



ELIXIR

Safeguarding the results of life science
research in Europe

Niklas Blomberg, ELIXIR Director

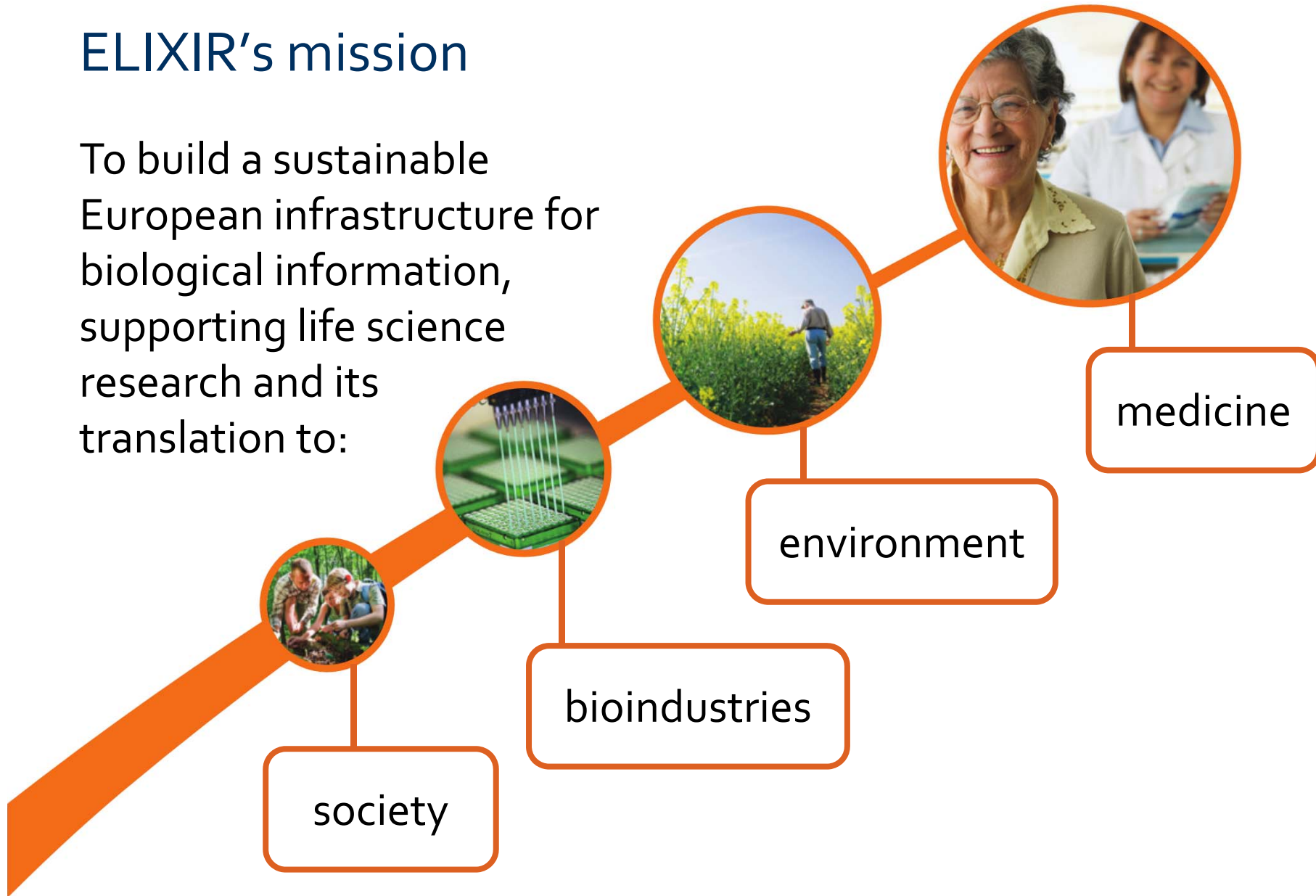
European Life Sciences Infrastructure for Biological Information

www.elixir-europe.org



ELIXIR's mission

To build a sustainable European infrastructure for biological information, supporting life science research and its translation to:



Life-science and data infrastructure in 5 years

- Data production and using at a large number of sites across Europe
 - (Illumina HiSeq sales up 40% in Q1 2013?)
- Human genomics projects but also plants, microbiota, environmental marker organisms
- Metabolomics & Proteomics coming of age
 - UK National Phenome facility
- 26 M people employed in health care sector?
 - Integrating genomics into nursing practice
- Be scalable to 1000s of sites
- Deal with incomplete, conflicting, and incorrect data



A distributed infrastructure that serves society, industry and users at a large number of sites

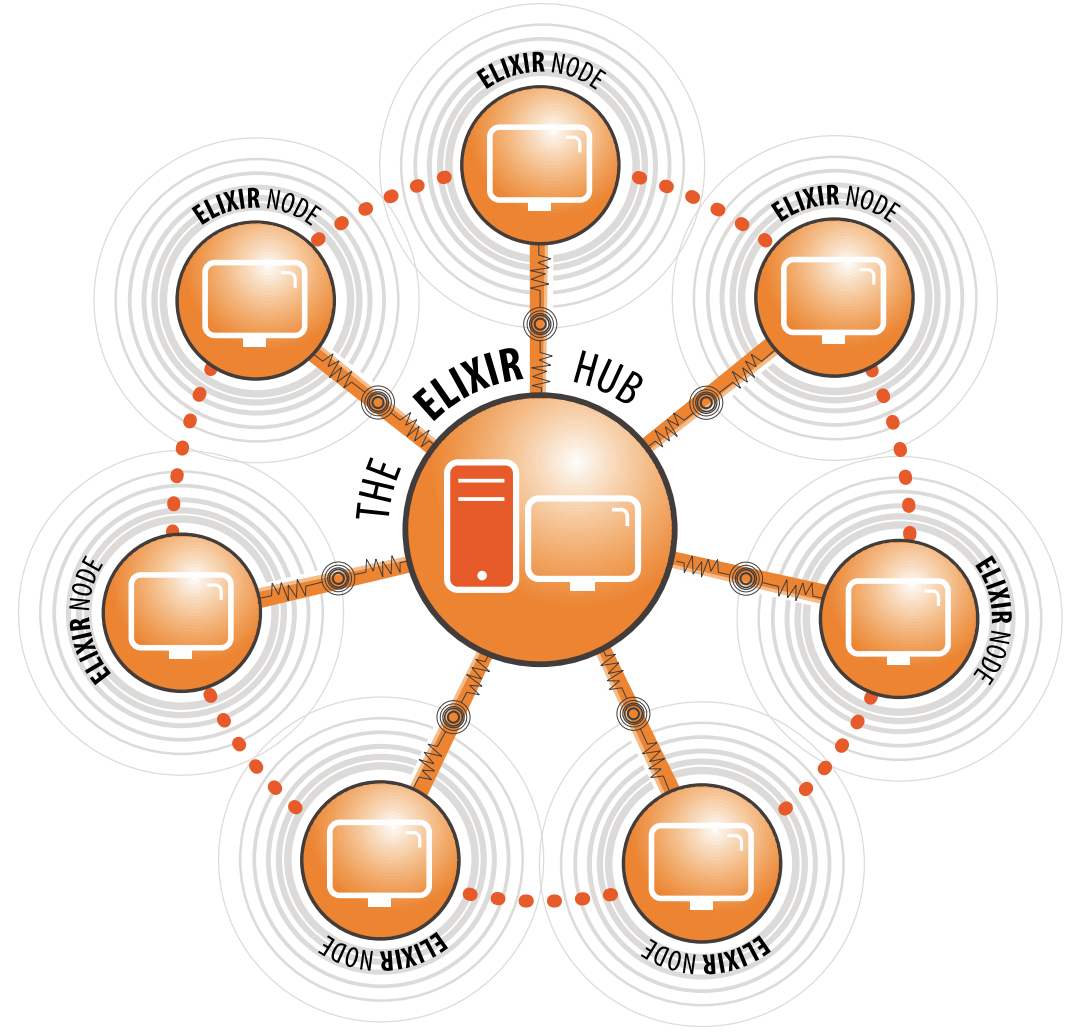


A distributed infrastructure to scale with the challenge

ELIXIR data infrastructure for Europe's life science research sector

ELIXIR Nodes build local bioinformatics capacity throughout Europe

ELIXIR Nodes build on national strengths and priorities

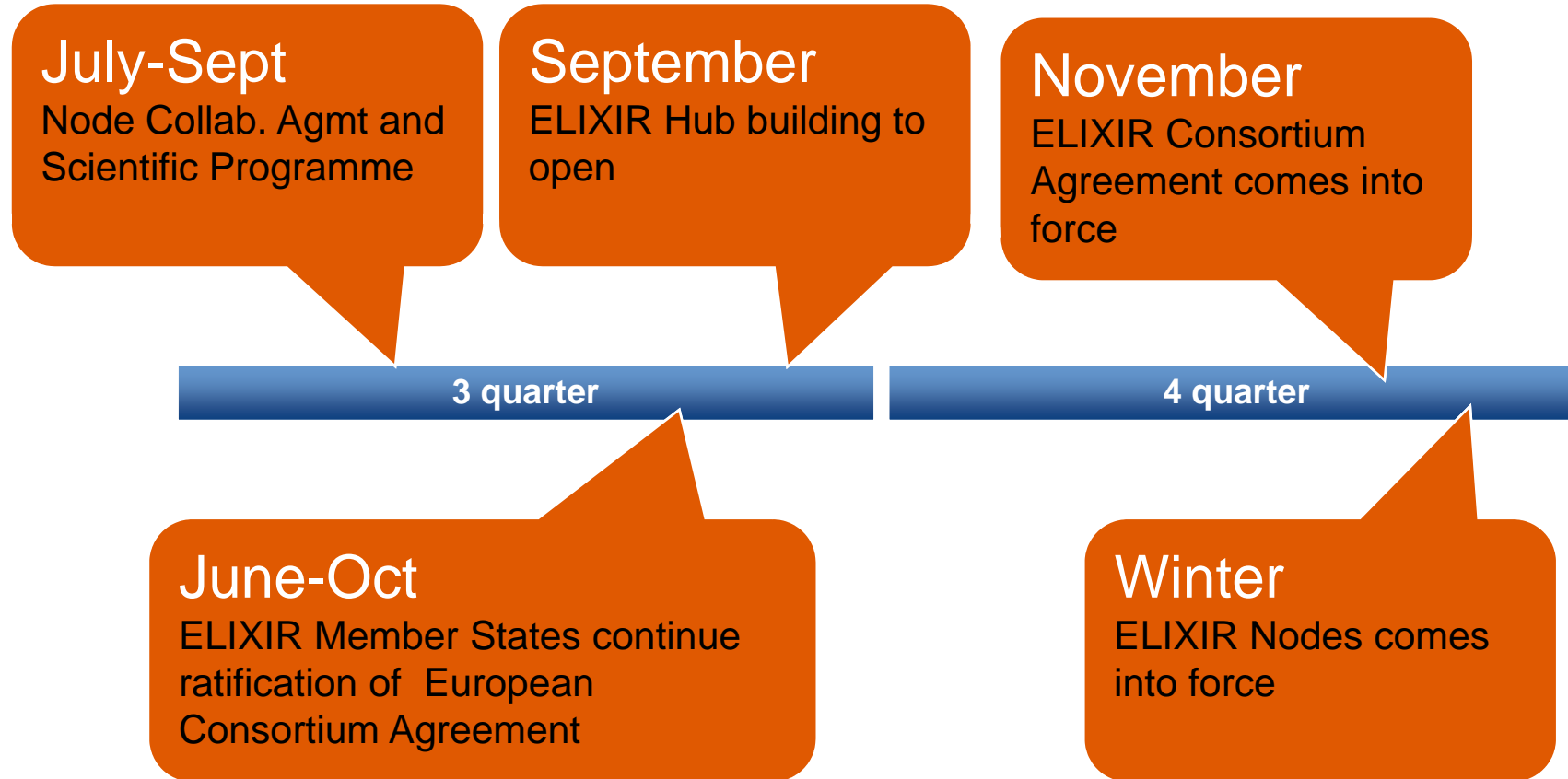


Fifteen countries have signed up

- 15 countries plus EMBL have now signed the Memorandum of Understanding (MoU) to participate
- More are expected to follow in the coming months...
- Countries will now work towards signing an ELIXIR Consortium Agreement (ECA)



ELIXIR Operational Timeline in 2013 (Q3-4)



Horizon 2020



How will ELIXIR Hub & ELIXIR Nodes work together?



