reads takes less than 24 hours, and one instrument can sequence up to 10,000 human genomes per year»



We generate data faster than we can deposit it (and analyze)

		Network file transfer rate
24 hours		100 mbps
	DNA sequencing ~6TB	~6 days
In Europe ~ 15.000 hospital ~ 4.000 universities ~ 2.000 life sciences research institutes		
	~4 TB	~4 days

Data growth in the life sciences

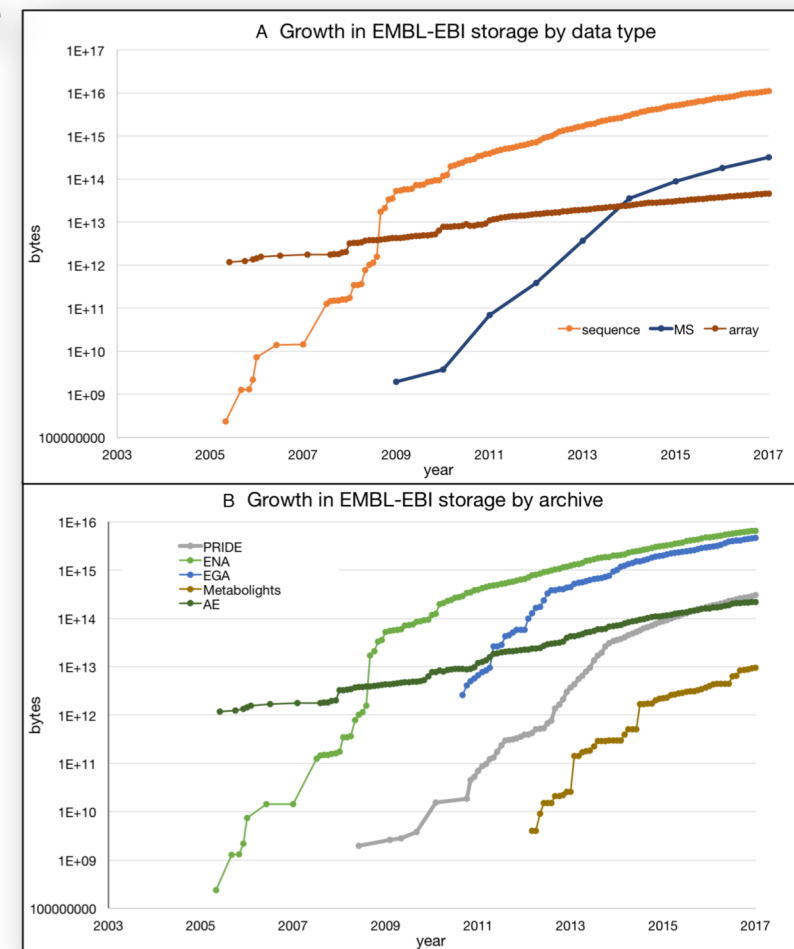


Source:
 Cook C.E. et al., *Nucl. Acids Res.* 2016;44:D20-D26
 Cook C.E. et al., *Nucl. Acids Res.* Nov 25, 2017 doi:10.1093/nar/gkx1154
 Sigmaz: <https://documentation.sigmaz.no/storage/nird.html>

Marine Metagenomics Workshop, Nov 26-30, 2018, Tromsø, Norway

2018
 ??? Petabyte

2017
 120 Petabyte



Data resources in life science

Non-vertebrate genomics

Structure

Metabolic and signalling p...

Nucleotide sequence

1685
databases

Protein sequence

Human and other vertebrate genom...

RNA sequence

Other

Human genes and diseases

Plant

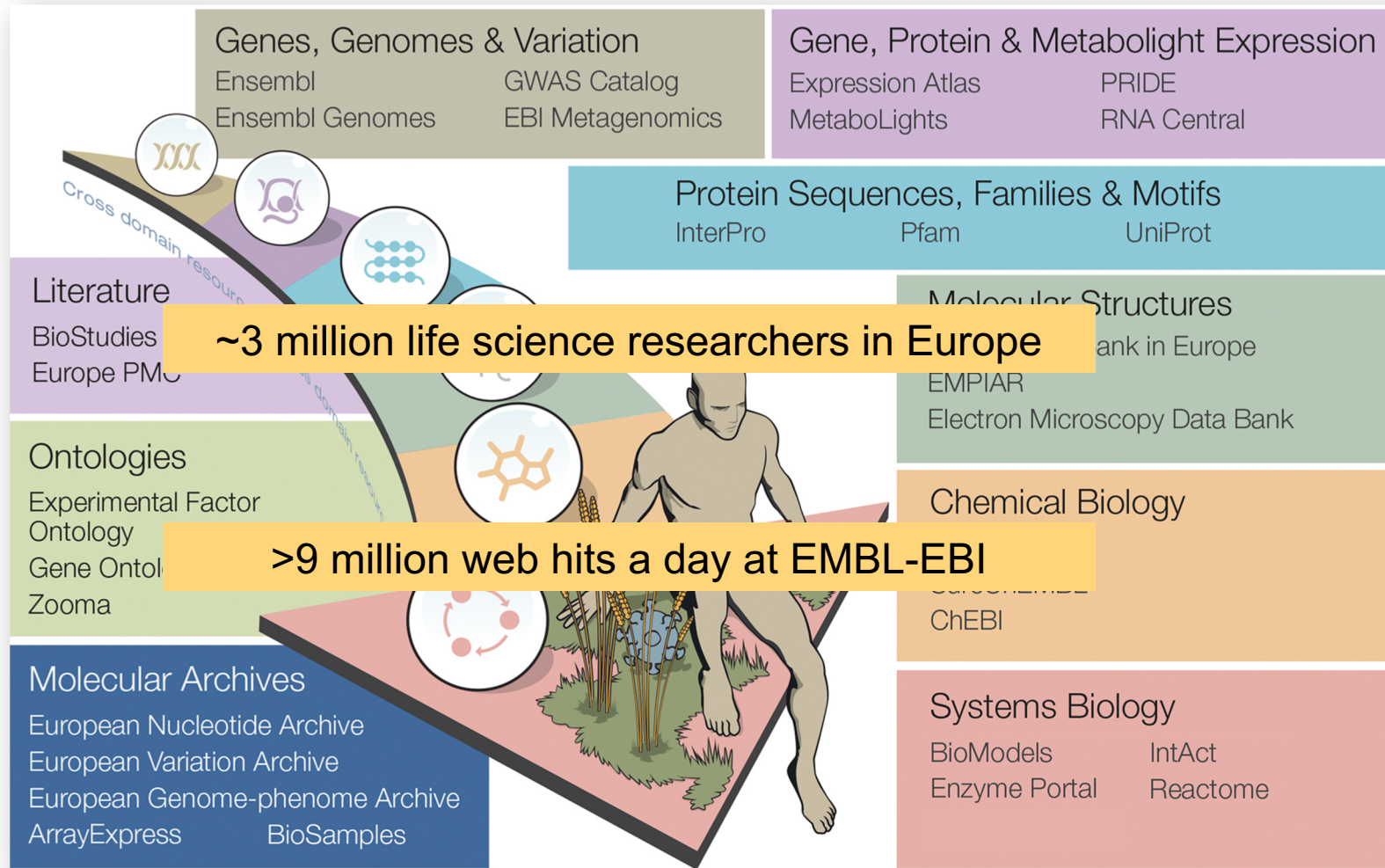
Microarray and...

