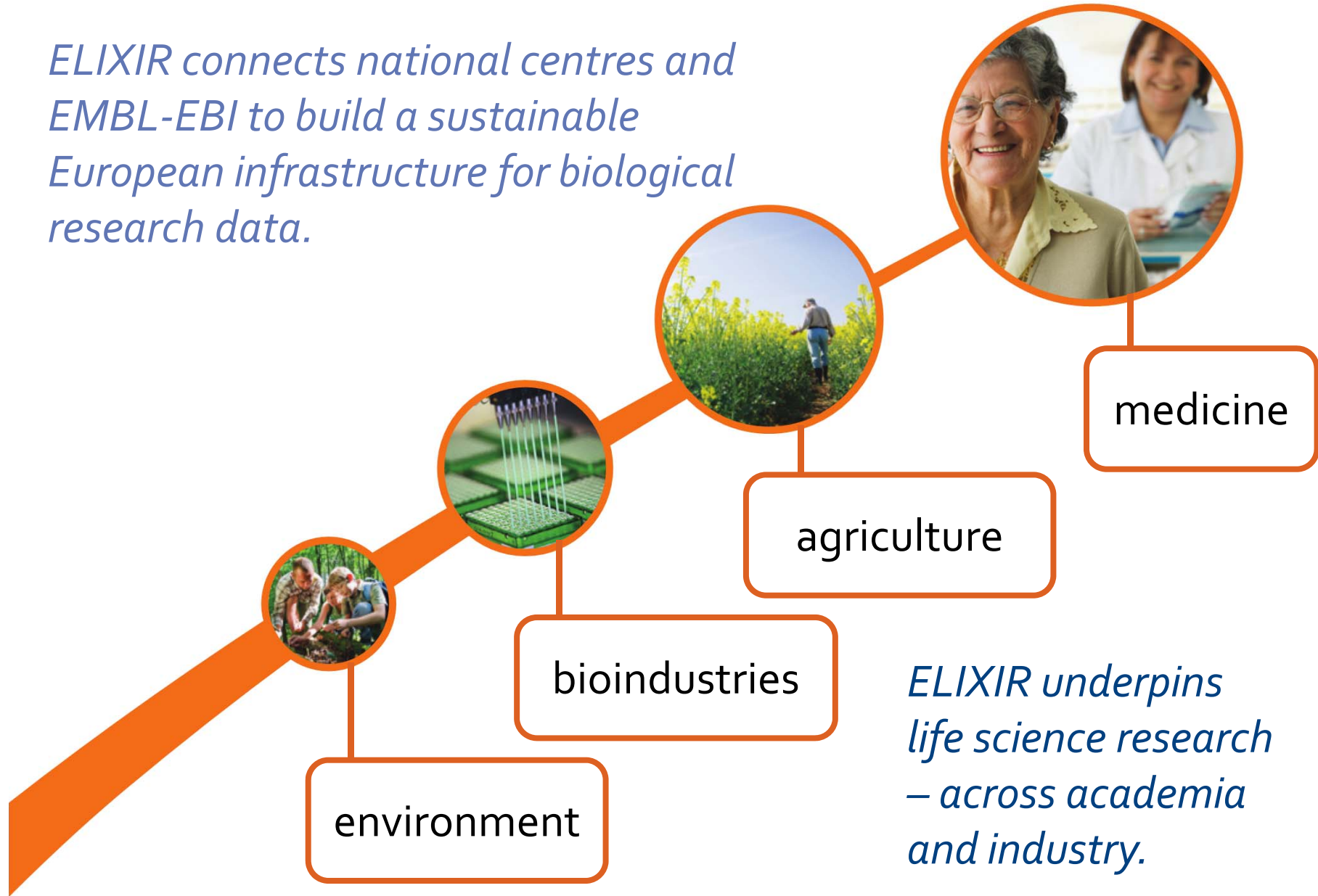


ELIXIR connects national centres and EMBL-EBI to build a sustainable European infrastructure for biological research data.