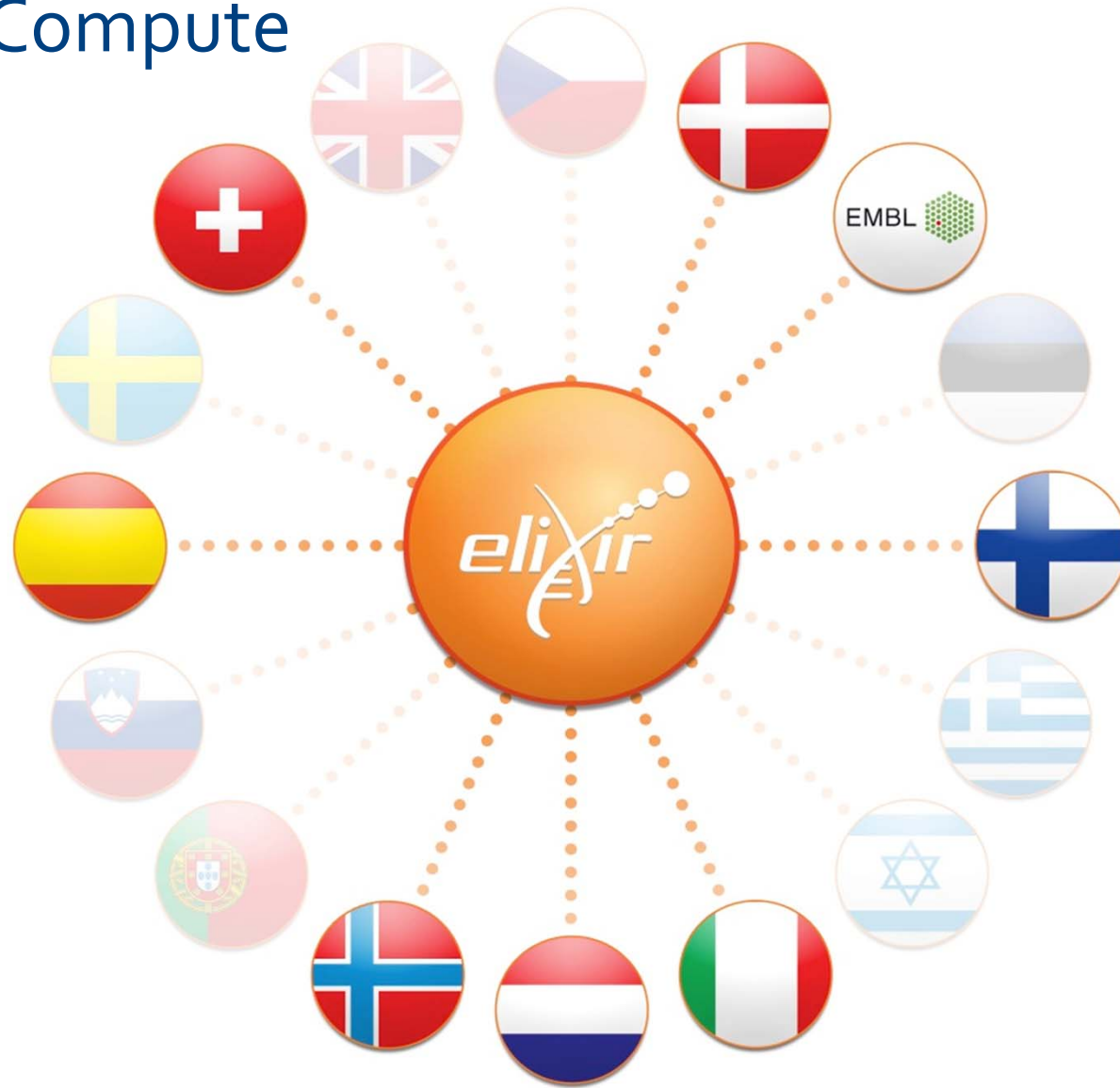
ELIXIR Node

- Research & Develop Bioinformatics services
- Deliver Services through own "brands"
- Management of core resources
- Develop & Deliver training activities
- Participation in international data consortia
- Industry Collaboration and support

ELIXIR Hub

- ELIXIR Mgmt & Secretariat
- Technical coordination across nodes
- Drive standards development & implementation
- Policy and Outreach
- Lead coordinated infrastructure investments
- Deliver services







Compute



- Czech Republic
- Denmark
- Estonia
- Finland
- Greece
- Israel
- Italy
- Netherlands
- Norway
- Portugal
- Slovenia
- Spain
- Sweden
- Switzerland
- United Kingdom


Tools



-  Czech Republic
-  Denmark
-  Estonia
-  Finland
-  Greece
-  Israel
-  Italy
-  Netherlands
-  Norway
-  Portugal
-  Slovenia
-  Spain
-  Sweden
-  Switzerland
-  United Kingdom

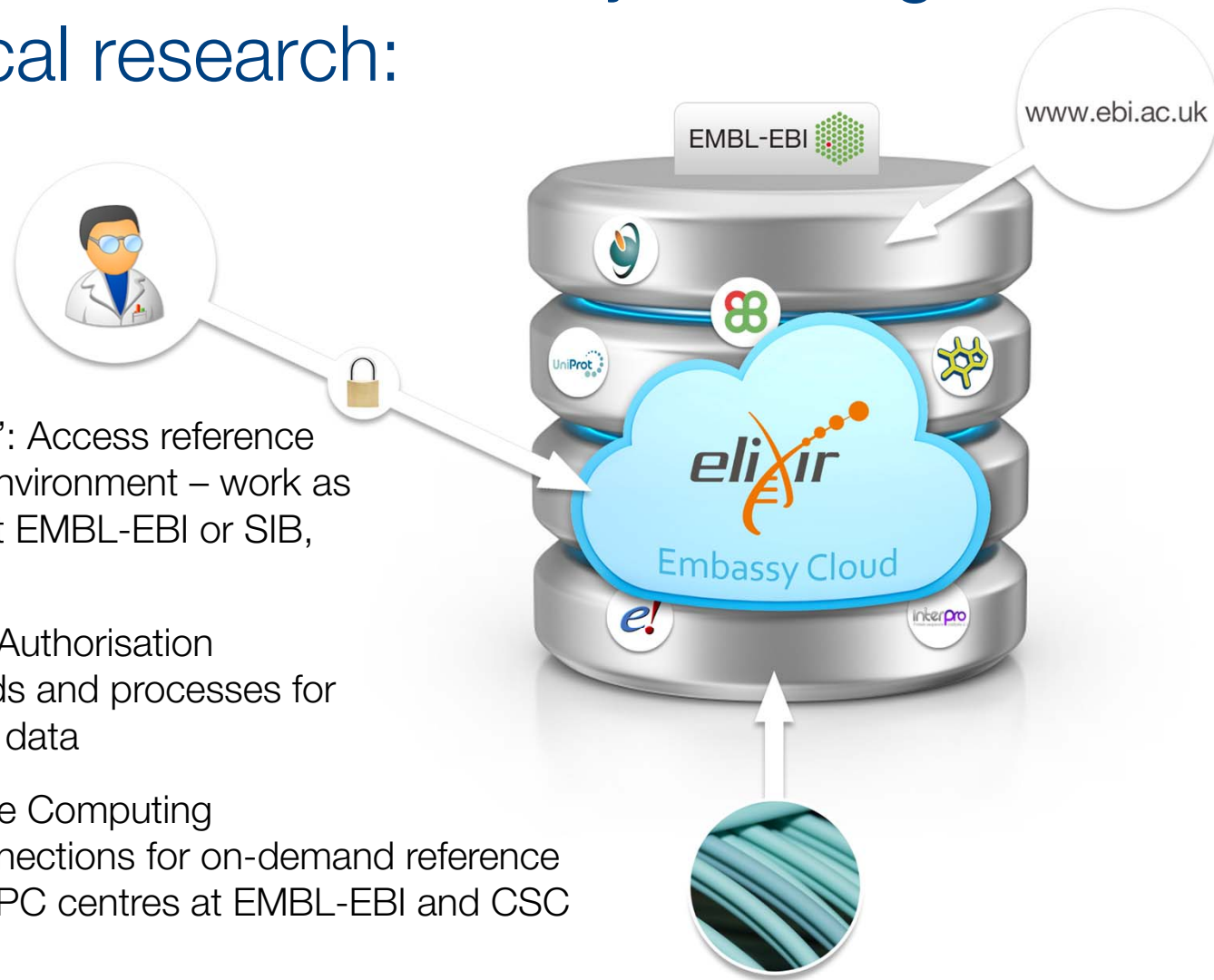
Standards



-  Czech Republic
-  Denmark
-  Estonia
-  Finland
-  Greece
-  Israel
-  Italy
-  Netherlands
-  Norway
-  Portugal
-  Slovenia
-  Spain
-  Sweden
-  Switzerland
-  United Kingdom

ELIXIR pilots to address key challenges in biomedical research:

1. Cloud computing
“Embassy cloud”: Access reference data in a virtual environment – work as though you are at EMBL-EBI or SIB, Switzerland
2. Authentication & Authorisation
Improved methods and processes for access to clinical data
3. High-Performance Computing
“Lightpath”: Connections for on-demand reference data to remote HPC centres at EMBL-EBI and CSC Finland



Thank you



ELIXIR unites Europe's leading life science organisations in safeguarding the biological data generated every day in publicly funded research. Learn more at www.elixir-europe.org



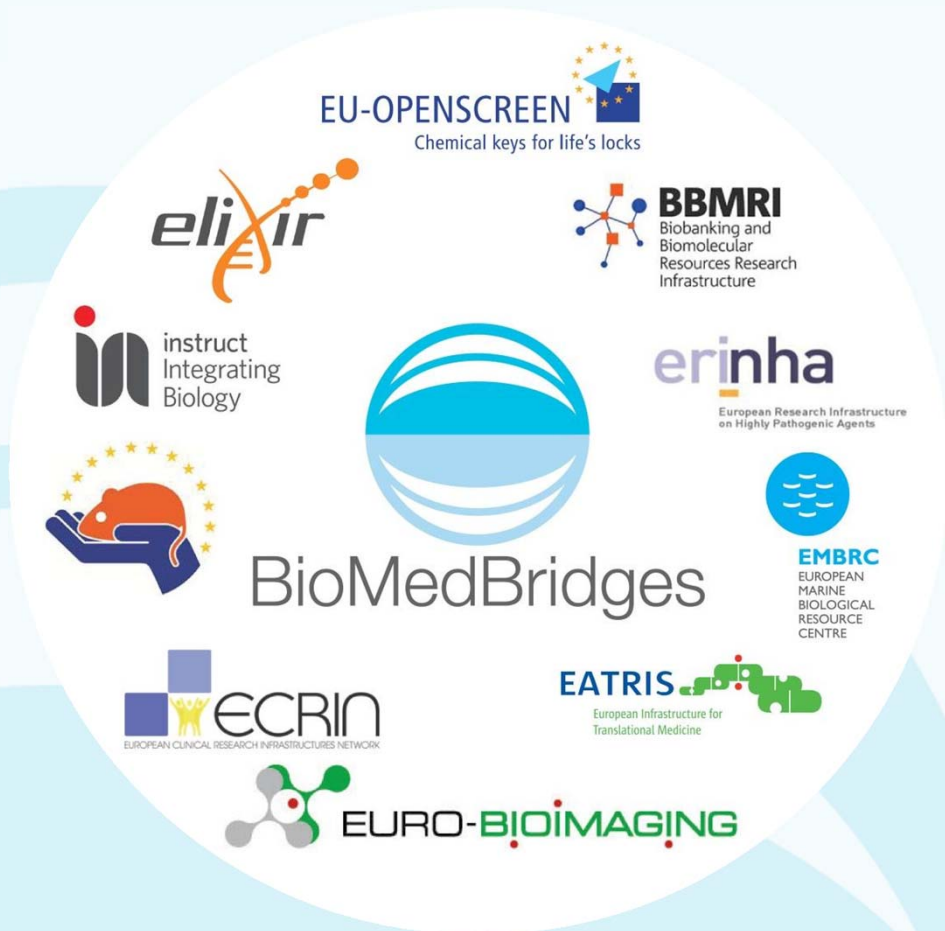
BioMedBridges

Providing data and service bridges
between the biomedical sciences
research infrastructures

www.biomedbridges.eu

Ten new biomedical sciences research infrastructures: stronger through common links