Organelle

Immunol...

Proteomics

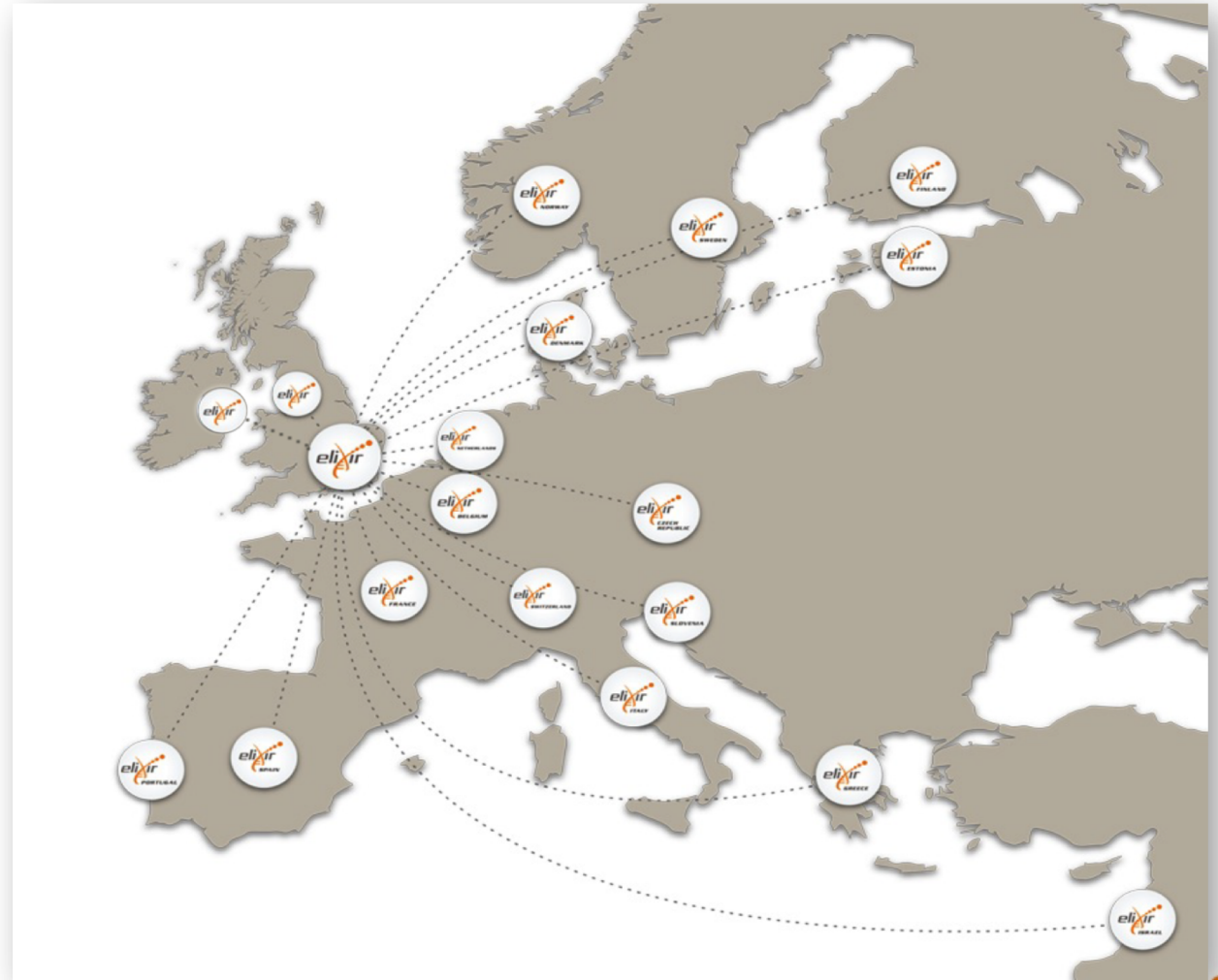
Data resources in life science





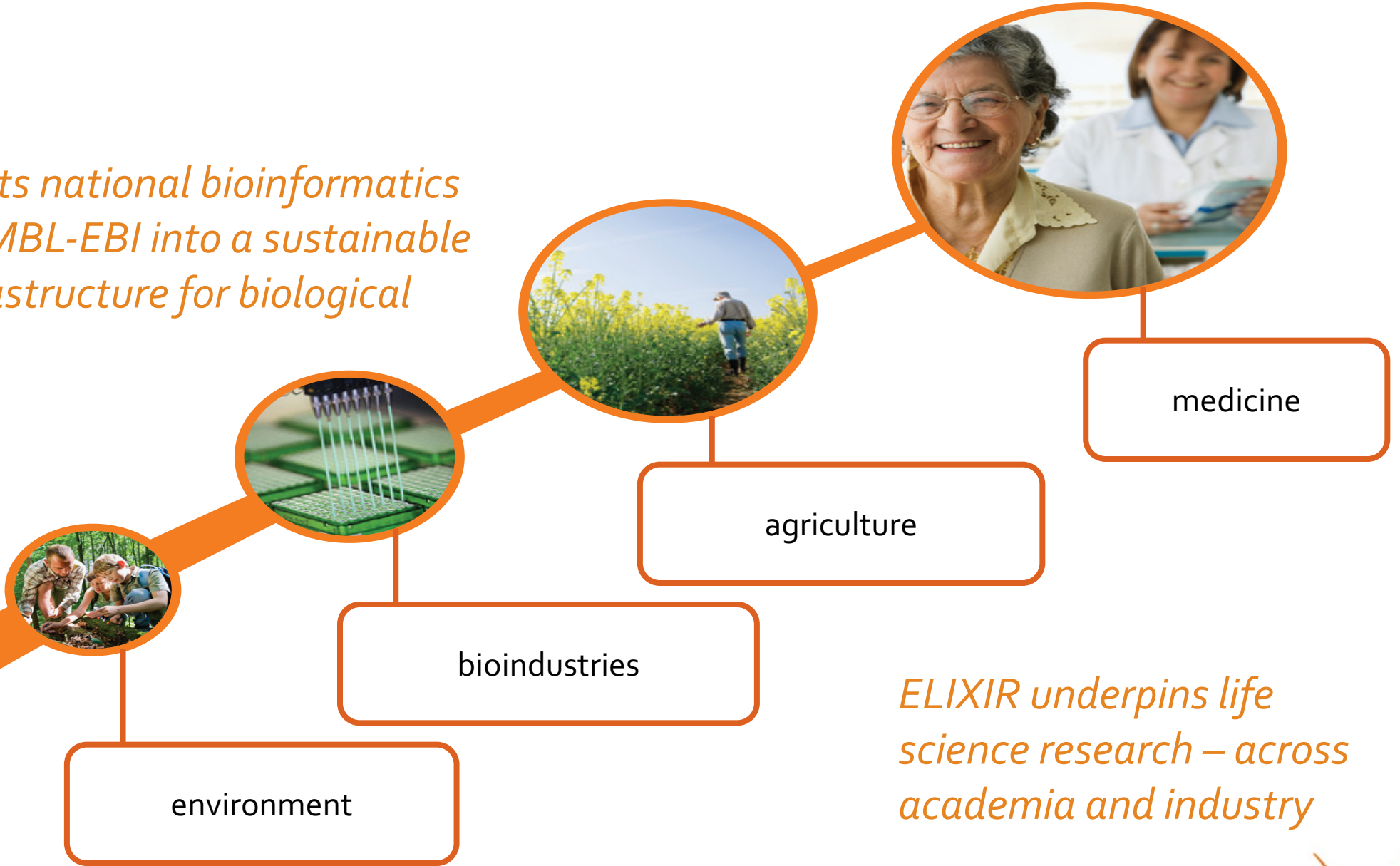
Mission:

*to establish an international
distributed infrastructure
for biological research data*



ELIXIR

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



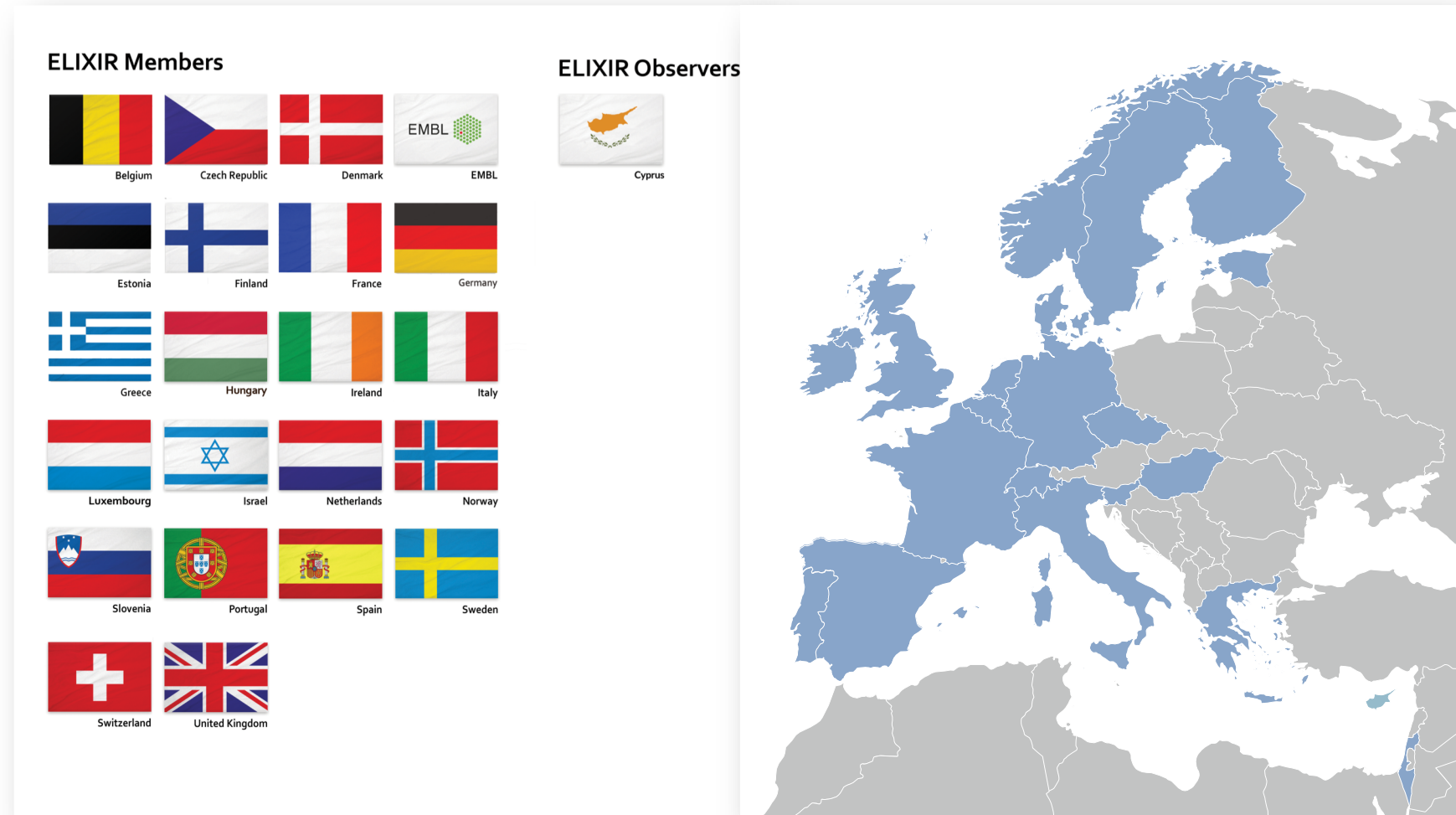
ELIXIR underpins life science research – across academia and industry



ELIXIR

ELIXIR members

- 21 national nodes (one observer) + EMBL-EBI
- > 160 institutions
- > 600 scientists involved
- Largest Research Infrastructure (RI) in Europe