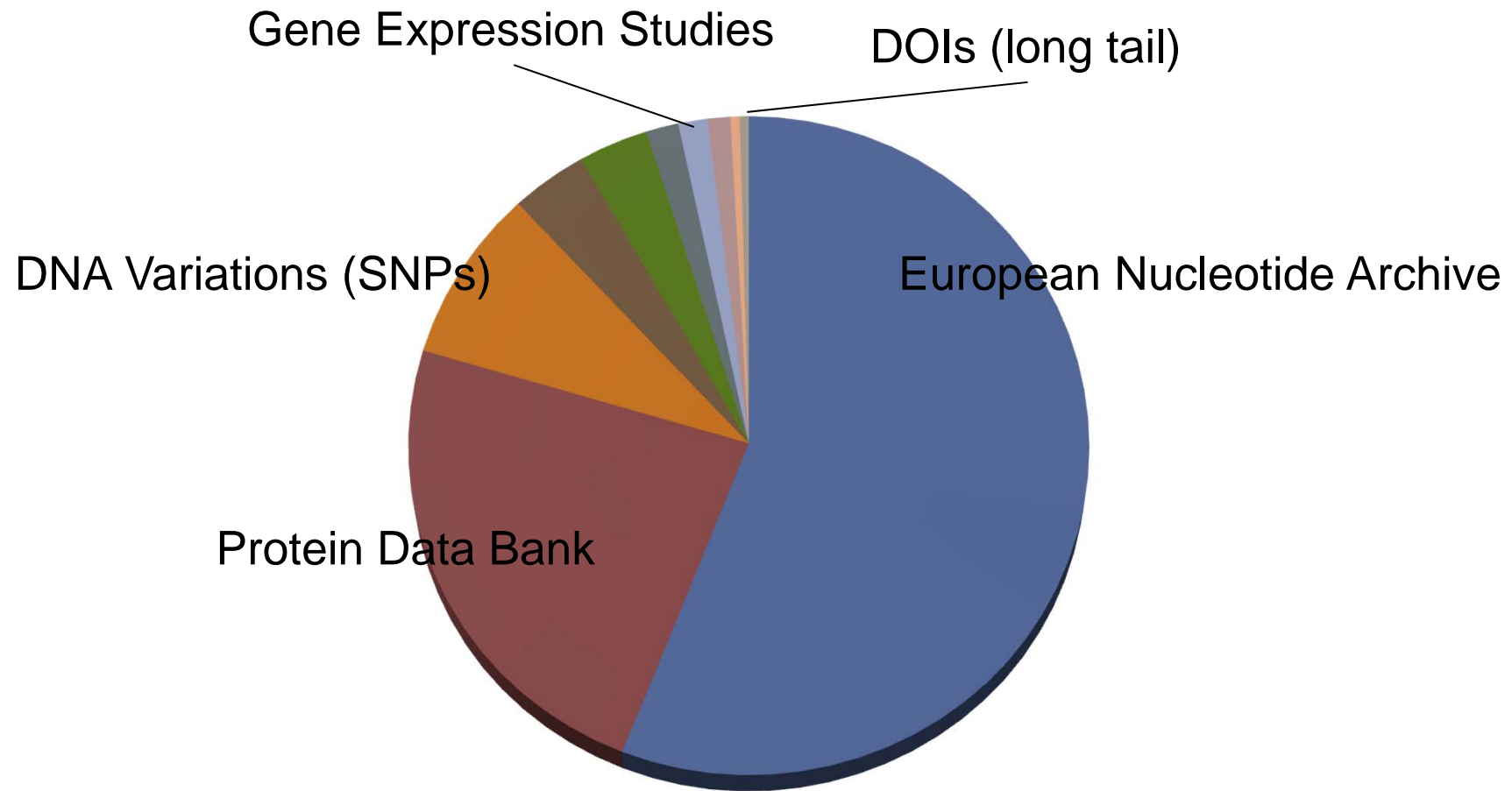


ELIXIR underpins life science research – across academia and industry.

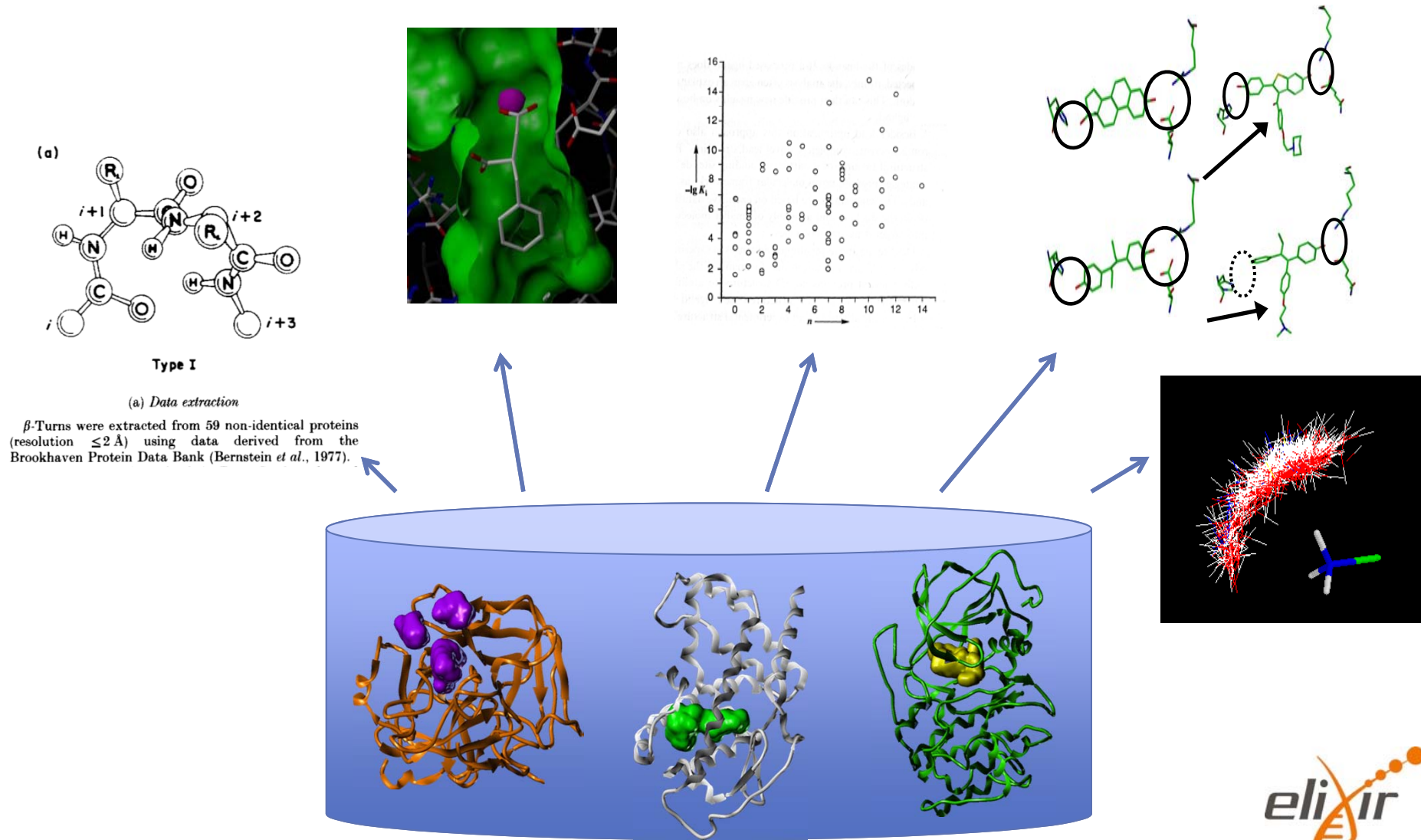
Data infrastructure for the life sciences