- Computational 'data and service' bridges between the BMS RIs
- Interoperability between data and services in the biological, medical, translational and clinical domains
- Link basic biological research data with clinical research and associated data

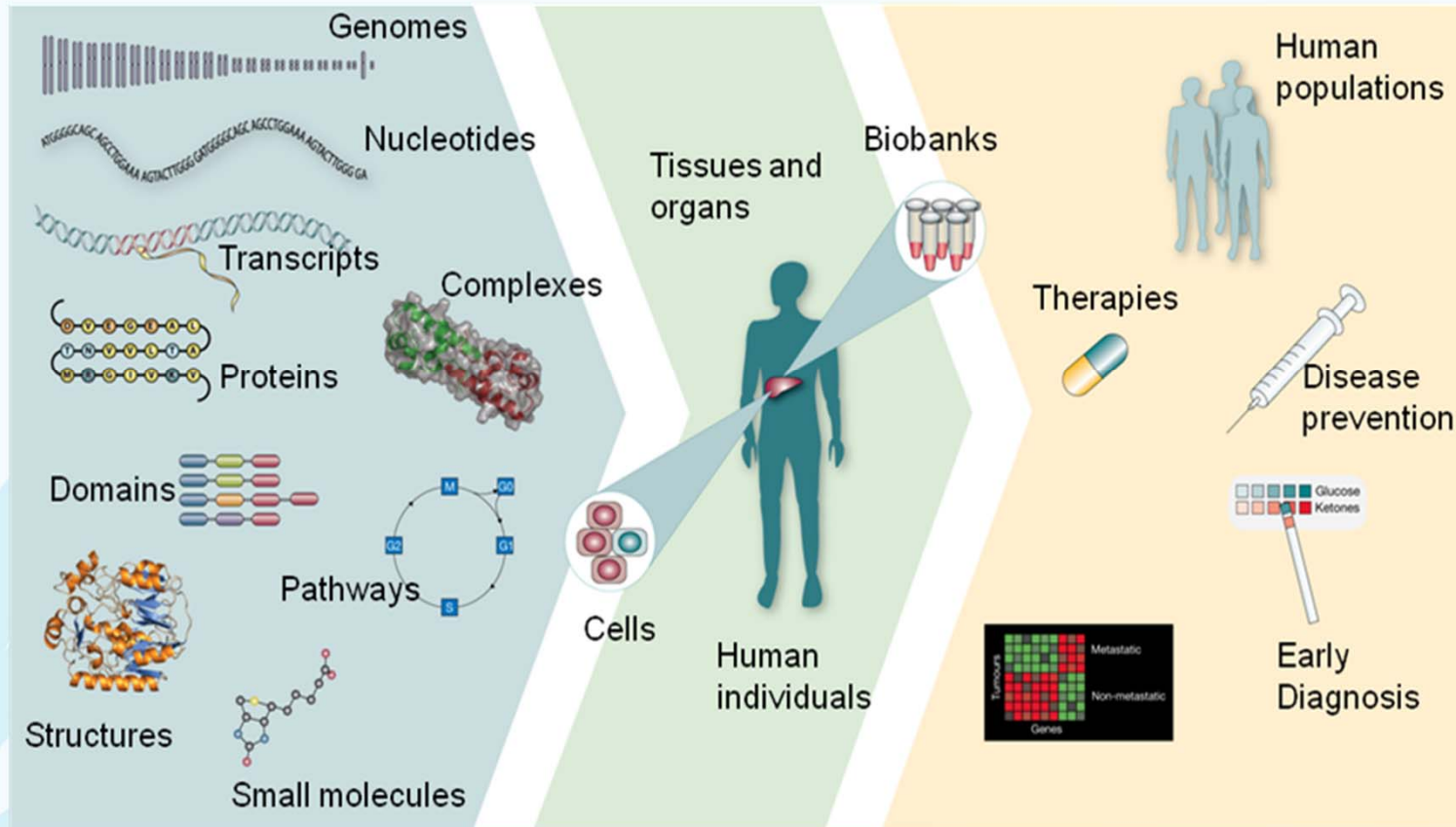


From Molecules to Medicine...

Molecular components

Integration

Translation



Three construction work packages to build the technical bridges

- “Secure Access” and “Standards” will facilitate the effective sharing of data
- “Technical integration” will construct the e-infrastructure

Standards description & harmonisation

Technical integration

Secure access

Five inter-related use cases

- To test the technical and data developments of the construction work packages
- Each use case is coordinated by a particular ESFRI Research Infrastructure

Interoperability of large scale image data sets



Improving the link between mouse models and human data



Personalized Medicine



Integrating structural data



Integrating disease-related data and terminology





BioMedBridges will deliver:



○ Data bridges

- Creating links between available data that were not linked before will hugely increase the potential for new discoveries

○ Interoperability bridges

- standards, formats, ontologies... and how to make it linkable!

○ Social bridges

- Connecting the biomedical research infrastructures



Image courtesy of MDOT's Photography

BioMedBridges Second Annual General Meeting

*10-12 March 2014
Florence, Italy*



- Day 1 – Frontier Science presentation from each Infrastructure
 - Interactive workshops, and poster session with reception.
- Day 2 – Open meeting to present the scientific achievements of the BioMedBridges project
- Day 3 – Closed meeting for project mid-term review

SIB: Swiss Node of ELIXIR



Swiss Institute of
Bioinformatics

- Established in 1998
- 46 research and service groups, more than 600 scientists
- in Basel, Bern, Fribourg, Geneva, Lausanne, Lugano and Zurich



- More than 130 bioinformatics resources for life science researchers worldwide
- All resources available on: <http://expasy.org>



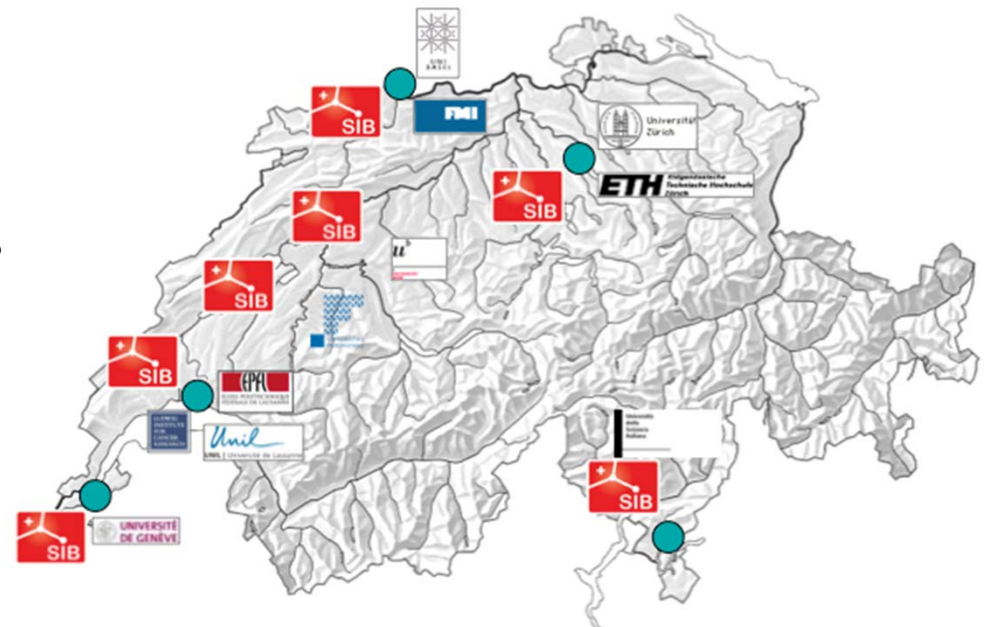
Vital-IT

High Performance
Computing Center

• Compute and competence centres

BC2

- Training facility, courses and workshops:
www.isb-sib.ch/education



Data



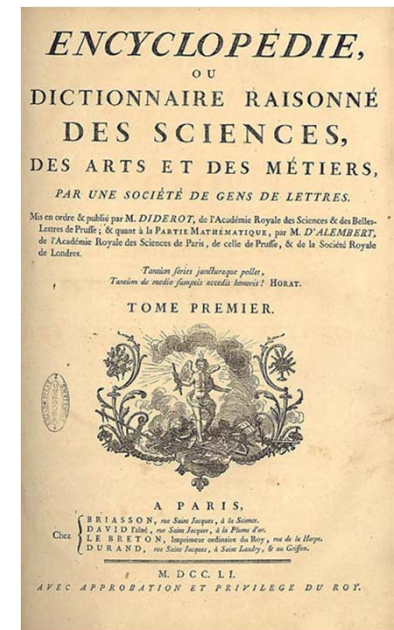
Knowledge

What is biocuration?

- Biological **knowledge** from the **literature** is **encoded** in **knowledgebases** using standard **identifiers**, **ontologies**, and **controlled vocabularies**.
- Biocuration of literature **recovers existing knowledge** that is **'lost'** in papers and **not evaluated within a global context**
 - => powerful queries within and across knowledgebases
 - => linking and exploration of existing data



What Swiss-Prot biocurators
Do is what
The monk did long time ago



What does a biocurator do?

Read

And read

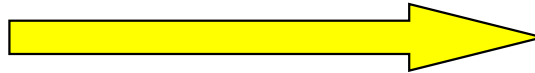
And read

And read

And read

And read

And synthesize !




They read publications and extract the **KNOWLEDGE** that will help explain the function of a protein.

The screenshot shows the UniProtKB entry for Cytochrome c (CYC). The search bar at the top contains 'Query'. The entry is identified as 'Reviewed, UniProtKB/Swiss-Prot P99999 (CYC_HUMAN)' and was last modified on March 18, 2006. The 'Names and origin' section lists the protein name as 'Cytochrome c', gene names as 'CYC', and organism as 'Homo sapiens (Human)'. The 'General annotation (Comments)' section describes the function of the protein as an electron carrier in the mitochondrial electron-transport chain and its role in apoptosis.

From January 2012 to January 2013

More than 9 000 articles have been curated to UniProtKB/Swiss-Prot by our biocurators.

