ELIXIR: A Life Science Infrastructure for Big Data and Open Data
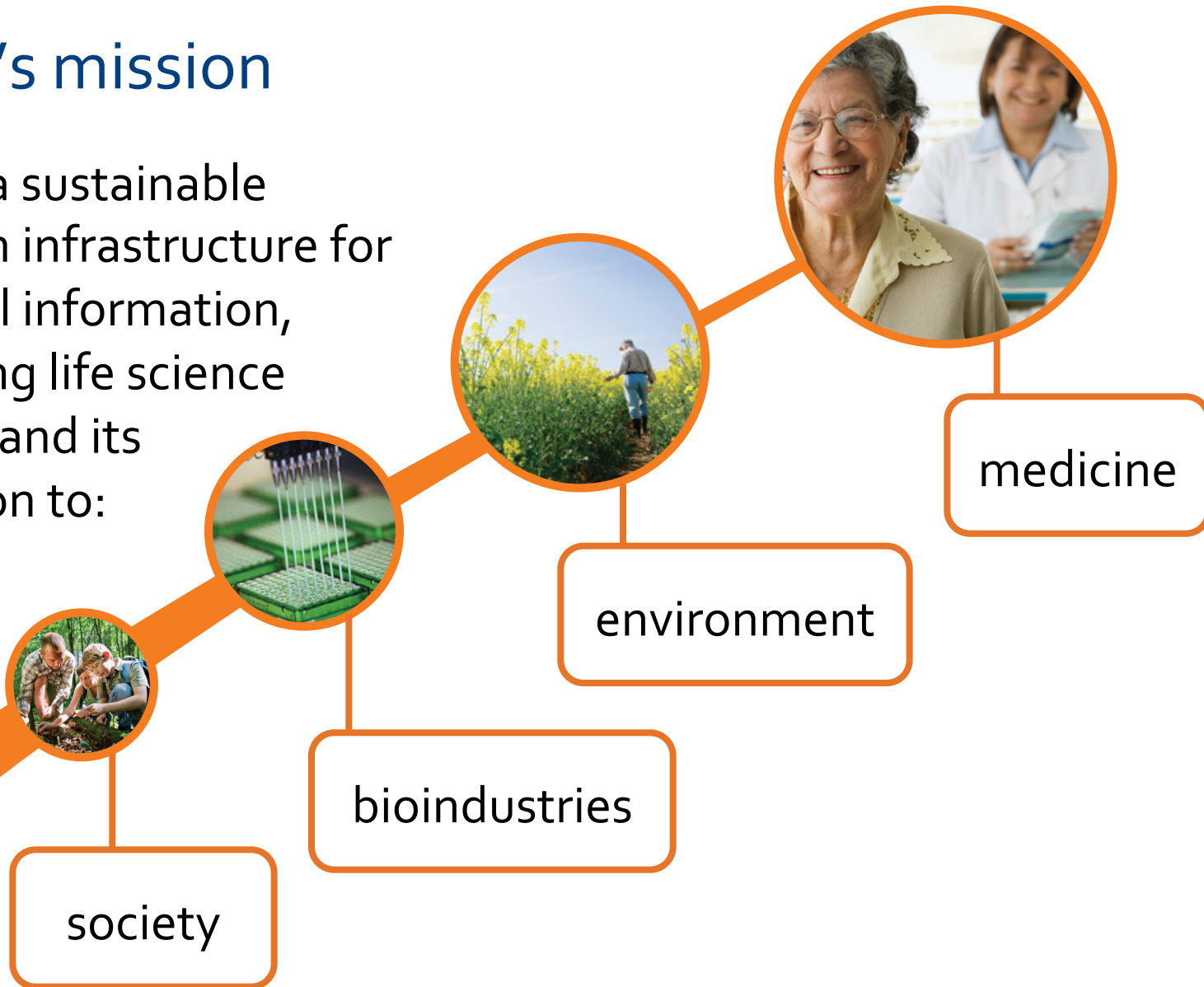
Alf Game, ELIXIR Chair

ELIXIR's mission

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

society

bioindustries

environment

medicine

Genome-wide analysis of crop plants

- Population growth and climate change are major challenges to food security.