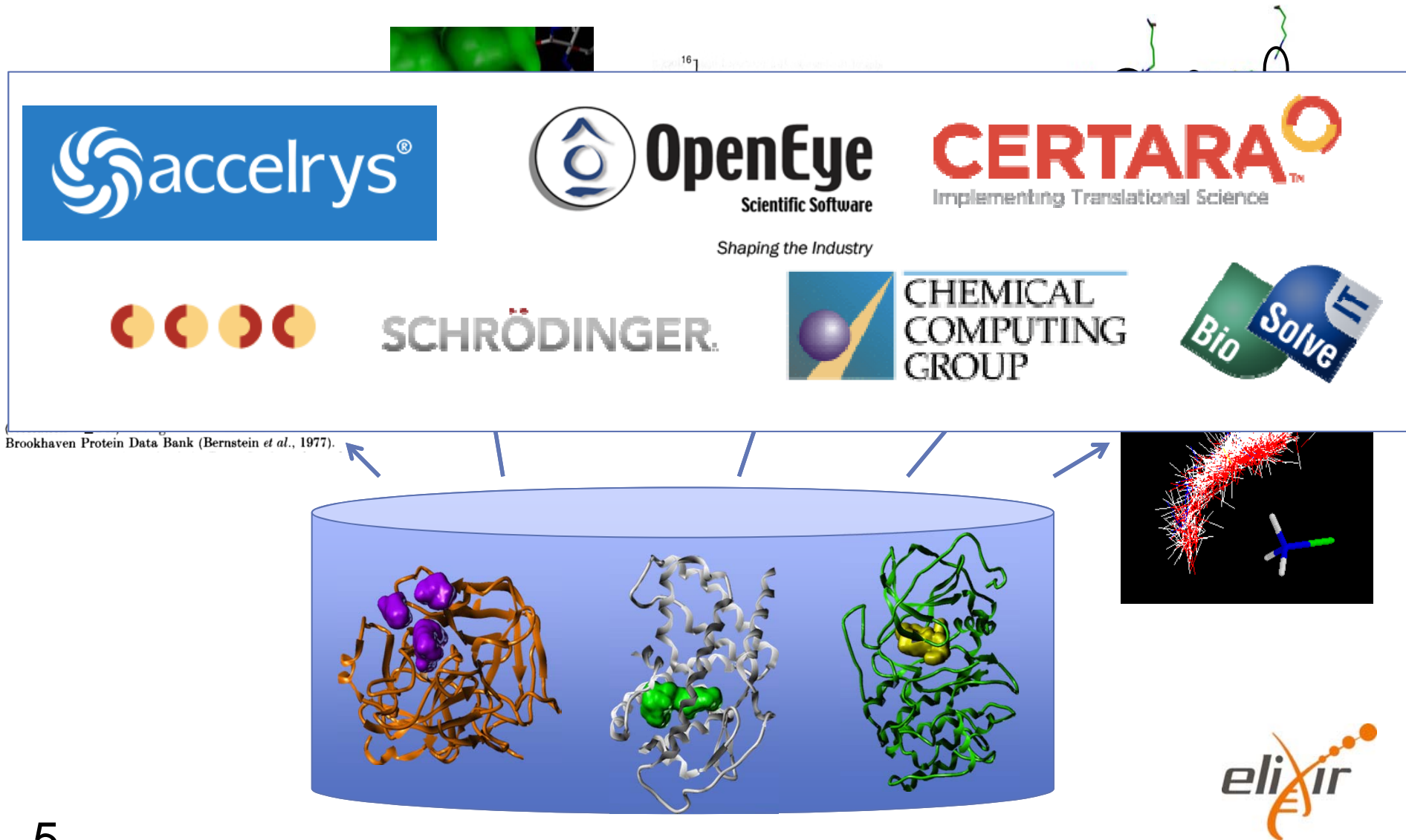
Data Citation in Europe PMC full text articles



Sustainable archives - value of data reuse?