Swiss-Prot biocuration efforts



Protein
sequence and
functional
annotation



Pathway and reaction
annotation



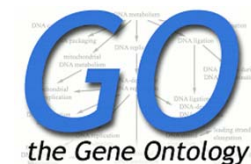
Protein interaction
annotation



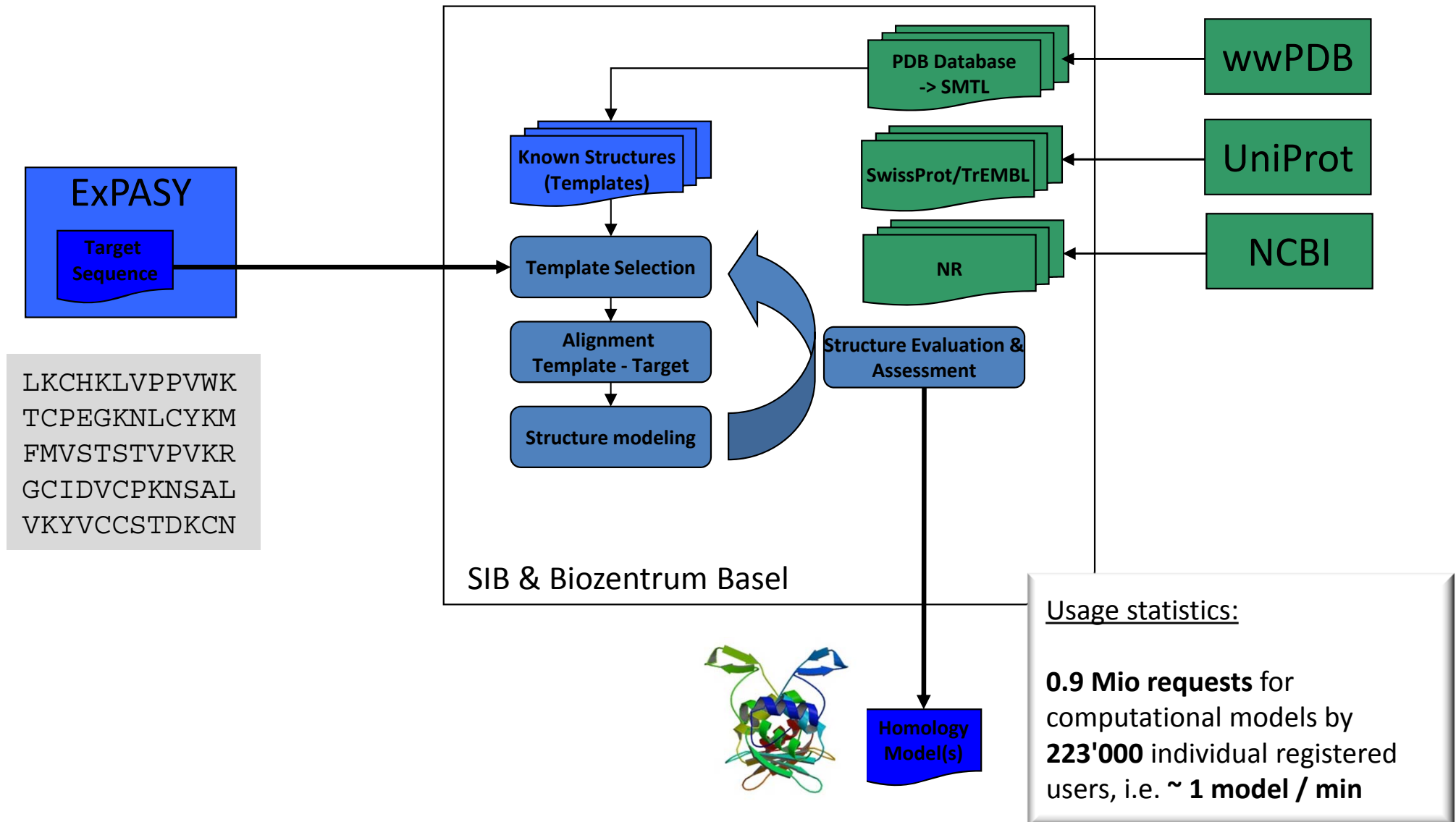
Evidence-based proteomics
annotation



Gene Ontology
annotation



SWISS-MODEL Workspace.



<http://swissmodel.expasy.org>

SWISS-MODEL Repository

SWISS-MODEL Repository - Model Details

Model Overview [+/-]

Click on the bars to get more details about individual [Models](#) or experimental [structures](#)

1

367

Sequence [+/-]

UniProt P0A9S5 Glycerol dehydrogenase
Escherichia coli (strain K12).
Database: Swiss-Prot (Reviewed) ★

STRING

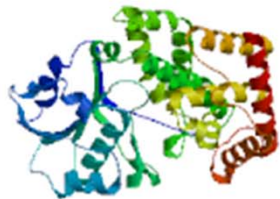


P0A9S5: gldA

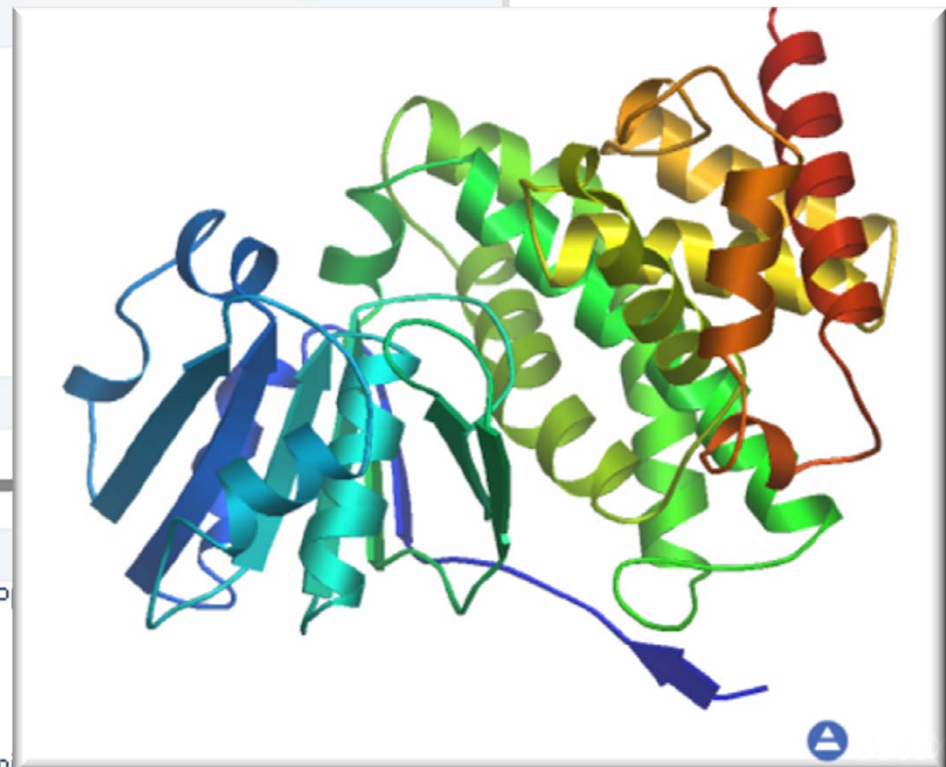
Domain [+/-]

Link to: [[InterPro](#)]

Model 3D Structure [+/-]



Based on template: [1kq3](#) [[SMTL](#)] [[P](#)
Sequence identity: 52%
Residue range: 1 to 364
Model date: 2010-01-07
Revision date: 2010-02-04
[[display](#)] [[download](#)] [[download pro](#)



Alignment [+/-]



[Search](#)
[Blast *](#)
[Align](#)
[Retrieve](#)
[ID Mapping *](#)

Search in: Protein Knowledgebase (UniProtKB)

P0A9S5 (GLDA_ECOLI) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified June 28, 2011. Version 55. [History...](#)

[Contribute](#)
[Send feedback](#)
[Read comments \(0\) or add your own](#)

[Clusters with 100%, 90%, 50% identity](#) |
 [Documents \(3\)](#) |
 [Third-party data](#)

[text](#)
[xml](#)
[rdf/xml](#)
[gff](#)
[fasta](#)

[Names](#) ·
 [Attributes](#) ·
 [General annotation](#) ·
 [Cross-refs](#) ·
 [Entry info](#) ·
 [Documents](#) ·
 [Customize](#)

[Domain \[+/-\]](#)
 Link to: [\[InterPro \]](#)

Names and origin

Protein names	<i>Recommended</i> Glycerol dehydrogenase Short name Short name EC=1.1.1.1
Gene names	

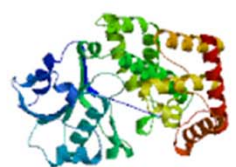
Organism	Escherichia coli
Taxonomic identifier	83333 [NCBI]
Taxonomic lineage	Bacteria > Proteobacteria > Enterobacteriales > Enterobacteriaceae > Escherichia > Escherichia coli

Protein attributes

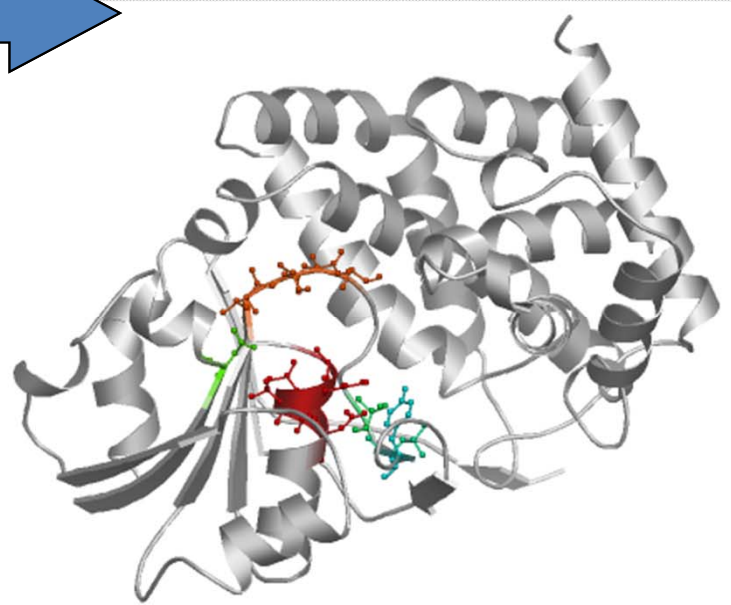
Sequence length	367 AA.
Sequence status	Complete.
Protein existence	Evidence at protein level

General annotation (Comments)

Model 3D Structure [+/-]



Based on template: [1kq3](#) [SMTL] [RCSB] [PDB] [SCOP] [CATH]
 Sequence identity: 52%
 Residue range: 1 to 364
 Model date: 2010-03-19
 Revision date: 2011-06-06
[\[display \]](#) [\[download \]](#) [\[download project \]](#)



Uniprot features:

<input checked="" type="checkbox"/>	Nucleotide binding	94 - 98	NAD (By similarity).
<input checked="" type="checkbox"/>	Nucleotide binding	116 - 119	NAD (By similarity).
<input type="checkbox"/>	Metal binding	171	Zinc; catalytic (By similarity).
<input type="checkbox"/>	Metal binding	254	Zinc; catalytic (By similarity).
<input type="checkbox"/>	Metal binding	271	Zinc; catalytic (By similarity).
<input checked="" type="checkbox"/>	Binding site	37	NAD (By similarity).
<input type="checkbox"/>	Binding site	121	Substrate (By similarity).
<input checked="" type="checkbox"/>	Binding site	125	NAD (By similarity).
<input checked="" type="checkbox"/>	Binding site	127	NAD; via carbonyl oxygen (By similarity).
<input checked="" type="checkbox"/>	Binding site	131	NAD (By similarity).
<input type="checkbox"/>	Binding site	171	Substrate (By similarity).
<input type="checkbox"/>	Binding site	254	Substrate (By similarity).
<input type="checkbox"/>	Binding site	271	Substrate (By similarity).

Conclusions

- Switzerland has committed funding to an ELIXIR-like structure through the SIB for the last 15 years
- Switzerland has funded services such as :
UniProtKB/Swiss-Prot, Swiss-Model,
STRING, NeXtProt
- The work has to be coordinated when a resource is becoming essential to more than 10k/months users, currently UniProt is used by more than 500k users/months worldwide....

A few thoughts

If you wanted to know whether your service is used...pull off the internet and check how much time is needed to have A user coming back to you by email or phone



ELIXIR

Special Session at ECCB

Summary of Key Topics

July 2013, Berlin

Key actions Dutch Node (BM)



European Life Sciences Infrastructure for Biological Information

www.elixir-europe.org

An Outstanding Group!



Overview of topics (in random order)

1. Detailing of user needs
2. Development of Elixir metrics



4. Funding approach



6. Data access & security approach

7. Certification and kite marking

8. Continued development of Elixir Infrastructures



tools,



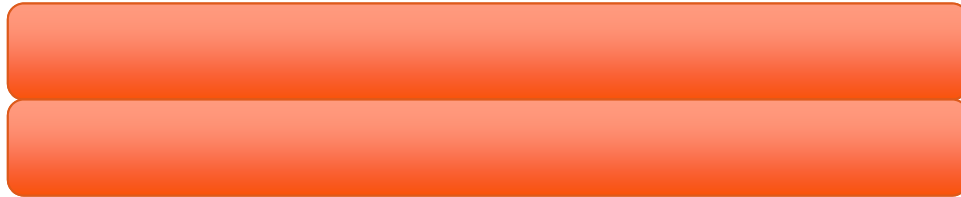
Follow up topics (in DTL-DISC context)

Elixir best practice

- Set up WG to determine non scalable best practices
- set up training facility (with GOBLET) to spread these in nodes.