- Traditional routes to crop improvement are too slow to keep up with this increase in demand.

- Understanding plant genomes helps us identify which species will be 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-27 will develop breast cancer before the age of 80.

- If we can identify patterns of genes that are active in different tumours, we can diagnose and treat cancers earlier.
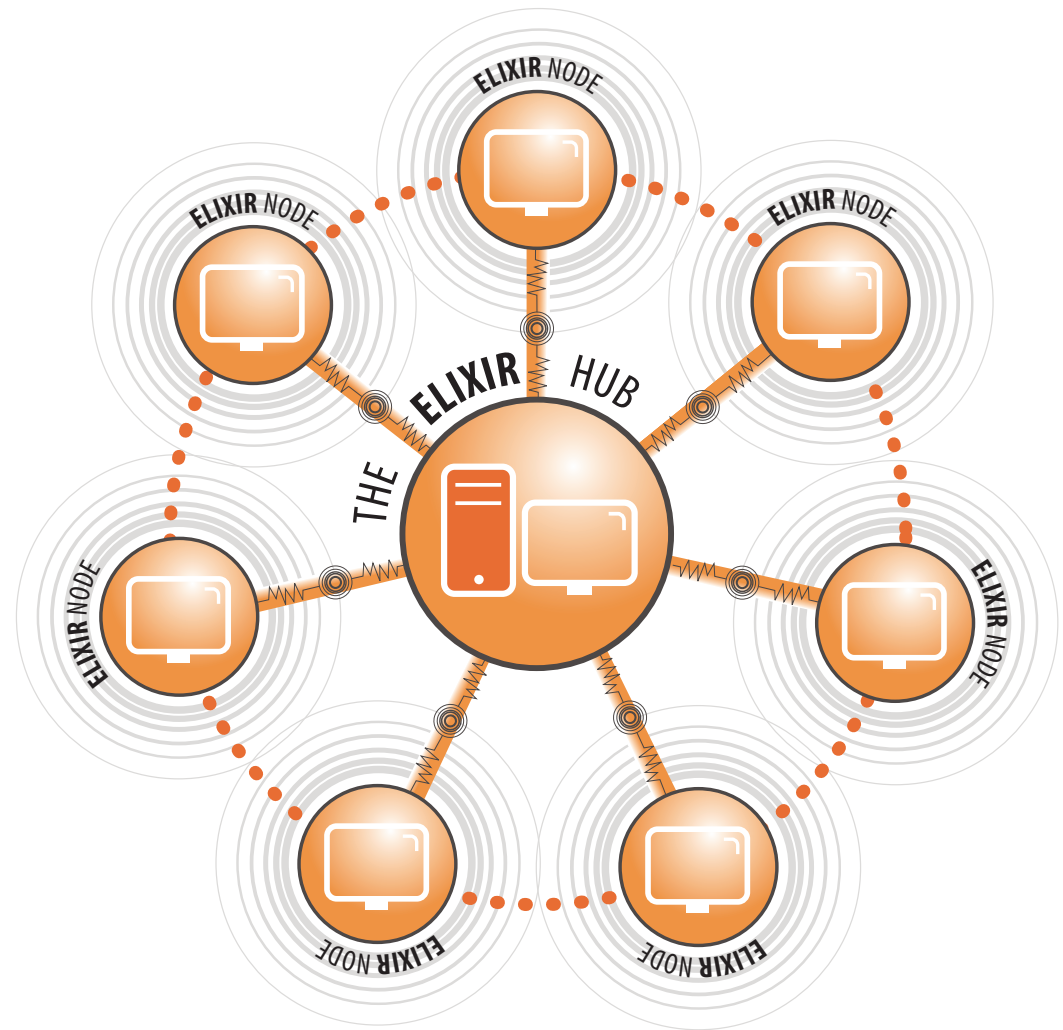
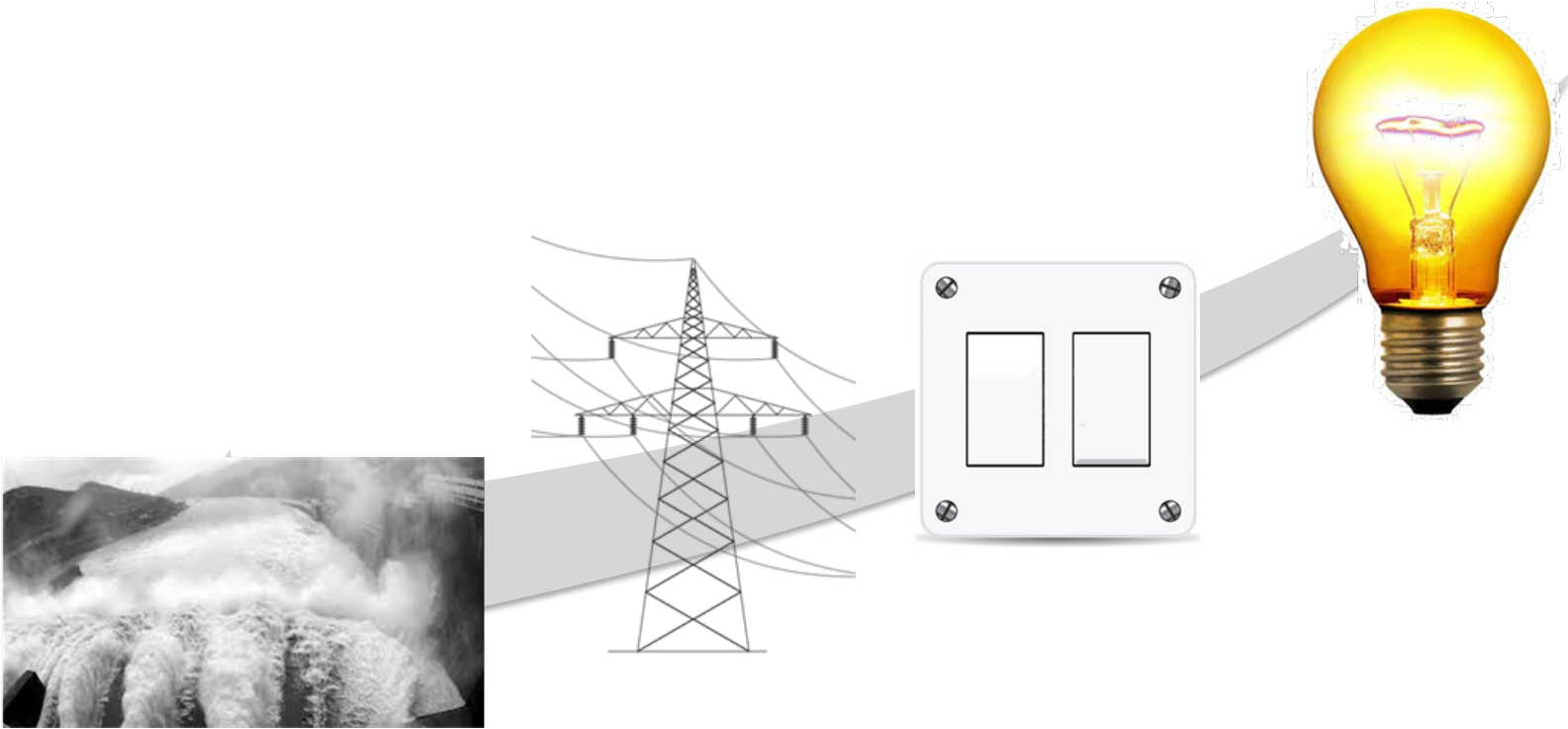# A distributed infrastructure to scale with the challenge