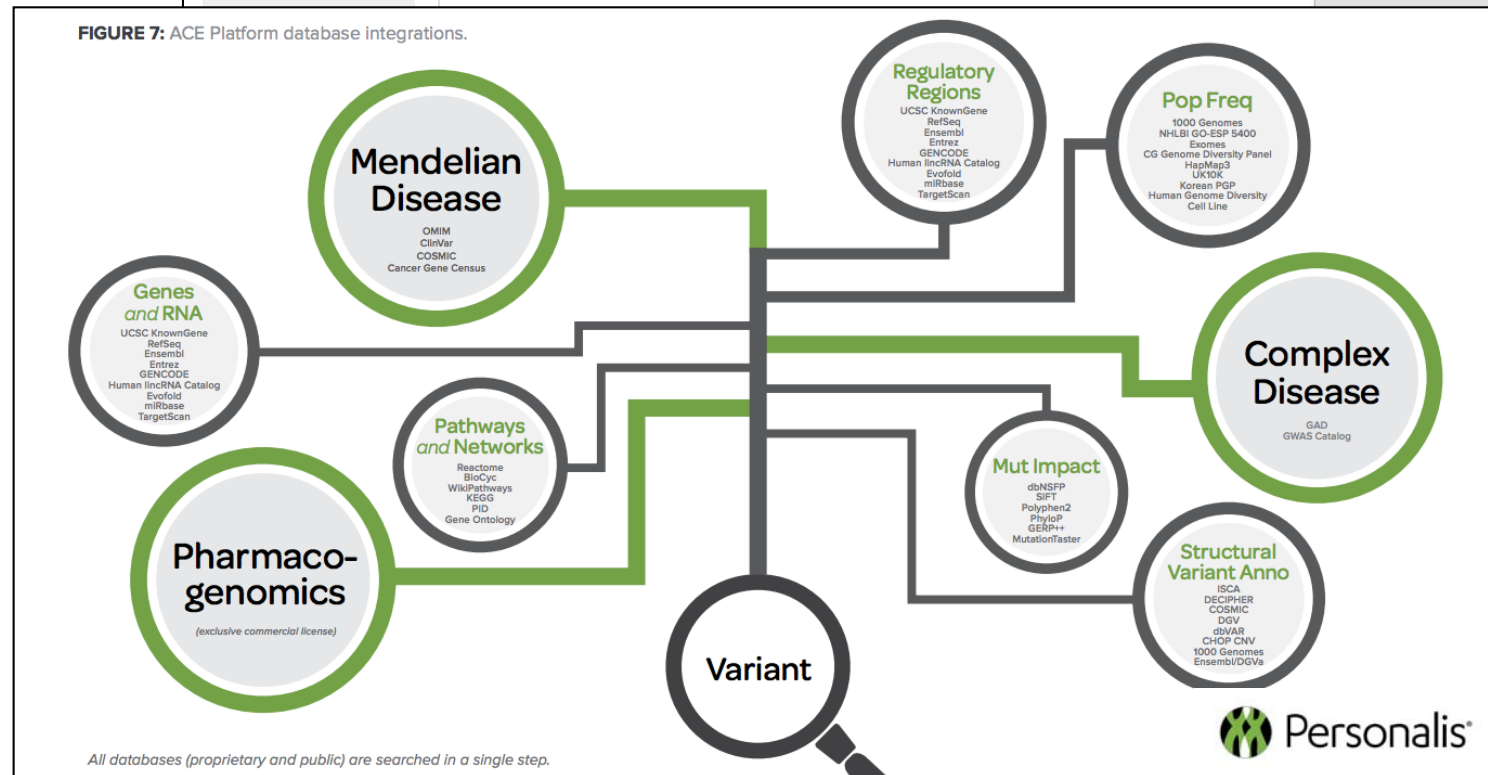


Sustainable archives - value of data reuse?



Public data: foundation for biotech innovation

The screenshot shows the NextBio website interface. At the top left, there is a navigation menu with links for Home, My Data, Bookmarks, Collaborations, and Inbox. Below this is an 'Import Your Data' button and a 'Literature' section with a 'FAQ' link. The main content area features a search bar with the text 'Enter Query Term' and a 'QuickView' button. Below the search bar, there is a list of search results, including one for 'Ontology-based meta-analysis of global collections of high-throughput public data'. The interface also includes various icons for different data types and a 'Mining (technology)' section on the right.