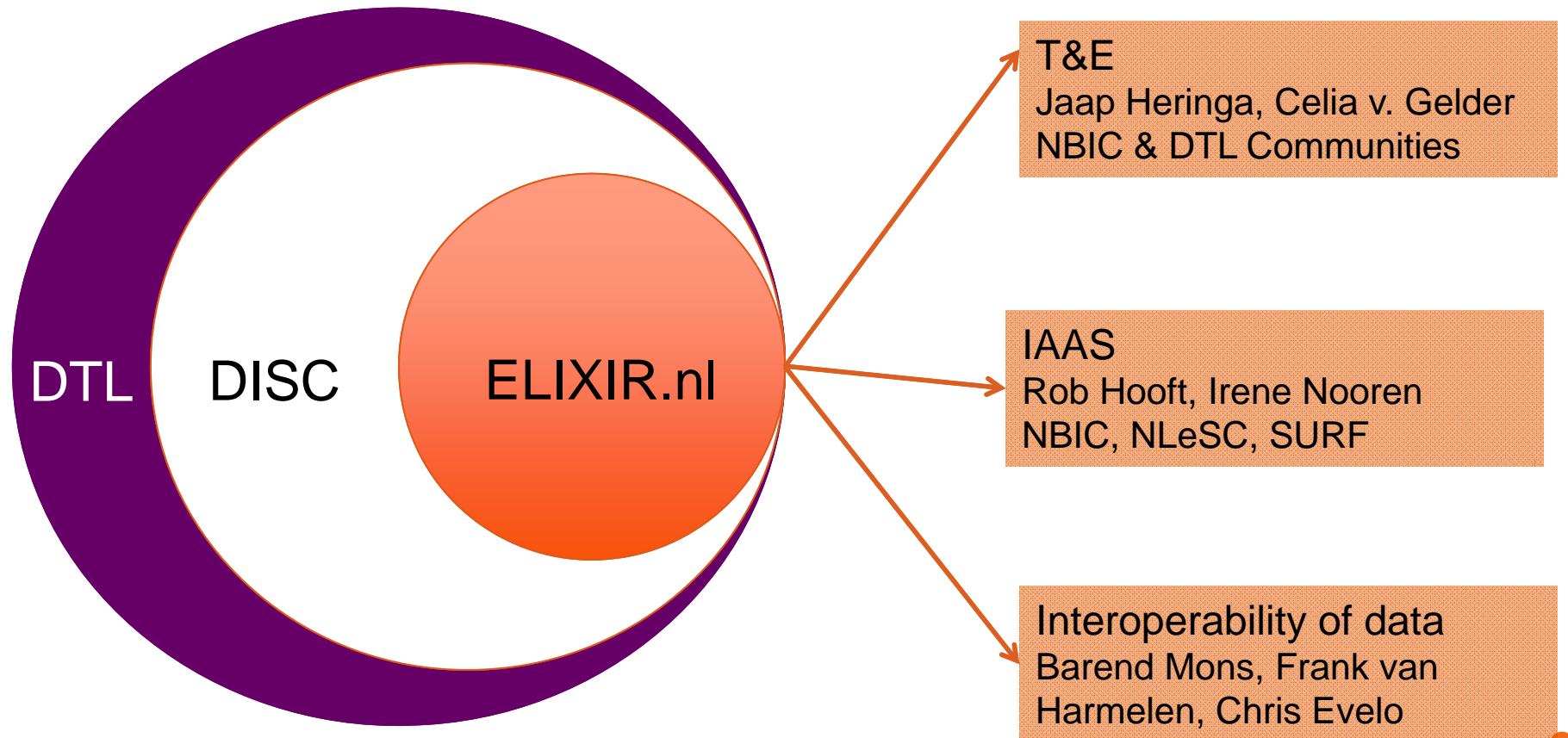
Overview of topics (in random order)



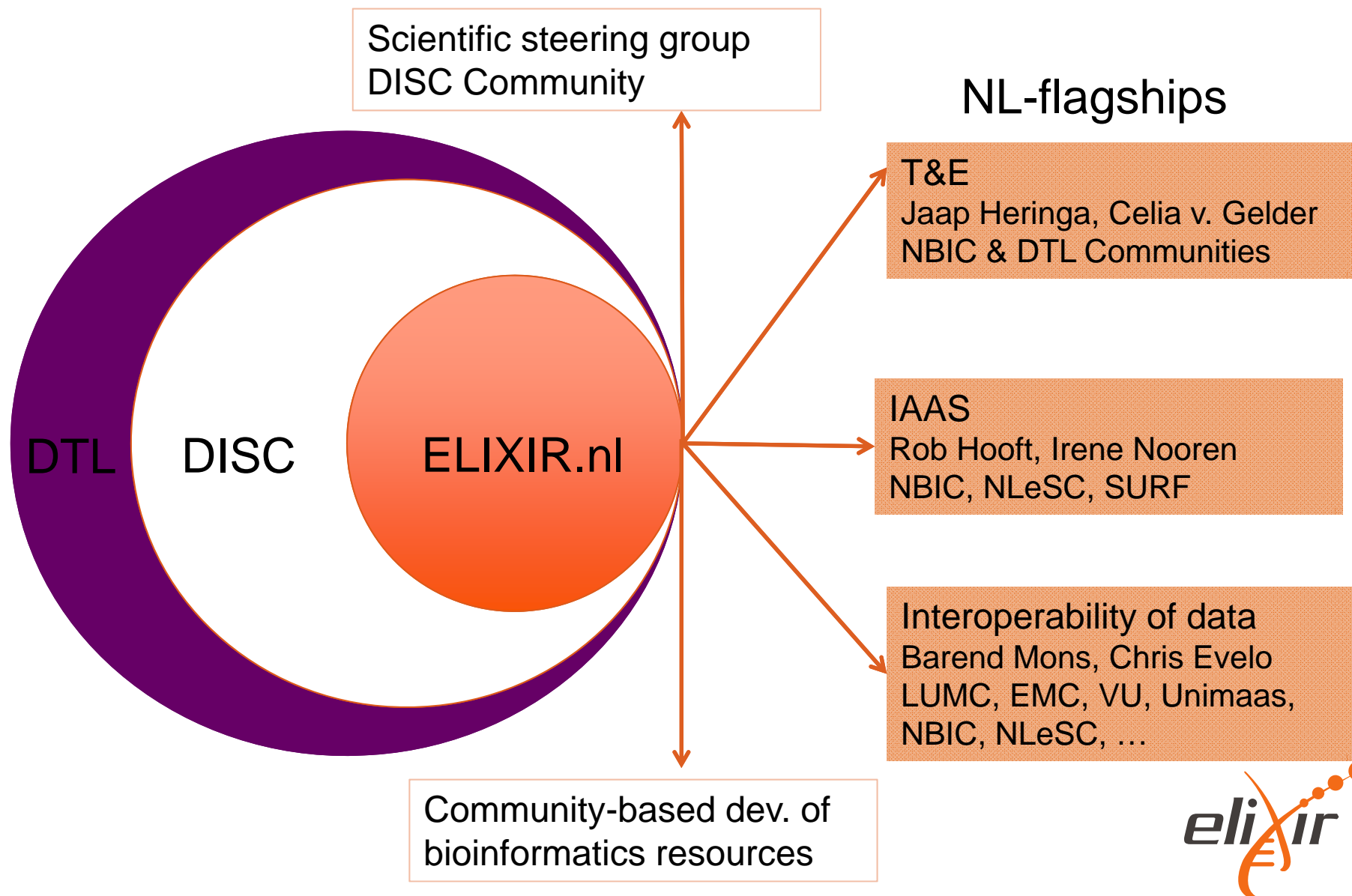
- Offer to participate in WG to determine actions with ESFRI meeting in Brussels (July 2013, Niklas invited)
- Offer to participate in WG Industry alignment.