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

What appears a simple, reliable user experience…

…is made possible by robust, non-trivial infrastructure that often goes unnoticed

The data challenge

- Data production and using at a large number of sites across Europe
  - (European Illumina sales up 20% 2013)
- Human genomics projects but also plants, microbiota, environmental marker organisms
- Metabolomics & Proteomics coming of age
  - UK National Phenome facility
- Be scalable to 1000s of sites
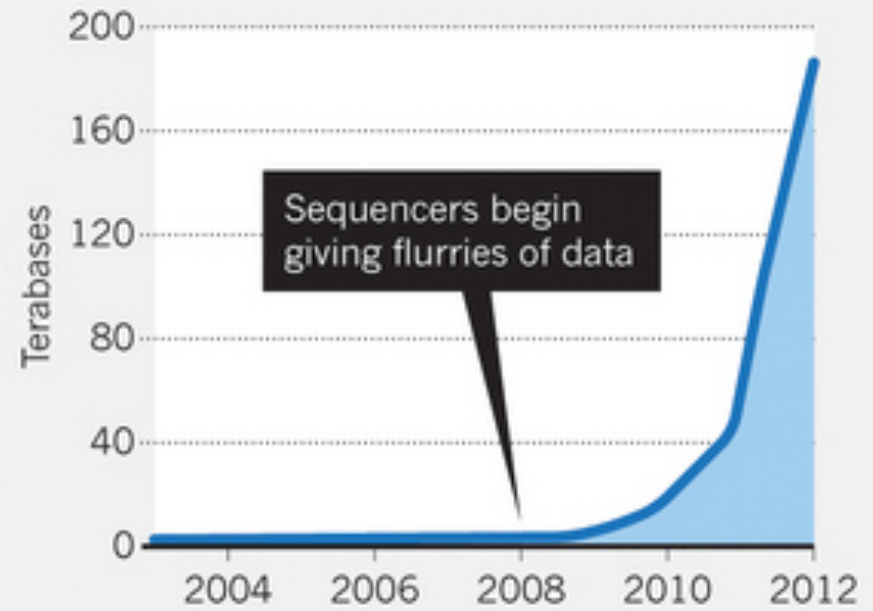- Deal with incomplete, conflicting, and incorrect data
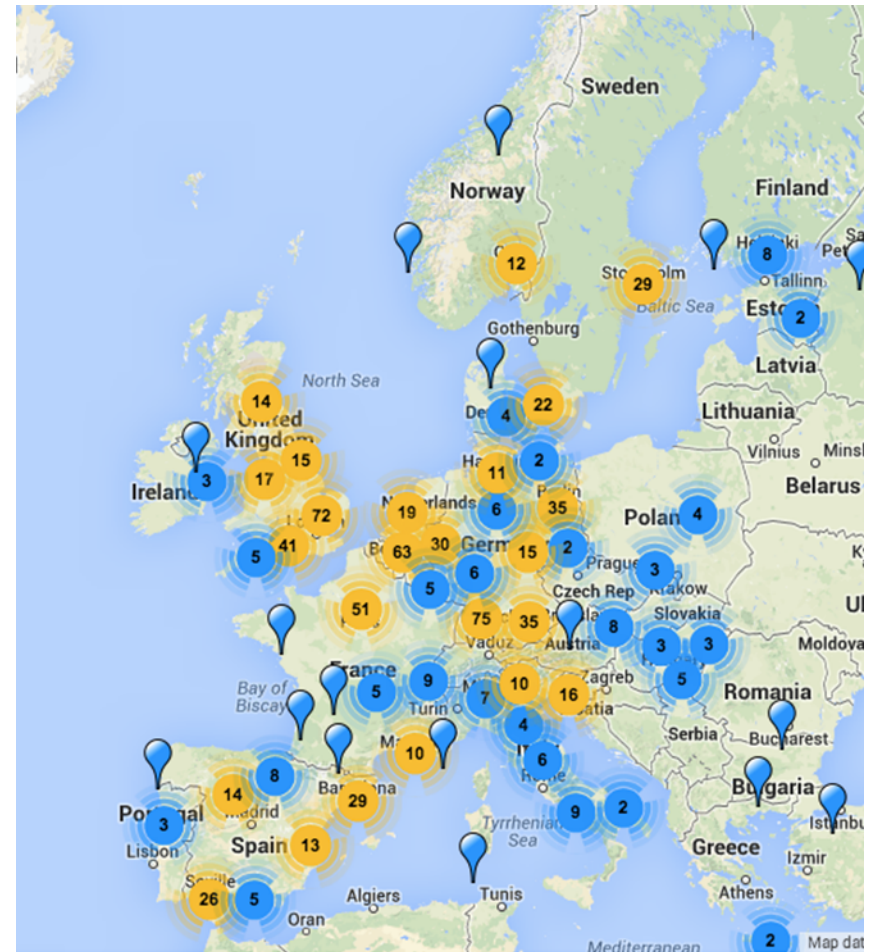
The data deluge