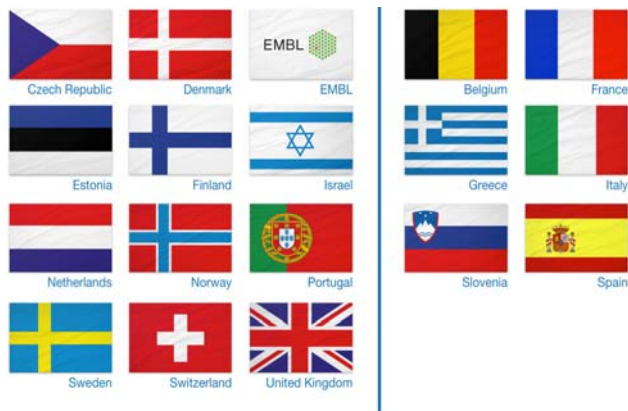
Life-science and data infrastructure

- Data production and using at a large number of sites across Europe
- Human genomics projects but also plants, microbiota, environmental marker organisms
- Metabolomics & Proteomics coming of age
- Be scalable to 1000s of sites
 - Over 500 000 Life scientists in Europe
- Deal with incomplete, conflicting data and rapidly evolving knowledge



ELIXIR: the European Research Infrastructure for biological data

- ELIXIR connects national infrastructures and EMBL-EBI
- 11 Member states + EMBL:
 - Czech Republic, Estonia, Denmark, Finland, Israel, Netherlands, Norway, Portugal, Switzerland, Sweden, UK
- 6 countries working towards membership:
 - Belgium, Greece, France, Italy, Slovenia, Spain



A distributed infrastructure scales with the challenges

- **ELIXIR** deliver services through national ELIXIR Nodes
- **ELIXIR** Nodes build on national strengths and priorities
- **ELIXIR** Nodes build local bioinformatics capacity throughout Europe



A distributed infrastructure scales with the challenges

- **ELIXIR** deliver services through national ELIXIR Nodes
- **ELIXIR** Nodes build on national strengths and priorities
- **ELIXIR** Nodes build local bioinformatics capacity throughout Europe



ELIXIR: The Norway Node

ELIXIR's Norway Node will provide competence and infrastructure building on key areas for Norway, in particular marine resources and medical research. Challenges related to processing and analysis of data from next-generation sequencing and other high-throughput methods are important both to basic research in these areas, and to the development of new enterprises. The node will also provide training and support toward researchers.

Collaborating organisations

University of Bergen
The coordinating partner of the ELIXIR Norway node. The main focus is on marine genomics and e-infrastructure. An early deliverable is a database developed in tight collaboration with the Sea User Research Centre.

University of Oslo
Emphasizes biomedical resource provision and analysis, leveraging public resources in integrative statistical genomics, with secure management of person sensitive data.

Norwegian University of Life Sciences
Main focus on providing genomic resources for species-oriented and comparative fish genomics. Provision of web-based solutions for services, toolboxes, and computational access to these data.

Norwegian University of Science and Technology
Tools and resources for analysing genome data, with focus on gene regulation, non-coding RNAs and epigenetics, but also bacterial genomics. Handling and analysis of data from human biobanks.

University of Tromsø
Tools and pipelines for analysing metagenomic (and genomic) data, with a particular focus on taxonomic classification and bioprospecting (functional and metabolic potential).

Marine research
The Norwegian ELIXIR Node will provide services and resources toward marine genomics including researchers, government, and industry. The Norwegian Node will offer several integrated packages geared towards large scale analysis of marine genomic and metagenomic data (e.g. fish genomics and marine bio-prospecting). This also includes provision of web-based solutions for services, toolboxes, and computational access to reference data provided by the ELIXIR infrastructure.

Health and biobanks
The Norway Node supports infrastructure for handling and analysis of data for medical research, including human biobanks. Such data may be sensitive, and must be stored with secure access. The node is developing infrastructure for sensitive data. Tools for data analysis are integrated into Next-Seq, the Norwegian e-infrastructure for life sciences. This provides user friendly solutions for example for human re-sequencing data and other genome-scale analysis.

Contact

Public contact person:
ELIXIR Norway Node
Computational Biology Unit
Department of Informatics
University of Bergen, Norway
mailto:elixir@iuh.uib.no

The Research Council of Norway



Data Roaming...