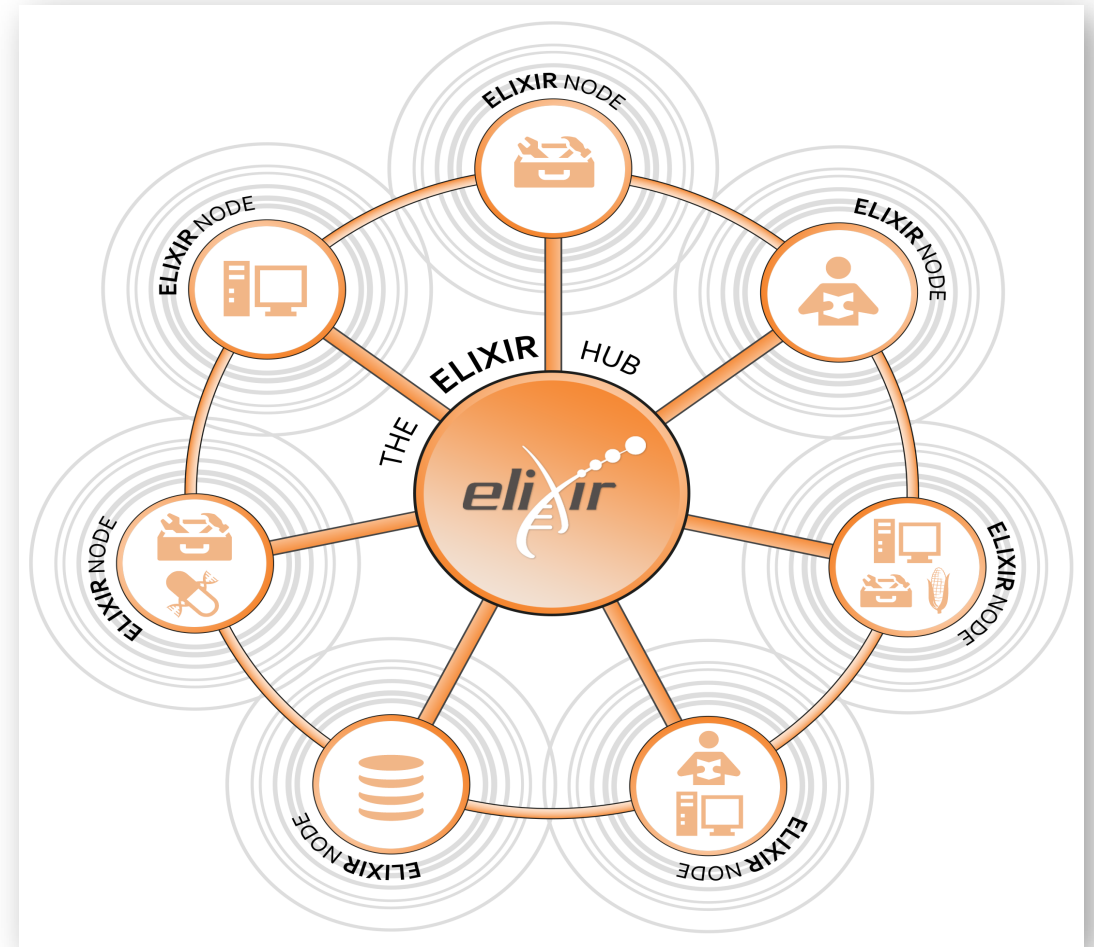


A distributed infrastructure to scale with the challenge

The **ELIXIR Hub** is like a headquarters and coordinates the work across ELIXIR.

The **ELIXIR Nodes** build local bioinformatics capacity throughout Europe

The **ELIXIR Nodes** build on national strengths and priorities





ELIXIR-EXCELERATE



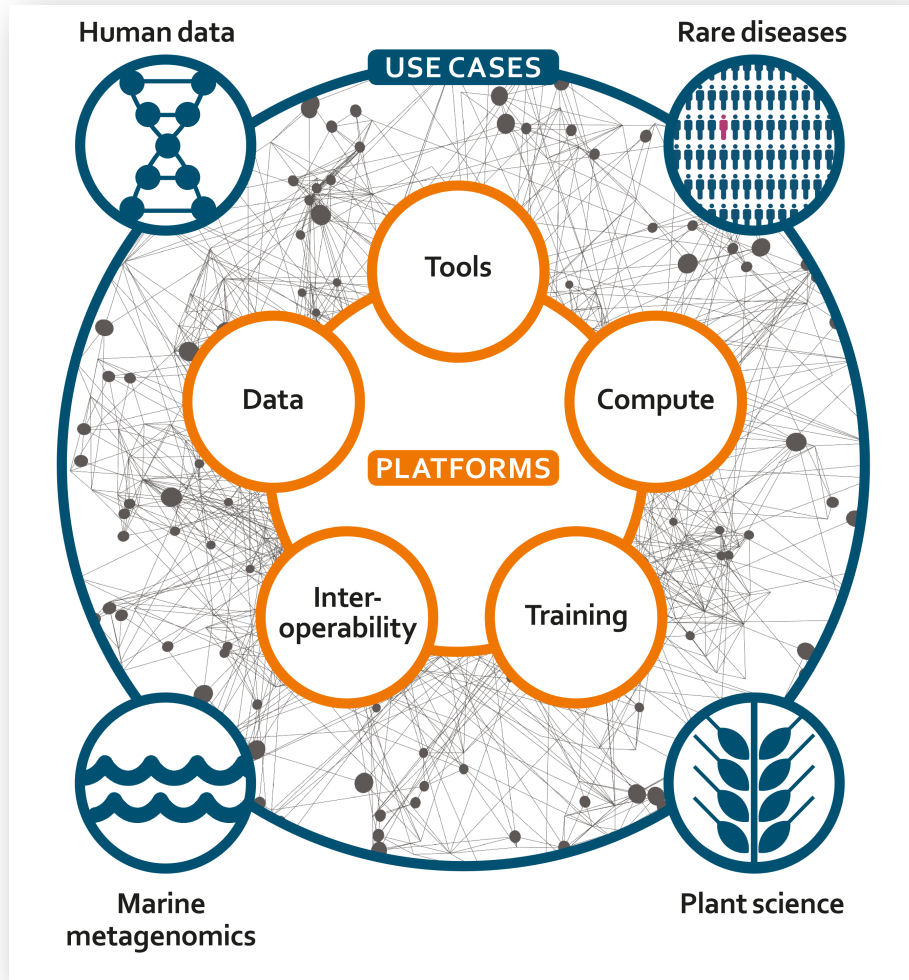
ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559.

www.elixir-europe.org/excelerate

ELIXIR EXCELERATE

- **Accelerate the implementation of ELIXIR across Nodes**
 - Develop and connect resources and services (from Nodes)
 - Build bioinformatics capacity across Europe (Capacity Building and Training)
 - €19 million, 4 year project (2015-2019), 46 partners across 17 countries





ELIXIR Structure

Five technical platforms for **Compute, Data, Tools** and **Interoperability**, in addition to the **training** platform

Complemented by four use cases for **marine meta-genomics, rare diseases, human data** and **plants sciences**

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research. See [About us](#).

News

- Bioschemas brings data closer to researchers**
4 Dec 2017
- ELIXIR data management helps "Unlock the Microbiome"**
27 Nov 2017
- ELIXIR event equips companies to advance personalised rare disease treatment**
20 Nov 2017
- ELIXIR and GA4GH agree on Collaboration Strategy**
20 Nov 2017

[More news »](#)

Services
ELIXIR services make it easier to discover, store, and analyse life science data.

Platforms
ELIXIR's activities are divided into five areas called 'Platforms'.

EU Projects
ELIXIR both collaborates in and coordinates EU projects.

Use Cases
Use Cases develop services that are specialised for particular life science domains.

Industry support
ELIXIR works with industry and SMEs to help them access life science datasets, tools, training and standards. See the [SME Programme](#).

Publications
You can view many of ELIXIR's peer-reviewed papers and posters on its [F1000 Channel](#).

Implementation Studies
ELIXIR also runs one-off short-term projects called [Implementation Studies](#).

Events

Tweets by @ELIXIREurope

Day 6 of #20DaysofNodes: @ELIXIRFinland manages a cloud service for #HPC and ELIXIR's AAI single sign-on with @ELIXIRCZ. Benefit at goo.gl/YGUDSH. Follow @CSCI and Head of Node @myronen for updates.

Day 5 of #20DaysofNodes: ELIXIR Estonia is maintaining, upgrading and integrating #bioinformatics tools and

Platforms

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. The Platforms are managed by Platform leaders and the work is carried out by groups within the Platforms.

- Data Platform**
Aims to identify key data resources across Europe and support the linkages between data and literature
- Tools Platform**
Helps researchers find the best software tools to analyse their data.
- Interoperability Platform**
Develops and encourages the adoption of standards to describe life science data.
- Compute Platform**
Develops services to make it easier to store, share and analyse large datasets.
- Training Platform**
Helps scientists and developers find the training they need, and also provides that training.

Find out more

- About us
- What we do
- Who we are
- Why ELIXIR is needed

Home »

USE CASES

- Human Data
- Rare Diseases
- Marine Metagenomics
- Plant Sciences

Use Cases

Use Cases coordinate ELIXIR's activities in four domains of life science: Human Data, Rare Disease, Marine Metagenomics and Plant Sciences. They bring together experts to develop specialised standards, services, workshops and Implementation Studies in these domains. The Use Cases also provide feedback on the Platform services, which helps ensure they are practical and useful.

Human Data
Develops long-term strategies for managing and accessing sensitive human data.