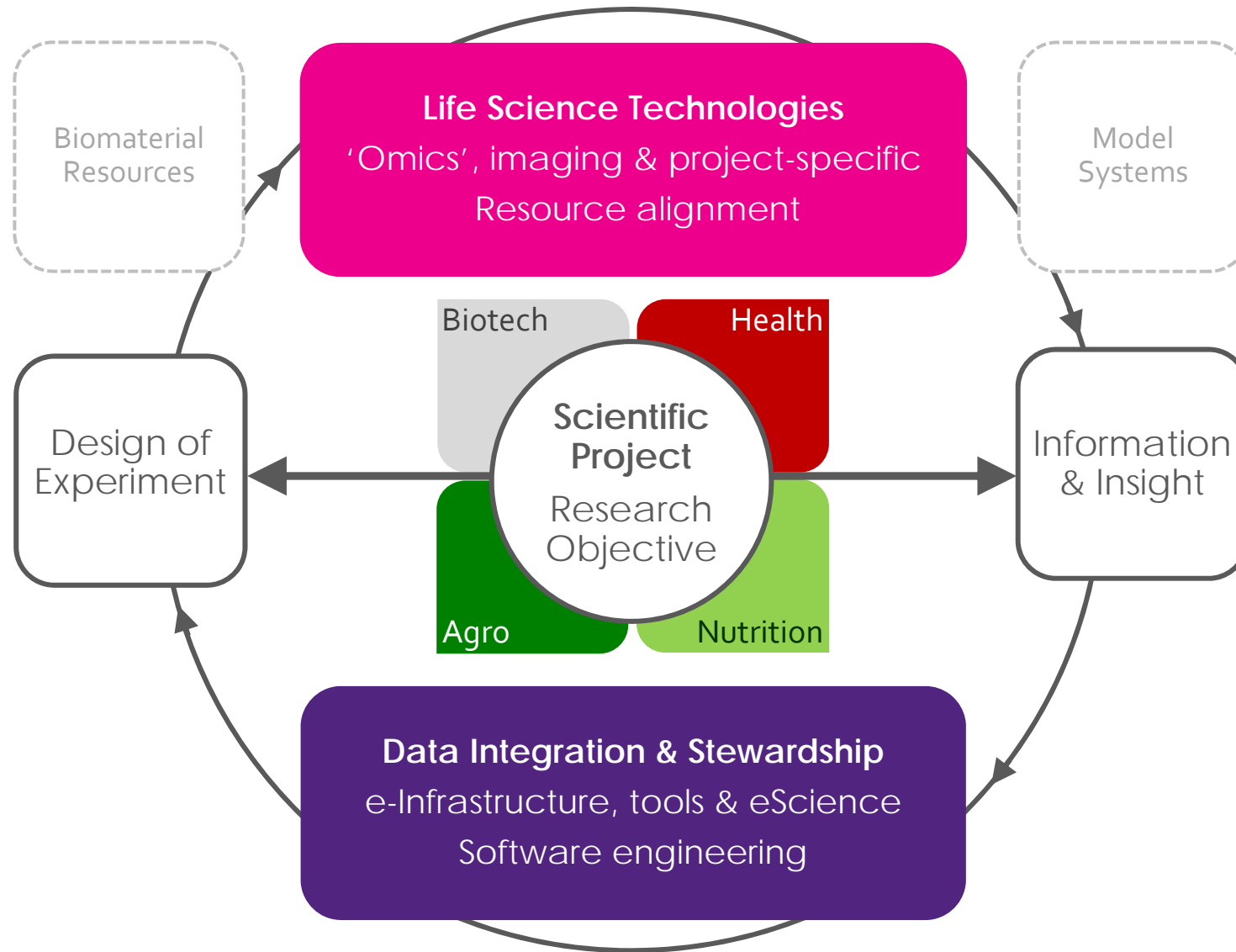


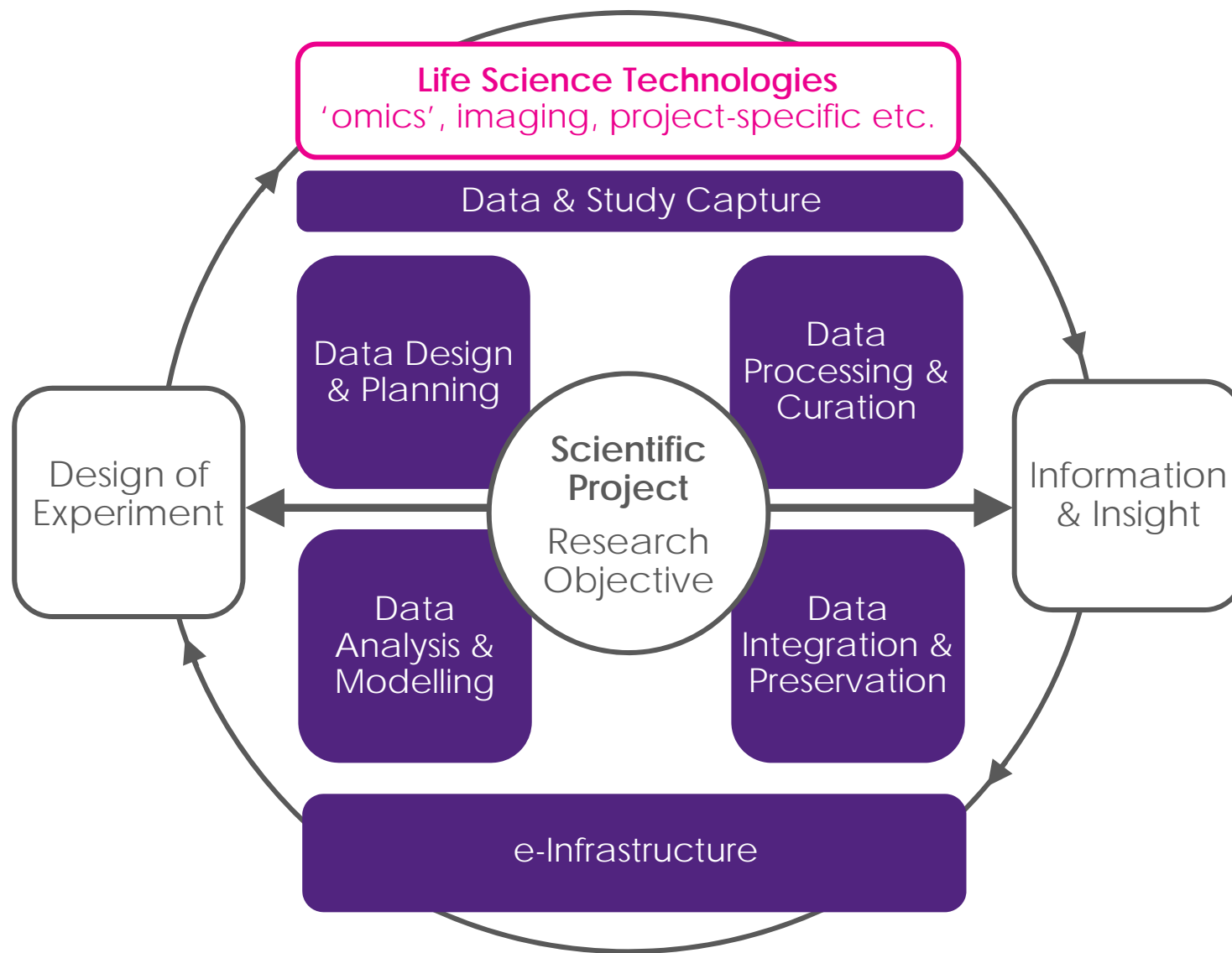
DTL-scope and ELIXIR.nl

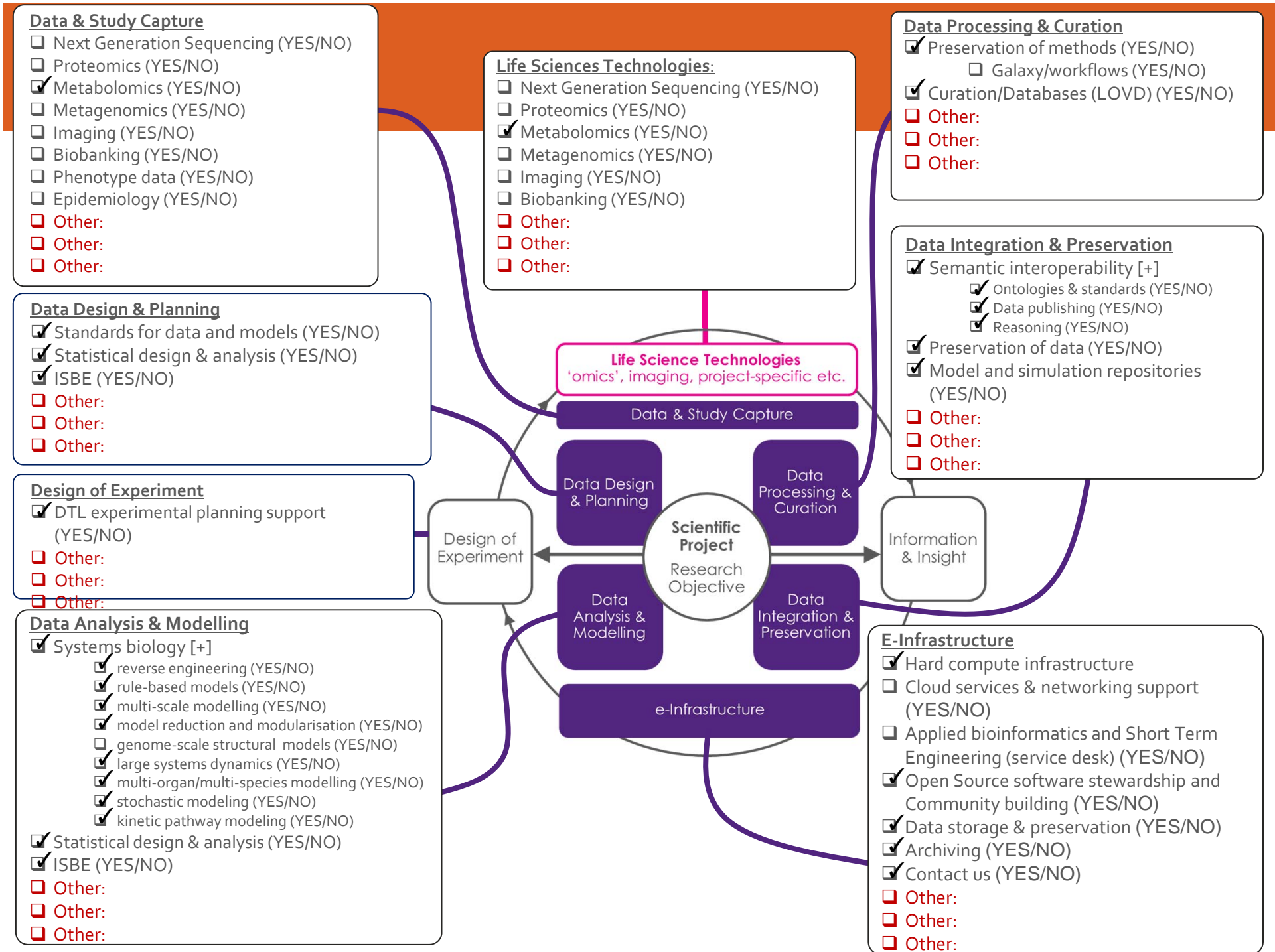


DTL-scope and ELIXIR.nl







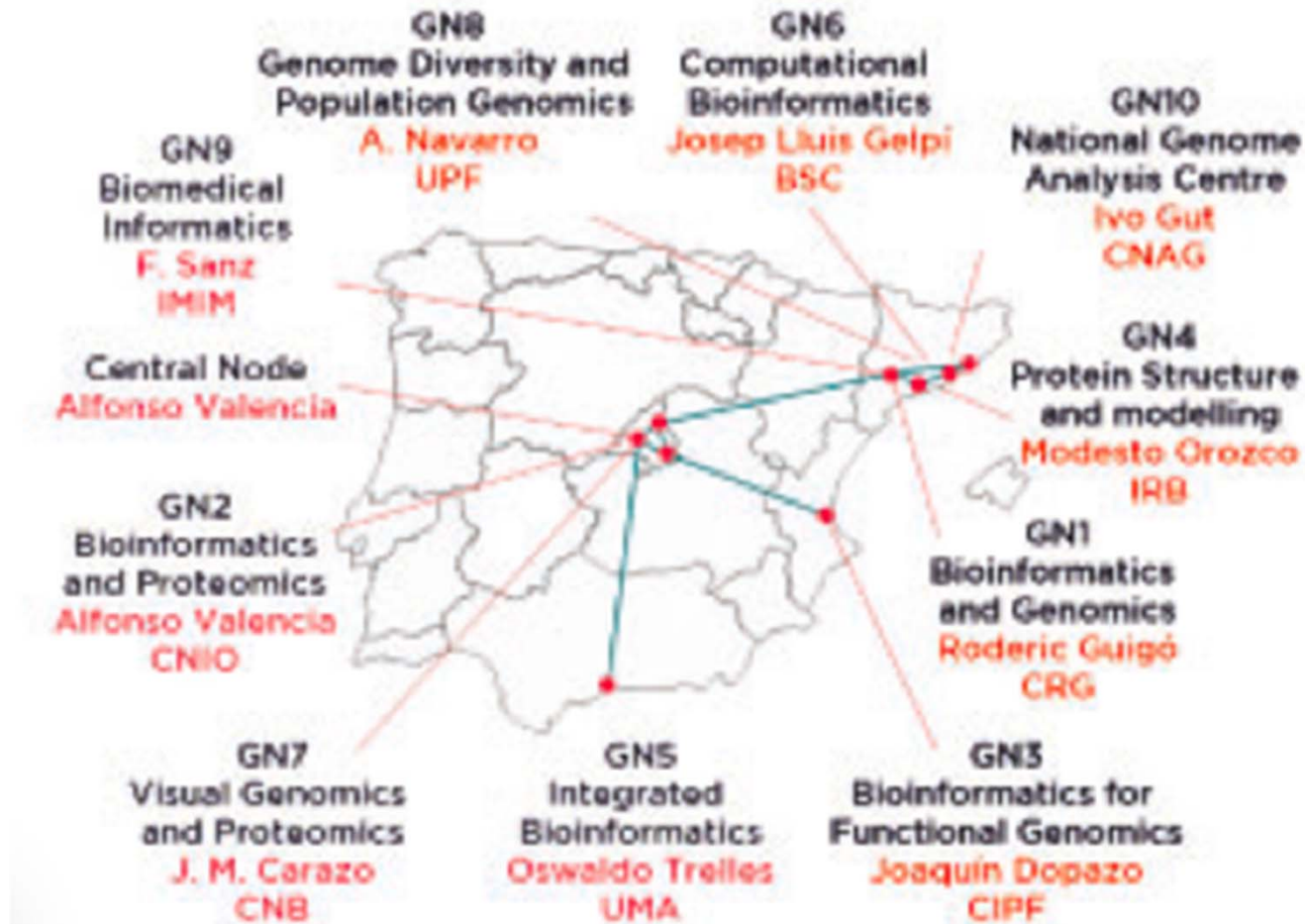


The Spanish National Bioinformatics Institute (INB) as **ELIXIR node**

Alfonso Valencia
Director INB

The Spanish National Bioinformatics Institute (INB)

2003-2013 / 2014-2018



10 nodes + central hub

**Technical Platform of the Spanish Health Institute
(ISCIII)**

Mission

“To generate and apply bioinformatics solutions to needs detected in development and implementation of genomics and proteomics focused projects”

- **To support Bioinformatics and Computational Biology development in Spain**
- To collaborate and provide scientific and technical support to national genomics and proteomics projects
- To contribute to the creation and establishment of local Bioinformatics groups with research and services components through bioinformaticians training
- To train bioinformaticians for genomics and proteomics research groups
- To develop pure Bioinformatics projects related with the Institute activities
- To support companies with activity in this sector in Spain
- **To represent Spain in international projects in the area of Bioinformatics**

about /

↳ The Spanish Institute of Bioinformatics ("Instituto Nacional de Bioinformática" or INB) supports significant bioinformatics resources and coordinates the participation of Spanish groups in important large scale projects

resources /

↳ The INB sustains important core bioinformatics resources developed in Spanish groups and keeps them available to the national and international research community

projects /

↳ INB groups participate in national and international projects that advance important areas like medical genomics and the sequencing of key organisms

quick links /

[INB nodes](#)