- Access of large, and growing, reference data sets
- Access human reference and study data
- Standardizing, curating and depositing large data-sets

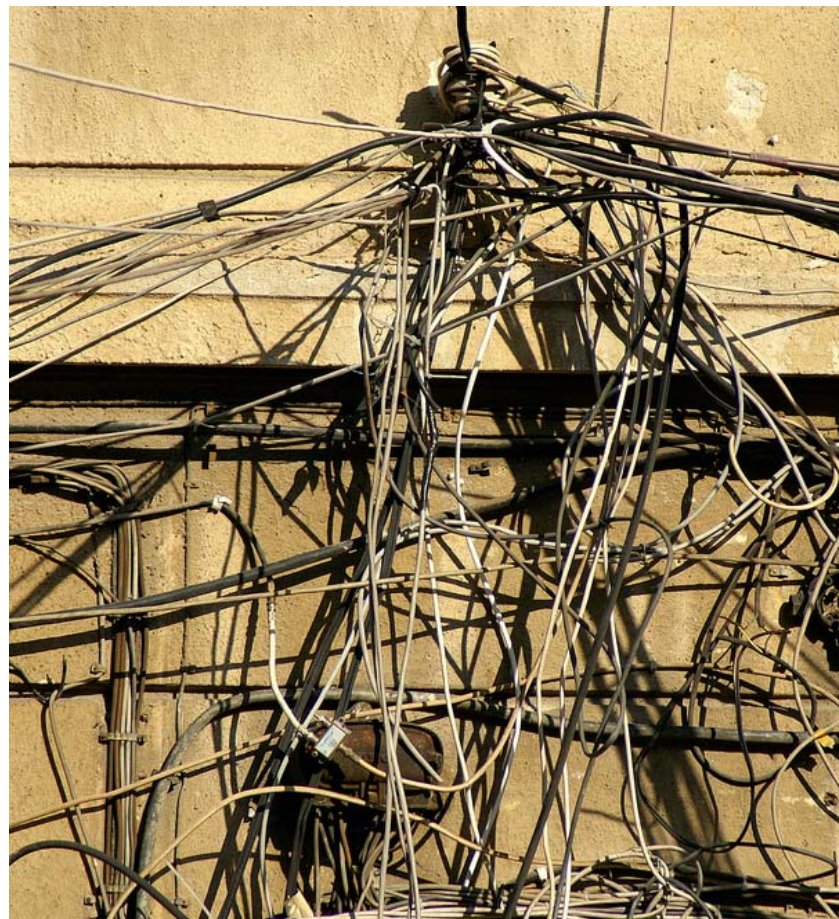


Integration and interoperability of resources

"... typically 40% of our effort in biomarker discovery is data integration"

J. Wojcek, CEO, Quartz Bio

- Formats, Ontologies, Guidelines, ...
- Professional skills for managing and exploiting data

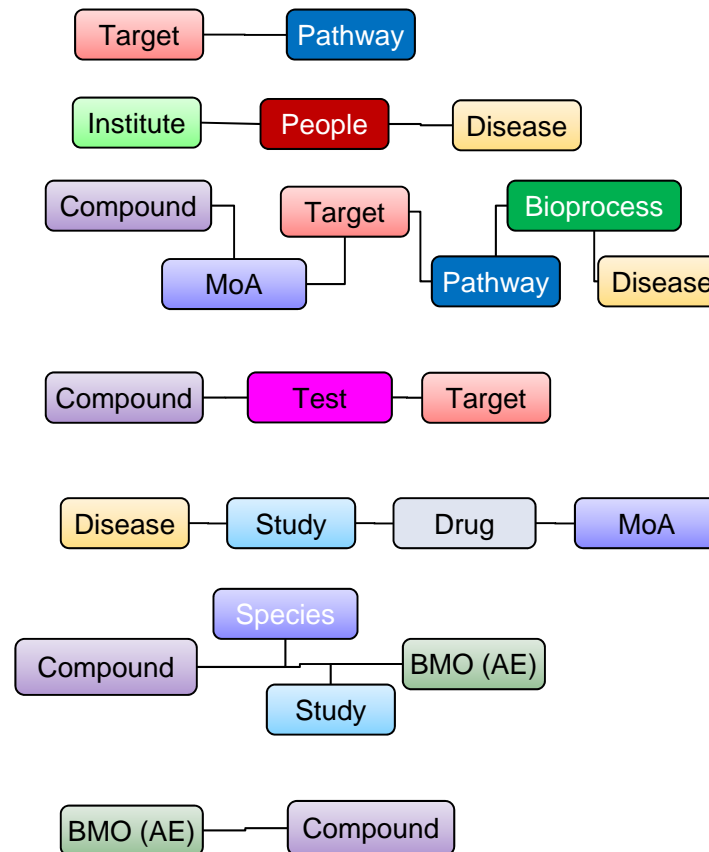


Different Questions, Common Language

Question

- Which proteins are in the EGFR pathway?
- What compounds bind *CCR1* with a K_d less than 5nM
- What mechanisms of action have been tested for the treatment of melanoma?
- What adverse events for propranolol have been observed in rats?
- What compounds cause cardiotox (screen development)?
- Can any of our CDs be prepositioned vs PAH?