Rare Diseases
Supports the development of new therapies for rare diseases.

Marine Metagenomics
Develops a sustainable metagenomics infrastructure to nurture research and innovation in the marine domain.

Plant Sciences
Develops an infrastructure to facilitate genotype-phenotype analysis for crop and tree species.

New Use Cases: Three new Use Cases have just begun: Proteomics, Metabolomics and Galaxy. For details see the [news story](#). More details about them will appear in this section shortly.

Goals of the Use Cases

- To tie together ELIXIR services into effective workflows that support the needs of the life science community
- To guide the coordination and enhancement of ELIXIR's resources in a specific domain (e.g. Human Data, Marine Metagenomics)
- To demonstrate the effectiveness of ELIXIR services
- To test the technologies and solutions developed by ELIXIR Platforms.

The four Use Cases began in 2015 with the start of the [ELIXIR-EXCELERATE](#) grant and will run at least until 2019. The [Support](#)

Home » Use Cases »

USE CASES

- Human Data
- Rare Diseases
- Marine Metagenomics
- Plant Sciences

Marine Metagenomics Use Case

Marine metagenomics is the study of genetic material recovered directly from the sea. It is a new and rapidly expanding area of research, and there is a danger that data is produced faster than users are able to share, analyse and interpret it. There is an urgent need to create a data management infrastructure dedicated to marine research.

This Use Case aims develop a sustainable metagenomics infrastructure to enhance research and industrial innovation within the marine domain. It will provide a web based portal through which researchers can access a marine reference database. The database will be populated with data from the [European Nucleotide Archive \(ENA\)](#), [UniProt](#) and other sources.

What the Use Case does

- Develops and implements data standards for the marine domain.
- Develops and implements databases specific to marine metagenomics.
- Evaluates and implements tools and pipelines for metagenomics analyses.
- Develops a search engine for the interrogation of marine metagenomics datasets.
- Organises training workshops for end users.

Services provided

- EBI metagenomics portal: portal for metagenomics analysis:** an automated pipeline for analysing and archiving metagenomic data.
- Marine Metagenomics Portal (MMP) (Beta version)**
 - MarRef:** a database for completely sequenced marine prokaryotic genomes.
 - MarDB:** a database of sequenced marine prokaryotic genomes regardless of level of completeness.
 - MarCat:** a catalogue of marine genes and proteins derived from metagenomics samples.
 - META-pipe:** a marine metagenomics analysis pipeline.
- ITSoneDB:** a database for fungal ITS1 sequences

[Support](#)

ELIXIR Communities

elixir ABOUT US SERVICES PLATFORMS **COMMUNITIES** EVENTS NEWS INTRANET

Home » Communities »

COMMUNITIES

- Human Data
- Rare Diseases
- Marine Metagenomics**
- Plant Sciences
- Metabolomics
- Proteomics
- Galaxy

Marine Metagenomics Community

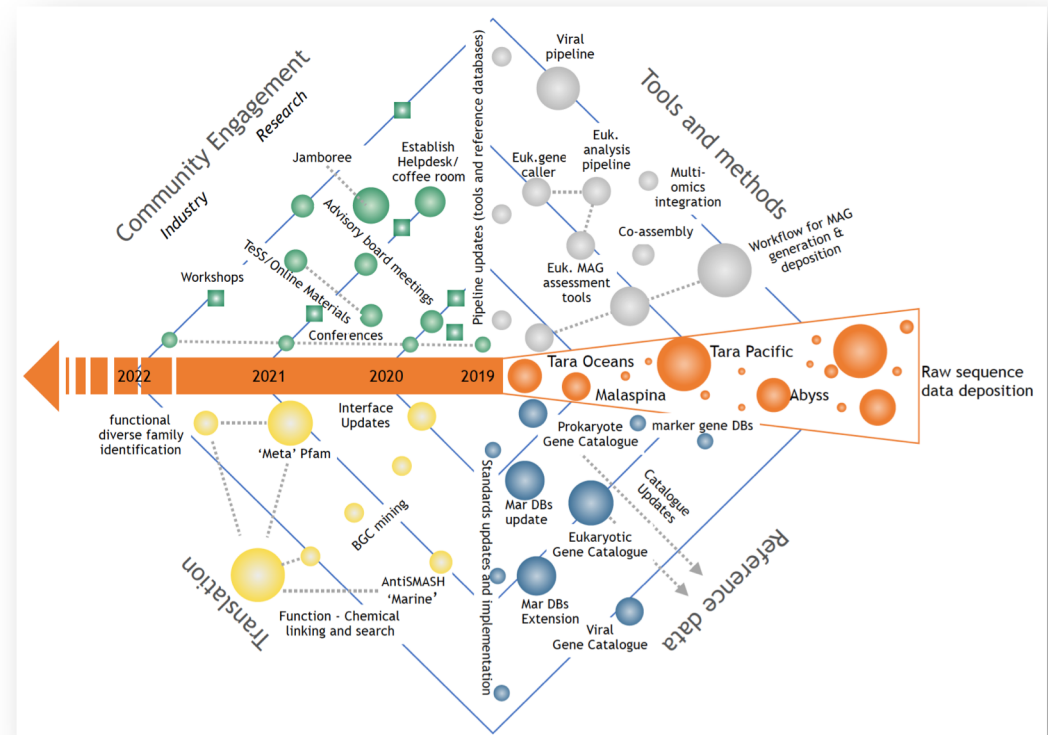
The Oceans represents the largest biosphere (~97%) on earth. The micro-organisms present within the oceans are responsible for fundamental processes of life, such as nutrient cycling, and produce 50% of the world's oxygen. However, these organisms are poorly represented in databases.

Marine metagenomics is the study of the genomes of these micro-organisms, as sampled directly from the ocean environment. It is a new and rapidly expanding area of research, and there is a danger that data is produced faster than users are able to analyse, interpret and share it. There is an urgent need to create a data management infrastructure dedicated to marine metagenomics research.

This Community aims to develop a sustainable metagenomics infrastructure to enhance research and industrial innovation within the marine domain.

Services offered

- [Marine Metagenomics Portal \(MMP\)](#): freely accessible microbial genomics and metagenomics reference databases that include MarRef, MarDB, and MarCat. The MMP also includes a search engine that you can use to query all three databases at once.
- [EBI metagenomics](#): an automated pipeline for analysing and archiving metagenomic data. You can use a range of services to search and retrieve hundreds of thousands of datasets and billions of annotations.
- [ITSoneDB](#): a database for fungal ITS1 sequences.





Data deposition:
ENA, EGA, PDBe, EuropePMC, ...



Compute:
Secure data transfer, cloud computing, AAI



Data management:
Genome annotation
Data management plans



Bioinformatics tools:
Bio.tools



Added value data:
UniProt, Ensembl, OrphaNet, ...