[Core INB projects](#)

[Support](#)

core INB resources /

[3DEM Benchmarking](#)

[aGEM](#)

[APPRIS](#)

[FlexPortal](#)

[GRAPE](#)

[SNPator](#)

Barcelona Supecomputing Centre (BSC-INB node)



BSC is part of PRACE

3 Engineers interface Supercomputing / Bioinformatics

(1/4 of the MN project in Biology/biomedicine, 2/3 of them using INB facilities / expertise)

ELIXIR as a natural expansion



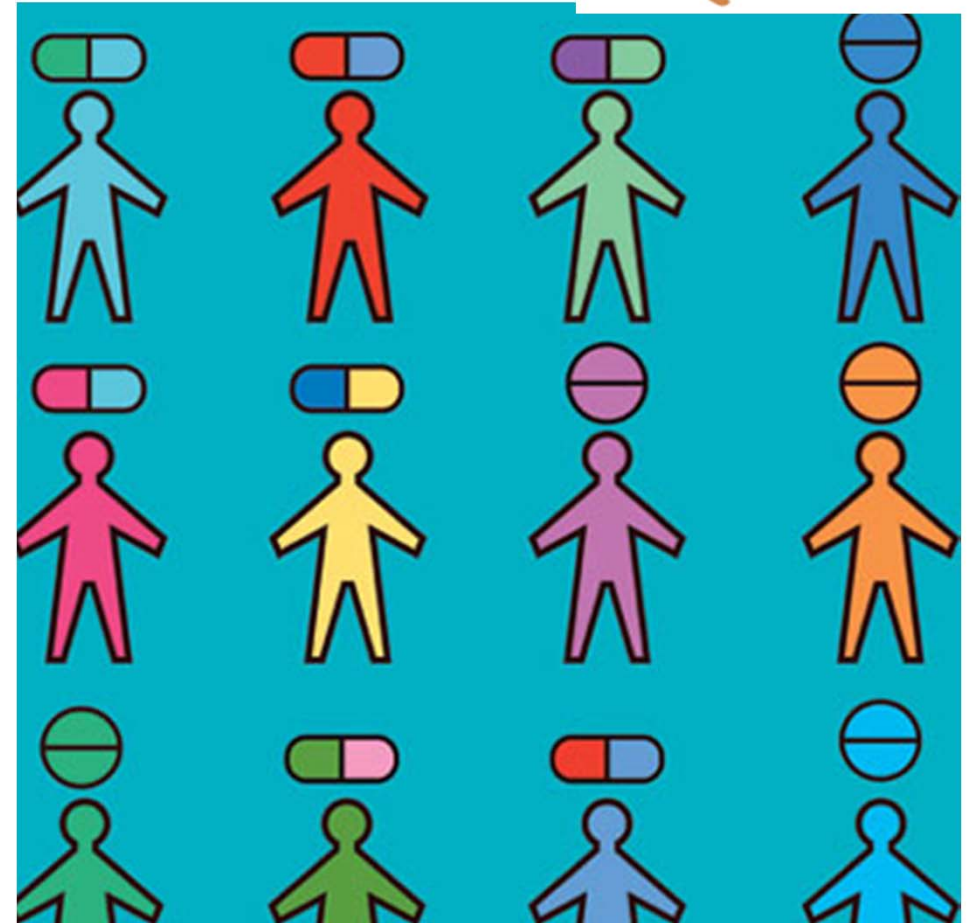
Proposed INB contribution to ELIXIR:

- *Provision of “(web) services” as a basic technology for ELIXIR.*
- *Evaluation of bioinformatics methods, with the related standards and repositories.*
- *“High-Performance Computing” in Biology with participation of the BSC-INB node, including ELIXIR pilot project, connection with PRACE*
- *Services based on literature and information extraction.*
- *Participation in large scale projects including ICGC, ENCODE, Blueprint International Consortium on Rare Diseases, as well as IMI*
- ***Alliance for the development of the European Genotype-Phenotype Archive (EGA) EBI-CRG(INB node) (ELIXIR pilot)***

ELIXIR and Medical information



- The era of personal genome sequencing is upon us.
- Sequence data will not cross national boundaries.
- Every national health system will need **expertise to interpret it** and treat patients accordingly.
- Individuals need to be sure that their personal biological data are in safe hands.





Swedish Elixir node

Integrating the Human Protein Atlas into the Elixir landscape

Bengt Persson



European Life Sciences Infrastructure for Biological Information

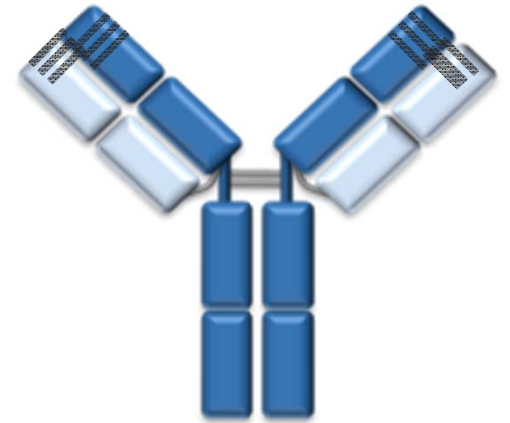
www.elixir-europe.org

- Initially we will contribute with ...
 - Human Protein Atlas project (<http://www.proteinatlas.org>)
 - human subcellular atlas
 - human tissue atlas
 - human cell line atlas
- Additional services to follow later ...
 - Tools
 - Databases

and integration into
other ELIXIR resources

The Human Protein Atlas project

- Started in 2003
- Systematic generation and use of antibodies to functionally explore the human proteome
- 2015 goal – first draft of human proteome
- 2020 goal – premium quality proteome atlas
- Funded by the Wallenberg Foundation (non-profit) and EU



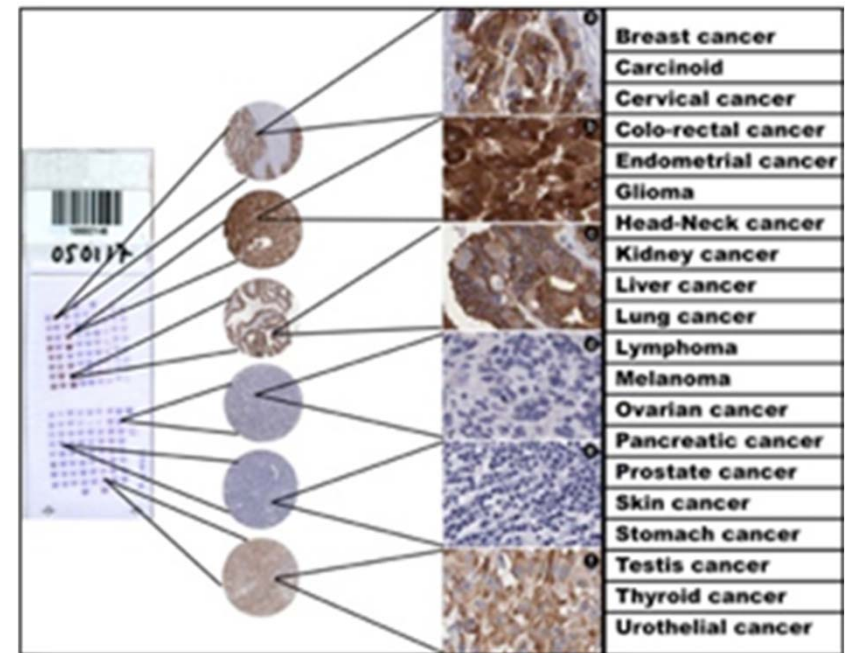
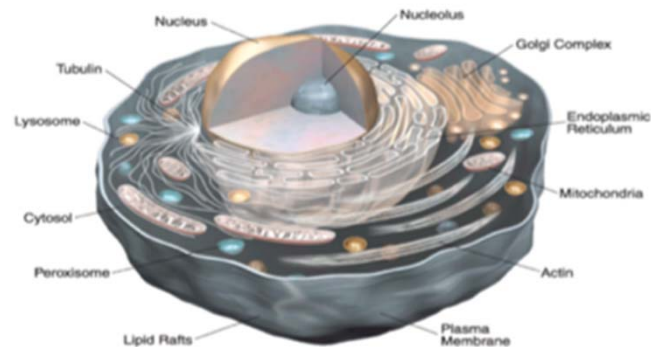
THE HUMAN PROTEIN ATLAS 

*Knut och Alice
Wallenbergs
Stiftelse*



The Human Protein Atlas Data

- Tissues (IHC)
 - 46 normal tissues and organs
 - 20 types of cancer tissues
- Cell lines (IF och IHC)
 - 47 cell lines
 - 12 clinical cell samples