- Computer speed and storage capacity is **doubling every 18 months** and this rate is steady

- DNA sequence data is **doubling every 6-8 months** over the last 3 years and looks to continue for this decade

Source: *Nature News & Comment*, June 2013

Challenges for life-science data services

Scale and Sustain funding

Distributed infrastructure with >1M users

Managing and interoperate big and heterogeneous data

Capacity Compute. Capability Storage

Integrating clinical and translational data

Privacy and ethical concerns

Algorithms to data – clouds, research environments…

ELIXIR's services

- Tools
- Standards
- Data
- Compute
- Training
- Industry
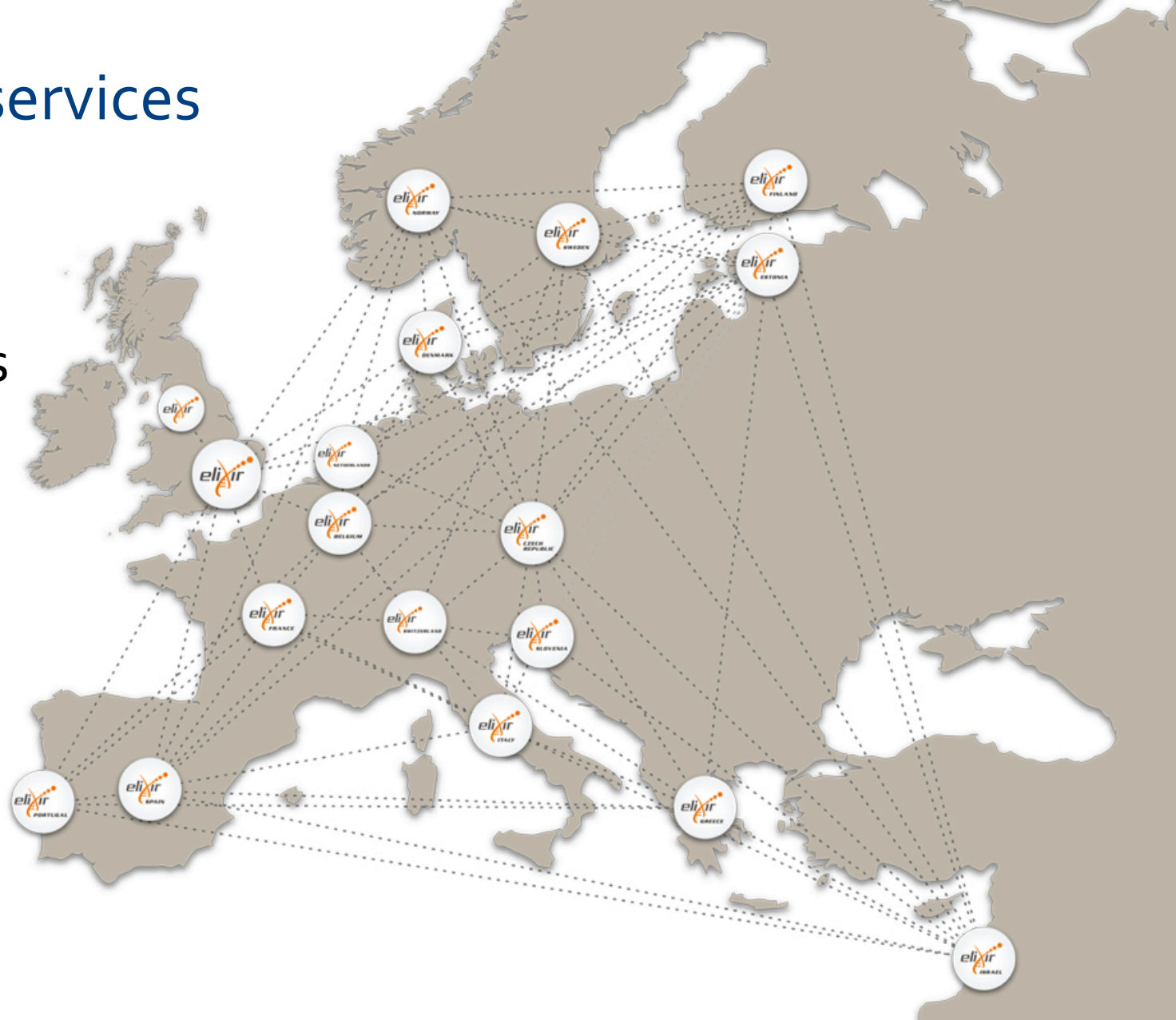
ELIXIR Pilot Projects