Industry:
Innovation and SME programme
bespoke collaborations



Data Interoperability:
BioSharing, identifiers.org and OLS



Training:
TeSS, Data Carpentry,
eLearning

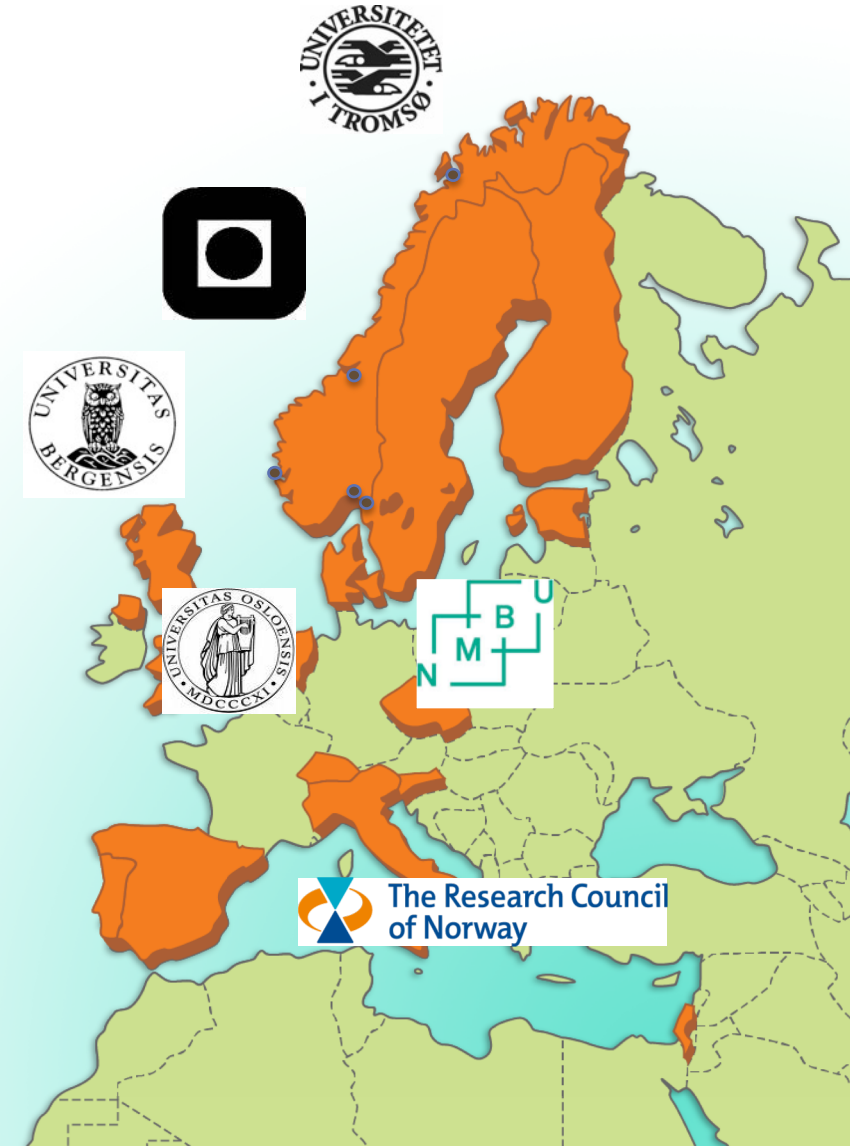


ELIXIR Norway

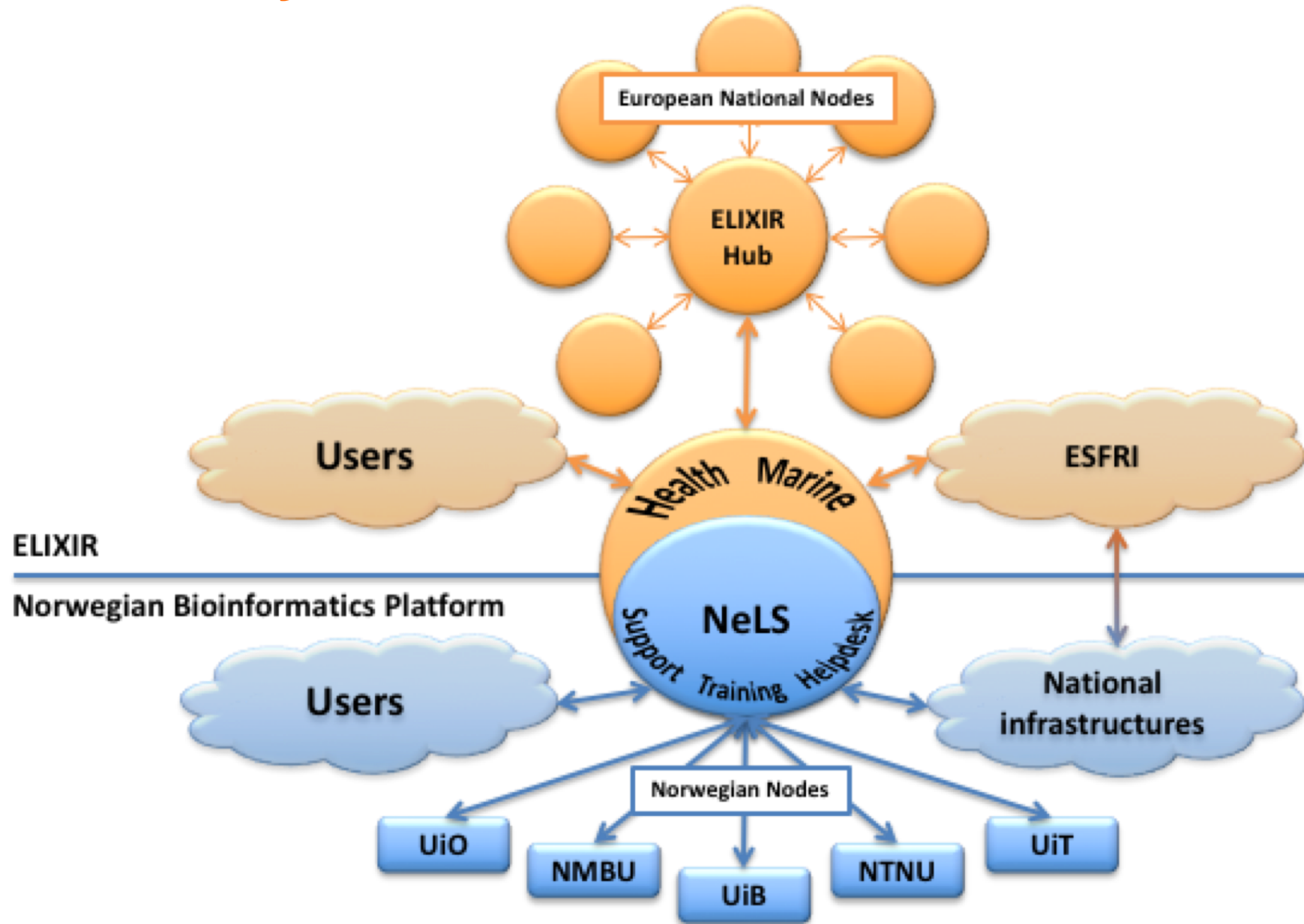
Build and offer an e-infrastructure for users within molecular life science (NeLS)

Provide state of the art bioinformatics support

Work tightly with other national and international technology platforms



ELIXIR Norway



ELIXIR Norway

Long term and temporary storage
Sharing of data with collaborators

The image shows two overlapping browser windows. The top window displays the NeLS homepage at <https://nels.bioinfo.no/welcome.xhtml>. The bottom window shows the NeLS Galaxy installation at <https://galaxy-uit.bioinfo.no>. The Galaxy interface includes a navigation menu with options like 'Analyze Data', 'Workflow', and 'Shared Data'. A central banner reads 'Welcome to the NeLS Galaxy installation in Tromsø'. A sidebar lists various tools such as 'GENERAL GALAXY TOOLS', 'Get Data', 'Send Data', 'Collection Operations', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Metagenomics', 'Statistics', 'NGS: QC and Manipulation', 'NGS: Picard', 'NGS: Mapping and Sequence analysis', 'NGS: GATK Tools', 'Assembly and Validation', 'Transcriptomics', 'NGS: SAM-tools', and 'Genome Diversity'. A 'Tweets by @elixirnorway' widget is visible on the right side of the Galaxy page.