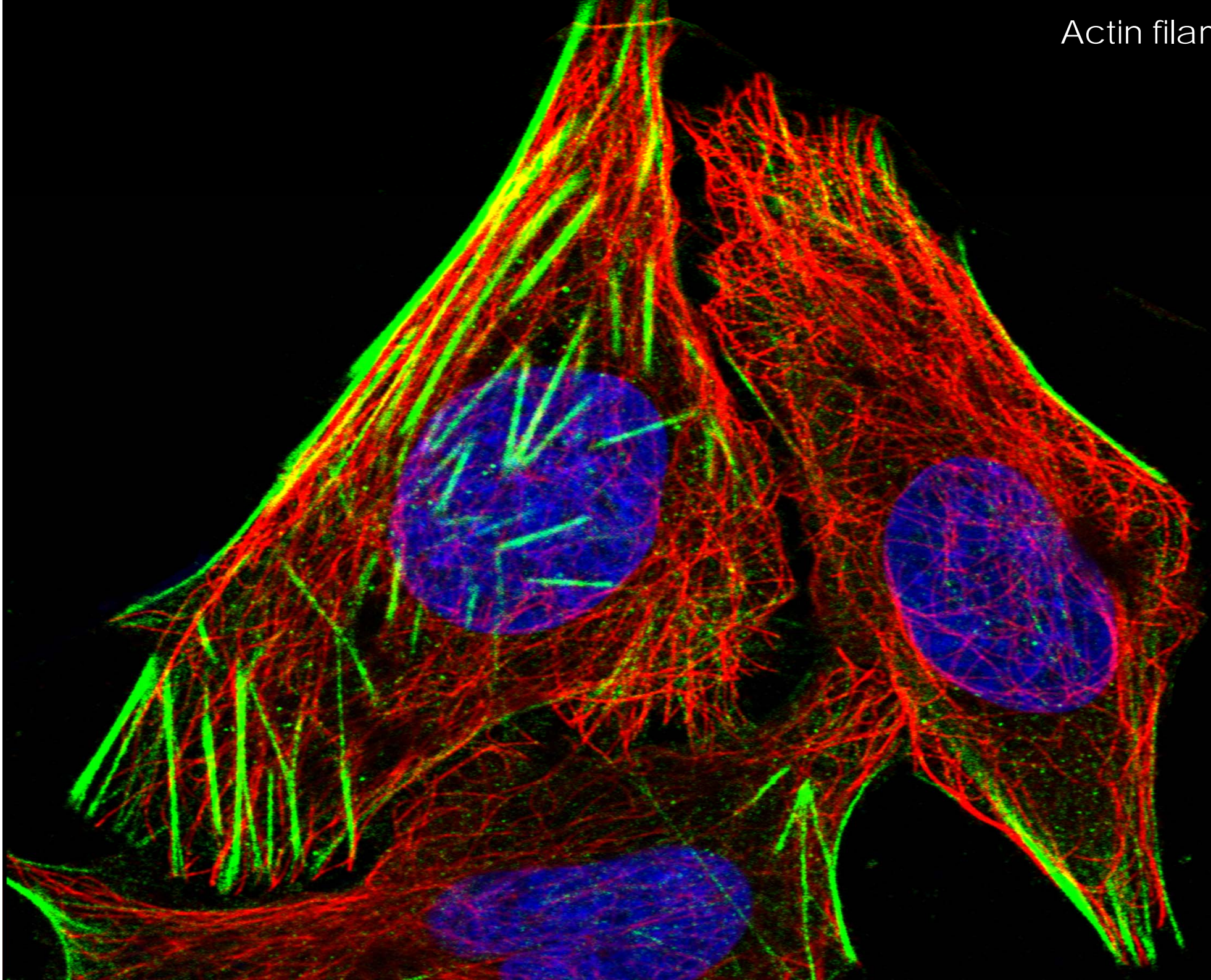
THE HUMAN PROTEIN ATLAS

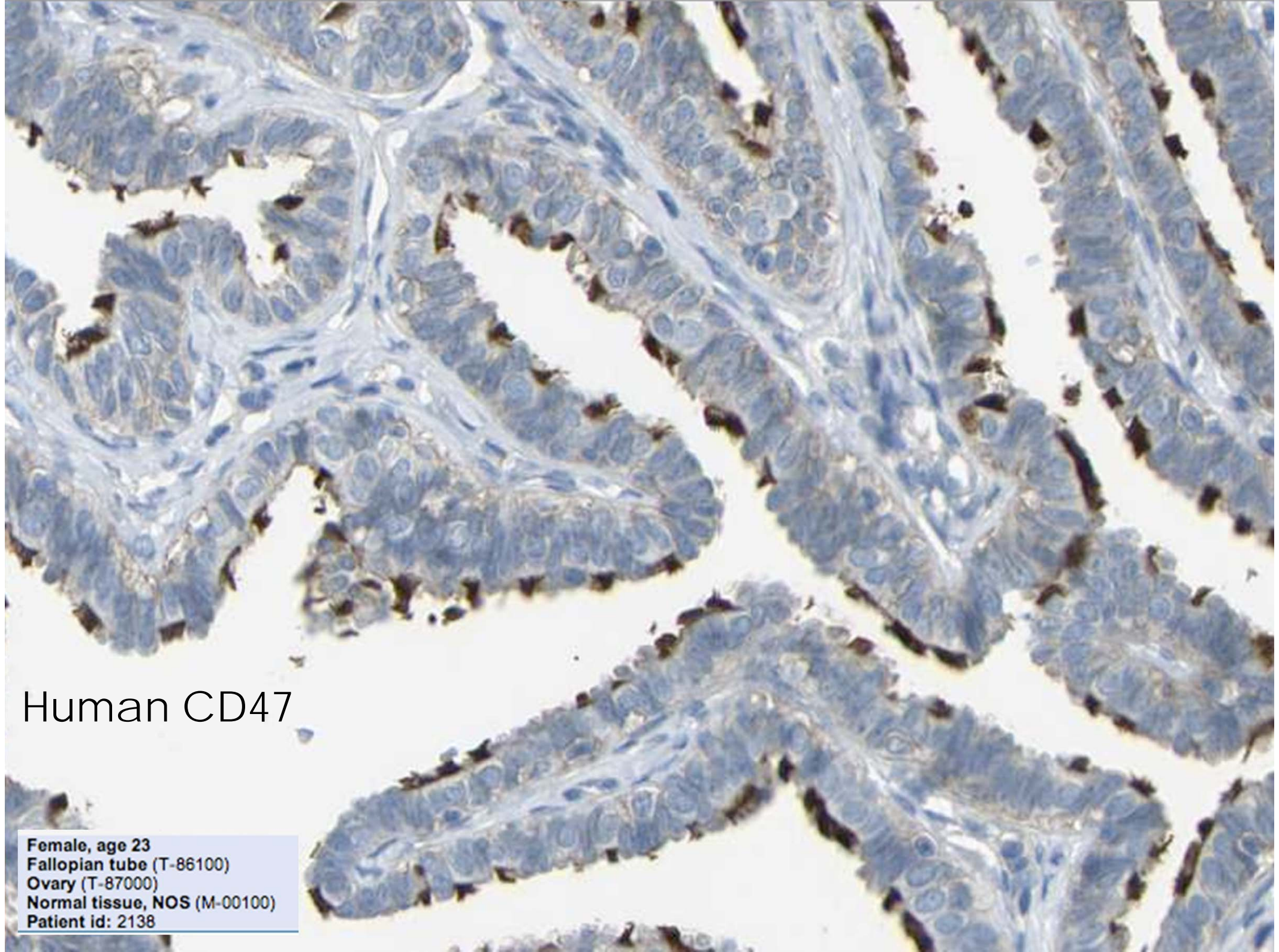
Knut och Alice
Wallenbergs
Stiftelse



elixir

Actin filaments





Human CD47

Female, age 23
Fallopian tube (T-86100)
Ovary (T-87000)
Normal tissue, NOS (M-00100)
Patient id: 2138

THE HUMAN PROTEIN ATLAS



ABOUT & HELP

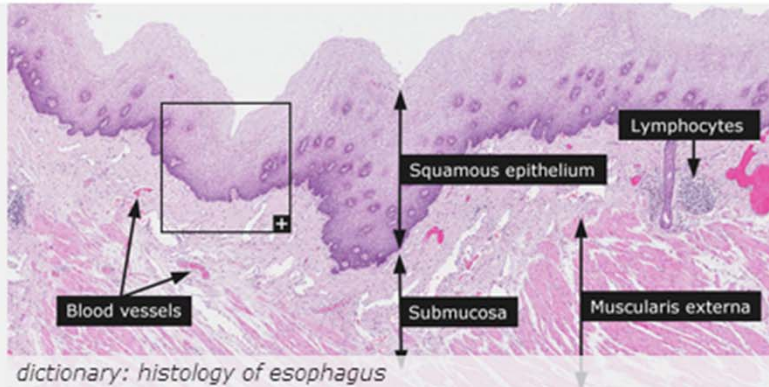
SEARCH ? »

Search

Clear

Fields »

e.g. [CD44](#), [ELF3](#), [KLK3](#), or use Fields to search specific fields such as [protein_class:Transcription factors](#) or [chromosome:X](#)



dictionary: *histology of esophagus*

News

Protein evidence according to [Fagerberg et al](#) is summarized in the [chromosome progress diagram](#).

Version: **11.0**

Atlas updated: 2013-03-11
[release history](#)

15156 genes with protein expression profiles based on **18707** antibodies.

*Knut och Alice
Wallenbergs
Stiftelse*

The Human Protein Atlas project is funded by the Knut & Alice Wallenberg foundation.



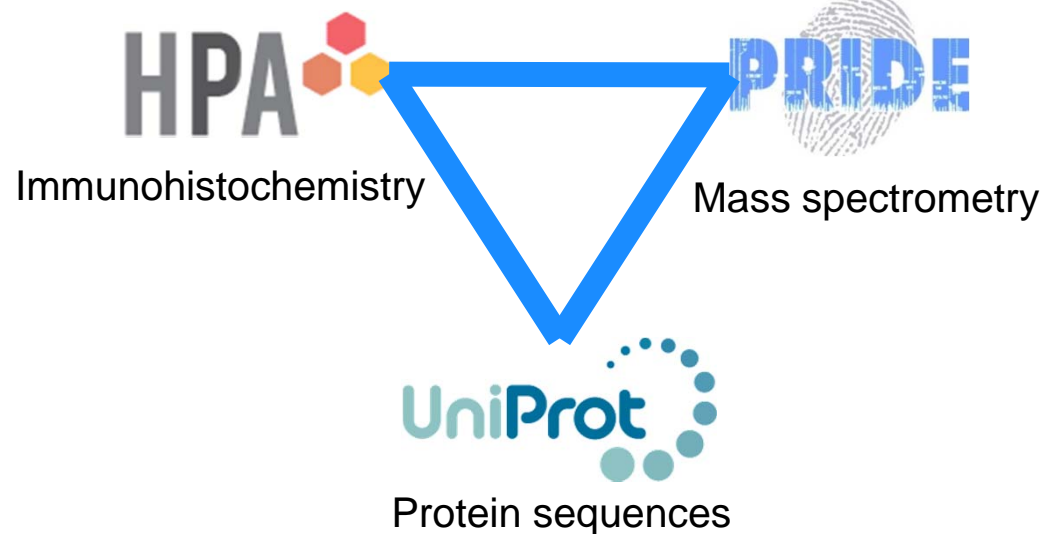
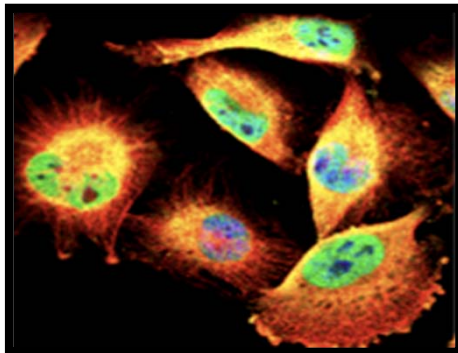
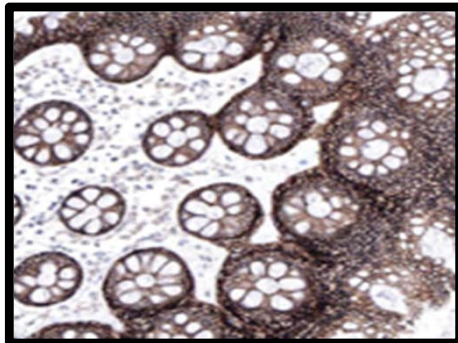
- Version 11 launched March 2013
- 18,707 antibodies from more than 50 providers
- 15,156 genes (~75% of all protein coding genes)
- ~12 million images
- All data publicly available (free)

www.proteinatlas.org



Integrating HPA into the Elixir landscape

- Aims to define and implement tools and standards to help the scientific community to access and integrate expression data



Methodology

- Add a service layer on top of existing database to provide data in standardized format
 - Keeps independence of original source
 - On demand integration
- Distributed Annotation System (DAS)
 - HPA DAS source <http://das.proteinatlas.org/das>
- DASTy web application integrating HPA data
 - <http://www.ebi.ac.uk/Tools/biojs/registry/Biojs.HpaSummaryFeatures.html>

The screenshot displays the Human Protein Atlas summary (HPA) interface. It features a header with the title "Human Protein Atlas summary (HPA)" and a sub-header "Antibody HPA001380". Below this, there are two main sections: "HPA001380 Cancer tissue immunohistochemistry summary" and "HPA001380 Normal tissue immunohistochemistry summary". Each section includes a list of bullet points summarizing the staining results. For example, the cancer tissue summary states: "Most malignancies were negative. Two (or more) antibodies yielding similar staining patterns which are partly consistent with gene/protein characterization data or consistent with limited gene/protein characterization data". The normal tissue summary states: "Cardiac myocytes showed strong membranous positivity. Glomeruli, smooth muscle and endothelium showed positivity. Two (or more) antibodies yielding similar staining patterns which are partly consistent with gene/protein characterization data or consistent with limited gene/protein characterization data". The interface also includes a small image of a tissue section and a link icon.

HPA data exchange standard – HPA-XML

- Structured data in XML format
- Served via normal REST web service
- Available as
 - Complete dataset
 - proteinatlas.org/download/proteinatlas.xml.zip
 - Subset of data corresponding to query
 - proteinatlas.org/search_download.php?format=xml&query1=P16070
 - proteinatlas.org/search_download.php?format=xml&query1=KLK3
- Verified and documented with XSD schema
 - proteinatlas.org/download/proteinatlas.xsd

HPA data exchange – current usage

- NextProt
 - HPA-XML (since 2011)
- UniProt
 - CSV-file (to be replaced by HPA-XML autum 2013)
- UniProt GO
 - CSV-file (to be replaced by HPA-XML autum 2013)
- UniProt xref
 - HPA-XML (since 2013)
- Ensembl
 - CSV-file (to be replaced by HPA-XML autum 2013)
- General usage on www.proteinatlas.org
 - HPA-XML (since 2011)

HPA-XML future

- XML content and structure will be extended to include all HPA data
- Use HPA-XML format for all collaborators
- Improve query interface, tailored to ELIXIR needs
- Provide extensive API documentation