- Five short-term Pilot Actions to act as test beds for integration of ELIXIR services:

1. ELIXIR Facing Cloud Support and Virtual Machines - with SIB

2. ELIXIR Data IO to pilot the continuous transfer of major archive resources to a remote European location - with CSC, Finland

3. Establishing EGA Distributed authentication - with CSC, Finland

4. Establishing EGA as joint venture – with CRG, Spain

5. Improving links between Human Proteome Atlas (HPA) and EMBL-EBI resources

Identifying new drug targets

ELIXIR pilot: Interoperability of high-resolution protein data at EMBL-EBI and HPA, Sweden
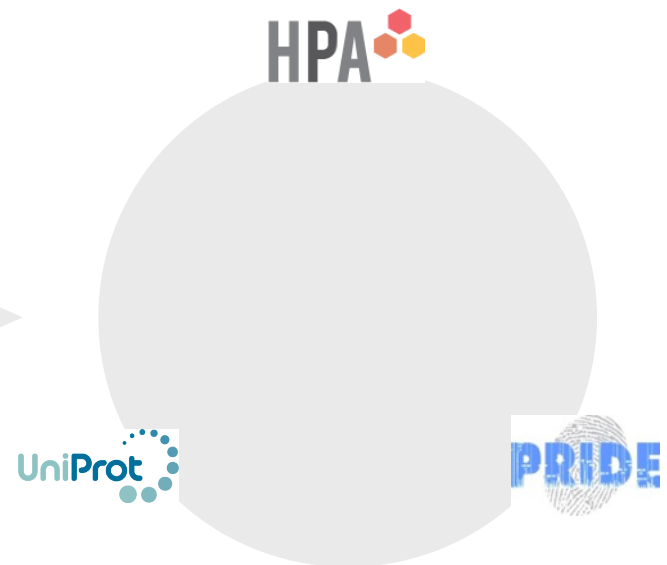

The Human Protein Atlas portal is a publicly available database with millions of high-resolution images showing the spatial distribution of proteins in 46 different normal human tissues and 20 different cancer types, as well as 47 different human cell lines.