NeLS
Norwegian e-Infrastructure for Life Sciences

Logged in as: Nils P V

Projects

You are at: /Projects

Name	Creation Date	Role in Project
Nofima_Karlsen_RNAseq_Metagenome_Aquaculture_salmon_2018	November 15, 2018	Power User (Manage File system: Add, Rename, Navigate, Download, Delete all Content)
SfB_sequencing_facility_core_sequence_storage_2018	October 5, 2018	PI (Manage File system: Add, Rename, Navigate, Download, Delete all Content) members
UiT_bioinformatics_platform_core_storage_facilities_2018	October 4, 2018	PI (Manage File system: Add, Rename, Navigate, Download, Delete all Content) members
UiT_Marine_sediment	August 22, 2018	Member (Add files & folders, Navigate & Download)
UiT_ONT	May 16, 2018	Power User (Manage File system: Add, Rename, Navigate, Download, Delete all Content)
Uio-Test-Project	March 19, 2015	Member (Add files & folders, Navigate & Download)
Ellixir_NO-Documents	November 28, 2014	Member (Add files & folders, Navigate & Download)

NeLS galaxy-uit.bioinfo.no

Analyze Data Workflow Shared Data Visualization Help User

Visualize datasets

Welcome to the NeLS Galaxy installation in Tromsø

Galaxy is a web-based platform for data intensive life science research that provides users with a unified, easy-to-use graphical interface to a host of different analysis tools. These tools can be run interactively, one by one, or combined into multi-step workflows that can be executed as a single analysis.

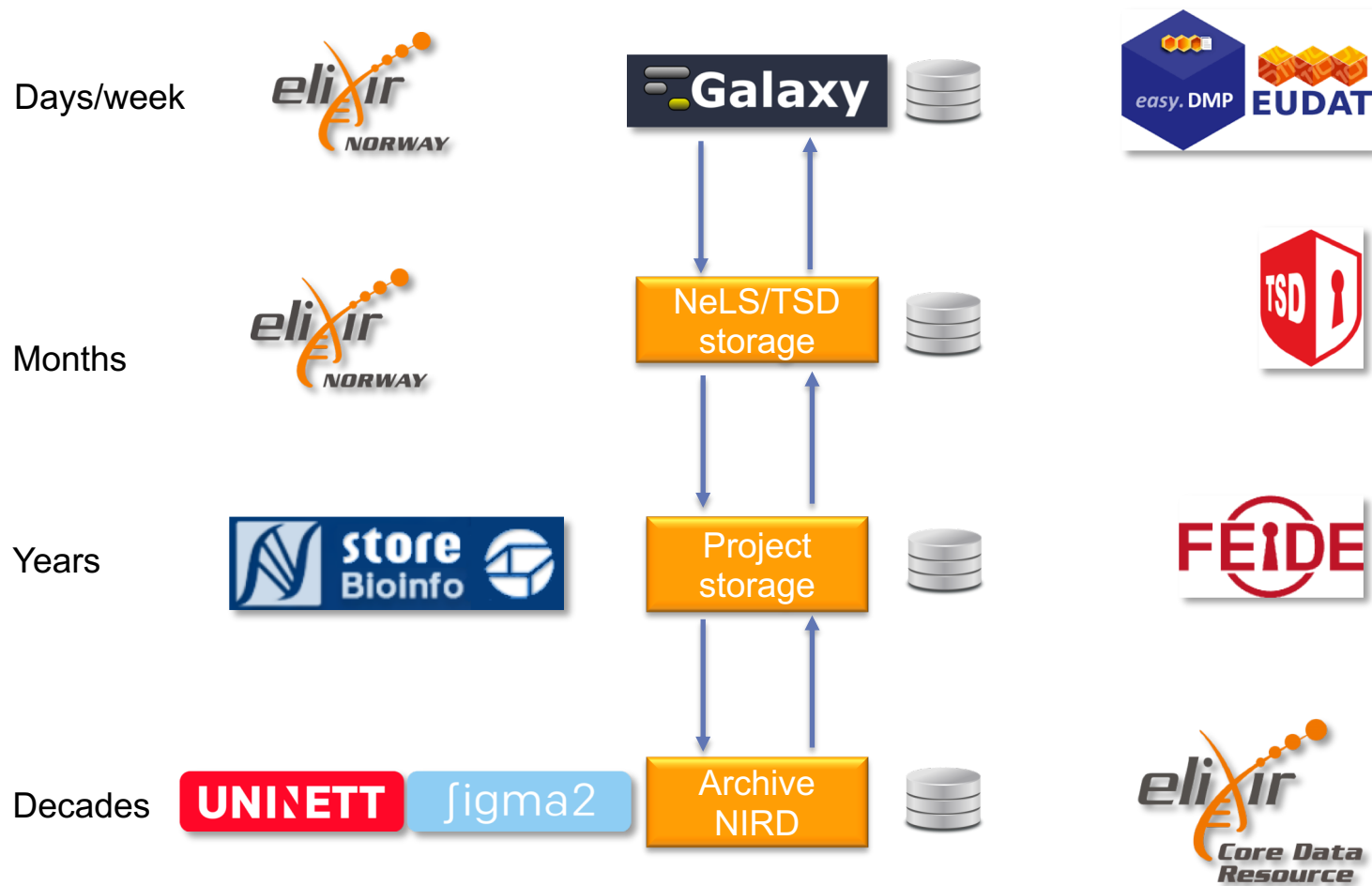
If this is your first time using Galaxy, you might want to have a look at this [Quick Start Guide](#)

Additional documentation and tutorials on using Galaxy can be found [here](#).