Concepts



Data interoperability – Human Protein Atlas

Thyroid gland

Annotated expression: Glandular cells

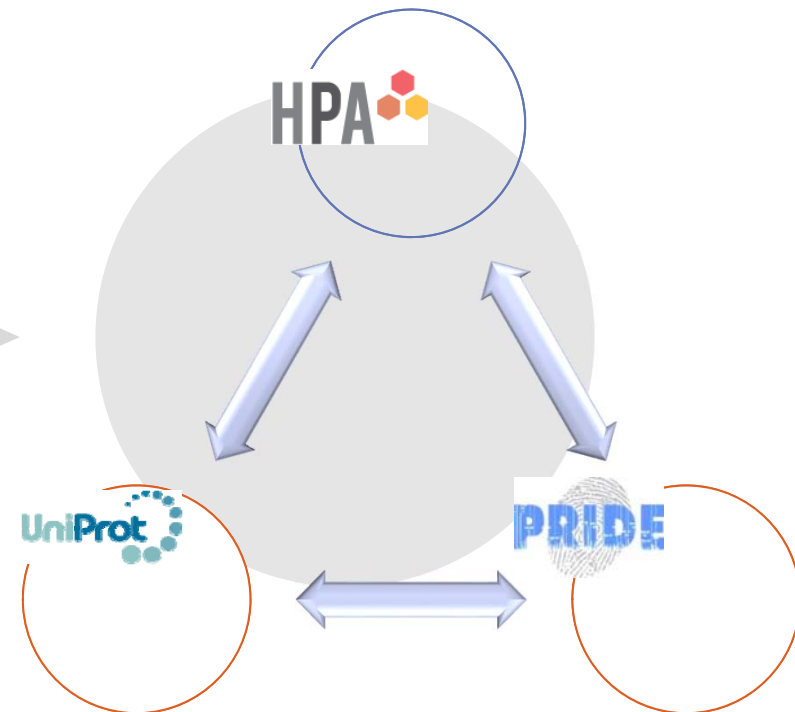
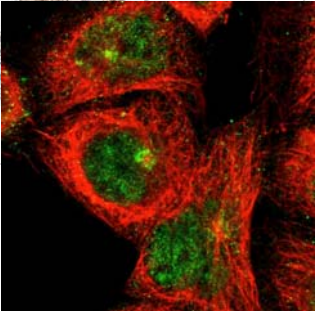
Cell type	Antibody ID: 102555	Antibody ID: 112557
Glandular cells	Glandular cells	Glandular cells
Intensity	Strong	Strong
Quantity	>75%	>75%
Location	Cytoplasmic/membranous	Cytoplasmic/membranous, nuclear
Antibody staining		

Level of antibody staining: Strong, Moderate, Weak, Negative

Level of annotated protein expression: High, Medium, Low, None

Dictionary: Thyroid gland

Gender: Female	Female
Age: 22	22
Tissue characterisation: Thyroid gland (T-96000) Normal tissue, NOS (M-00100)	Thyroid gland (T-96000) Normal tissue, NOS (M-00100)
Patient: 2146	1712
Gender: Female	Female
Age: 75	44
Tissue characterisation: Thyroid gland (T-96000) Normal tissue, NOS (M-00100)	Thyroid gland (T-96000) Normal tissue, NOS (M-00100)
Patient: 1501	3005
Gender: Male	
Age: 61	
Tissue characterisation: Thyroid gland (T-96000) Normal tissue, NOS (M-00100)	
Patient: 2072	



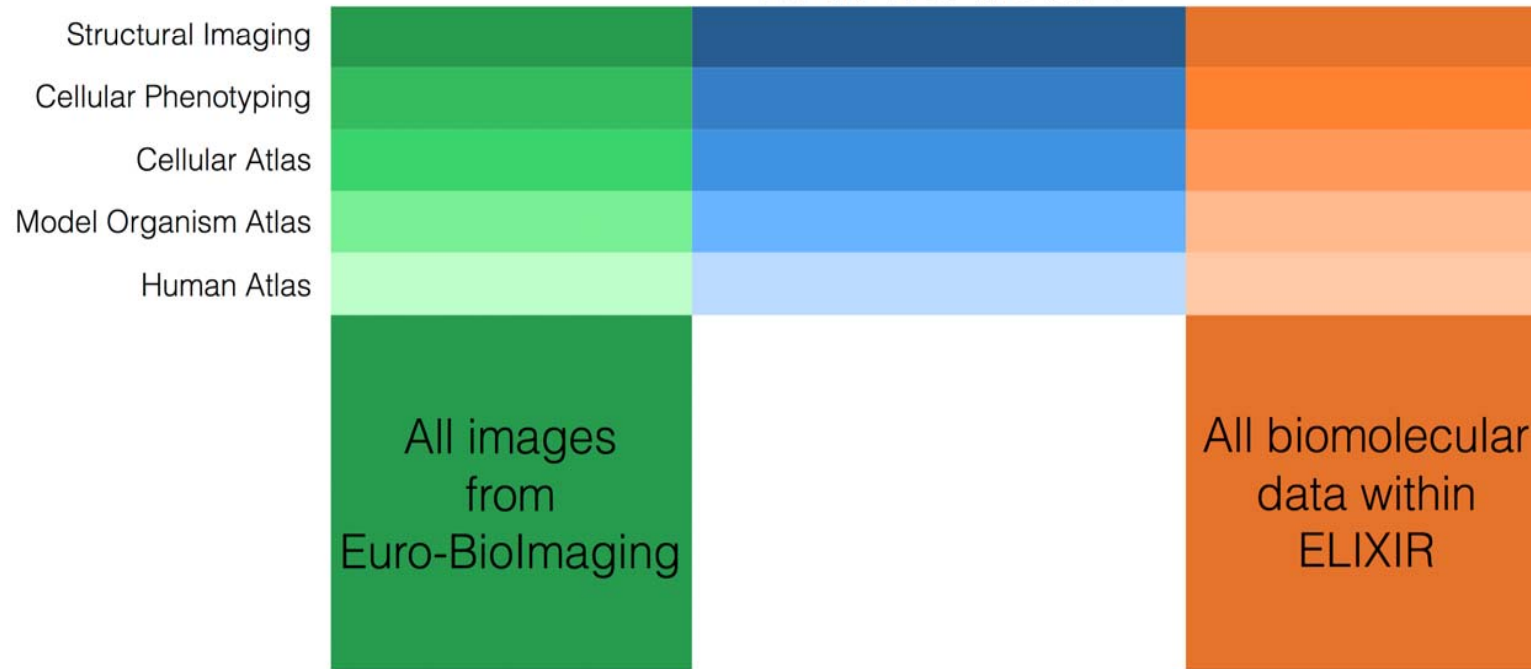
“Bring your own data” - Hands on interoperability