Industry and Innovation

- Challenge to access and understand services
  - EU wide engagement programme - Build on strong local interactions through Nodes!
- Small companies – Big Data
  - ELIXIR Cloud for SMEs
- EFPIA and IMI
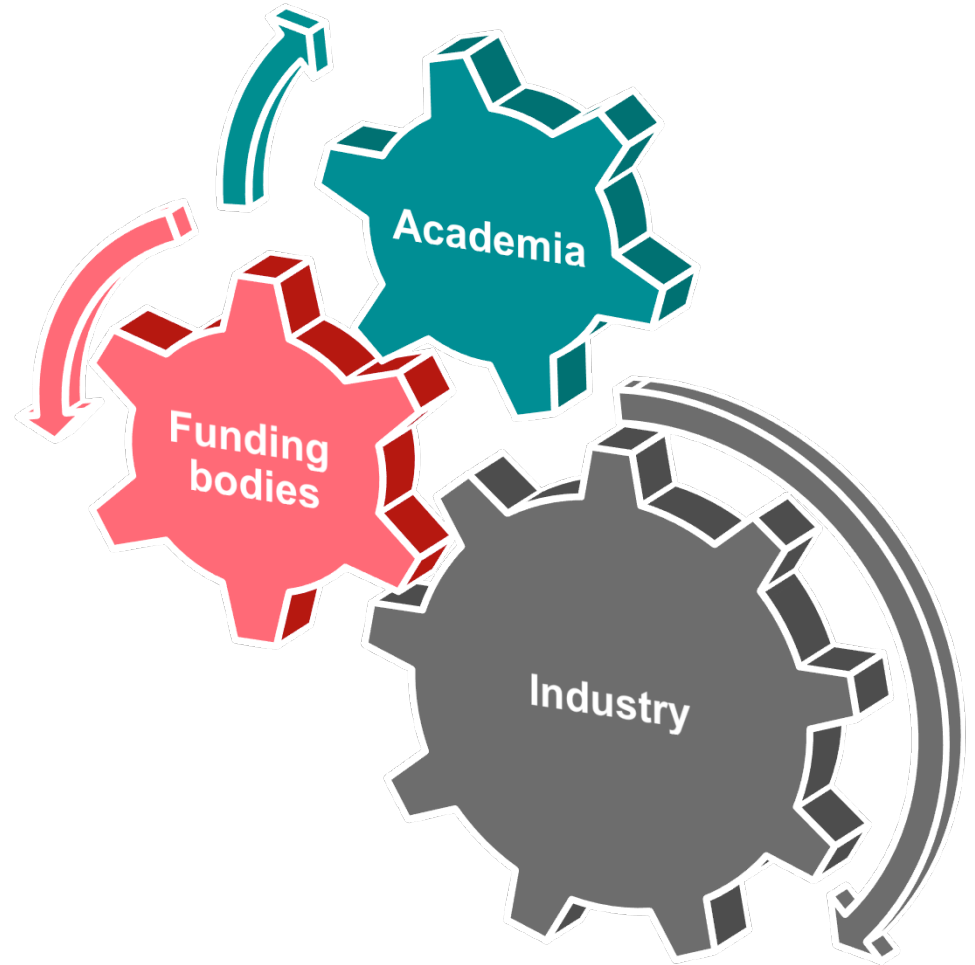
BioMedBridges
Ten new biomedical sciences research infrastructures stronger through common links

- 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

Thank you

Belgium

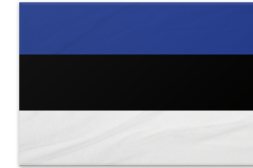
Czech Republic

Denmark

EMBL

Estonia

Finland

France

Greece

Israel

Italy

Netherlands

Norway

Portugal

Slovenia

Spain

Sweden

Switzerland

United Kingdom