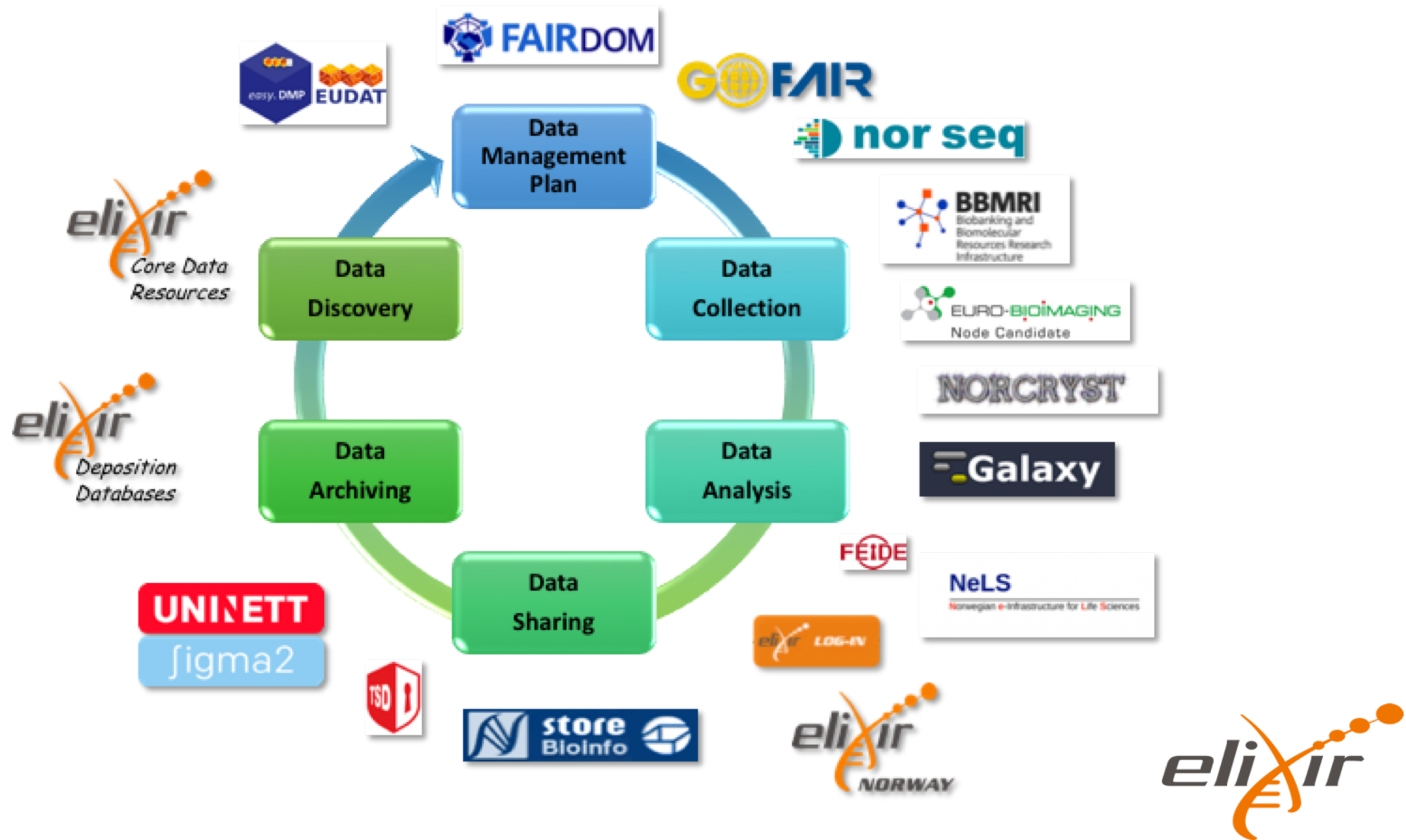
Tweets by @elixirnorway

- ELIXIR Norway Retweeted
- ELIXIR Belgium @ELIXIRnodeBE
- Follow our one-day workshop which will introduce you to the Galaxy user interface using NGS data: bit.ly/2CIKQdH
- Oct 26, 2018
- ELIXIR Norway Retweeted
- Rafael C Jimenez
- Embed View on Twitter

ELIXIR Norway



ELIXIR Norway



ELIXIR Norway

Data resources

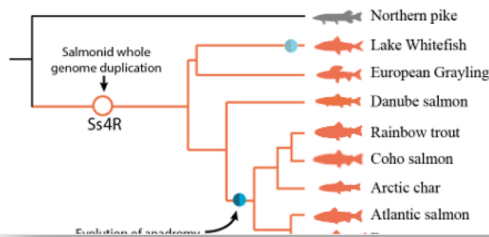
https://salmobase.org

SalmoBase - Integrated molecular resource for Salmonid species

Home Species Related Projects Download About Contact Release Updates

SalmoBase

The Atlantic salmon (*Salmo salar*) is arguably one of the most iconic and societal, economic, and ecological importance. Its mammalian sized genome since a whole genome duplication event ~80 million years ago, which cre... process of rediploidization. Today, a genome assembly is publically avail... (N50 = 3.18Mb) are ordered into 29 chromosome files. On the SalmoBas... exploring the chromosome sequence. Using public RNAseq data we hav... high confidence protein coding genes, wherein almost half show splice v... links to expression levels, gene and SNP information.



https://licebase.org

Search Genes and Sequences In-situ images RNAi Forum About BLAST Dow

Sea Lice Research Centre LiceBase

Feide login

Search

Documentation

NAVIGATION

- Browse in-situ images
- Forums
- Galleries
- RNAi Gallery
- Search Data
 - Features
 - Organisms
 - Publications
 - Advanced Feature search
 - Features, basic

Welcome to LiceBase

An Open Letter to the Steering Committee of SeaLice 2018

- BLAST
- RNA interference
- GBrowse
- Biological data
- Publications

Sea lice (*Lepeophtheirus salmonis* and *Calligus* spp.) are the major pathogens affecting the global salmon farming industry and have a significant impact in many areas. The annual costs for sea lice control have recently been estimated to exceed €500 million (only for Norway) and the aquaculture industry relies heavily on a few chemotherapeutants. Emerging resistance development against these drugs increases the necessity to develop new treatment methods (biological, prophylactic and drugs) and tools to avoid increased loss due to sea lice to ensure a sustainable salmon farming industry in the future.

- More information on the Salmon Louse.
- If you are looking for more data on Salmon or the Atlantic salmon which is the host of the salmon louse, please visit SalmoBase, our sister database.

https://mmp.sfb.uit.no

Statistics: 2141 MMP and META-pipe portals unique visitors

SERVICES DOCUMENTATION COMMUNITY HELP CONTACT HELPDESK

MARINE METAGENOMICS PORTAL

The mission of Marine Metagenomics Portal (MMP) is to provide the marine scientific community with high-quality curated and freely accessible microbial genomics and metagenomics resources.

- MAR DATABASES**
The MAR (MARine) databases are richly annotated and manually curated contextual and sequence databases. MarRef contains completely sequenced marine prokaryotic genomes. MarDB includes all non-complete marine prokaryotic genomes regardless of level of completeness. MarCat is a catalogue of marine genes and proteins derived from metagenomics samples.
Browse
- META-PIPE**
META-pipe is a complete workflow for the analysis of marine metagenomic data. It provides assembly of high-throughput sequence data, functional annotation of predicted genes, and taxonomic profiling. META-pipe is not released as an ELIXIR service yet. For now you may use the NeLS META-pipe service.
Run
- MAR BLAST**
MAR BLAST provides BLAST search on all genes and protein coding sequences from the marine databases MarRef, MarDB and MarCat.
BLAST

MMP is part of the ELIXIR infrastructure | © 2018 SFB - UIT | Terms and conditions



ELIXIR Norway

Service

Support

Access to data, tools, compute & storage

Collaboration

Small & large scale projects

Provide connection to other international ELIXIR nodes

Training

Training courses

Hands-on workshops





contact@bioinfo.no

ELIXIR Norway

ELIXIR

The platform is a node in ELIXIR, a pan-European distributed infrastructure for life-science information.




www.elixir-norway.org

www.bioinfo.no
contact@bioinfo.no



Common point of access:
contact@bioinfo.no