- Problem-centered workshops
- Integration experts - Data resources –Users
- ELIXIR funds external trainers
- Great feedback –turning into programme

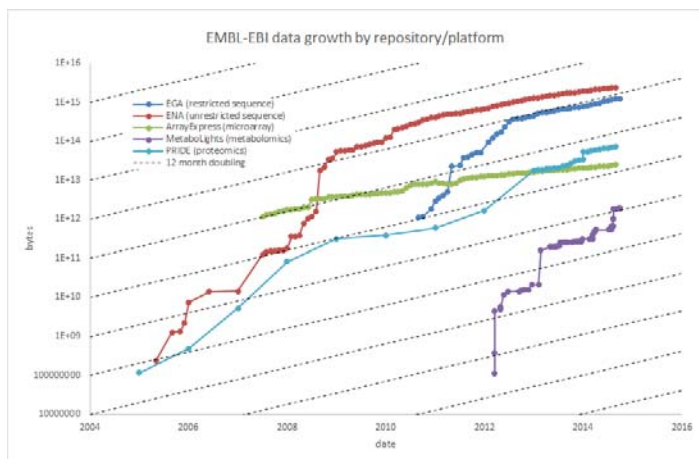


ELIXIR -EuroBioImaging Technical Strategy

Euro-BioImaging — ELIXIR DATA BRIDGE



Industry values ELIXIR resources



>110M annual hits in EBI resources in 2013

Aggrecan is a member of the chondroitin sulfate proteoglycan family, which also includes versican, brevican and neurocan. An exemplary aggrecan 1 sequence can be found under the NCBI gene ID 176 (ACAN; Ensembl: ENSG00000157766; HPRD:01123; MIM:155760). Exemplary versican sequence can be found under the NCBI gene ID (VCAN; Ensembl: ENSG0000038427; UniProtKB: P13611). Exemplary brevican sequence can be found under NCBI gene ID (BCAN; Ensembl: ENSG00000132692; UniProtKB: Q96GW7). Exemplary neurocan sequence can be found under the NCBI gene ID (NCAN; Ensembl: ENSG00000130287; UniProtKB: O14594). This gene family is highly homologous and exhibits similar protein functions containing extensive protein domains of greater than 50% amino acid identity. ?Aggrecan?, as used herein also includes any naturally-occurring variants and splice variants of aggrecan, versican, brevican and neurocan, and any variants of aggrecan, versican, brevican and neurocan due to splicing by different cell types.

SureChEMBL patents in 2014:

- 590 quoting UniProt
- 197 quoting Ensembl



Data, Ontologies and resources behind industry driven KM efforts



ELIXIR is strengthening the local innovation environment

SME specialised events through ELIXIR nodes:

- Understanding of ELIXIR resources
- Meeting PI's of the resources
- Leverage the best out of ELIXIR resources



Industry Advisory Committee members

- Philippe Sanseau, GlaxoSmithKline, UK
- Martin Ebeling, Hoffmann-La Roche, Switzerland
- Wendy Filsell, Unilever R&D, UK
- Jakob de Vlieg, Bayer CropSciences Innovation Center, Belgium
- Mark Forster, Syngenta, UK
- Iain Hrynaszkiwicz, Nature Publishing Group - Macmillan, UK
- Angel Pizarro, Amazon Web Services, USA
- Natalia Jiménez Lozano, Bull, Spain
- Montserrat Vendrell, Biocat, Spain
- Claus Stie Kallesøe, Gritsystems A/S, Denmark
- Anita Eliasson, Biocomputing Platforms Ltd, Finland



Thank you!

