BioMedBridges: Providing research and data links between the ESFRI BMS RI

e-IRG Workshop, Poznań, Poland

Wednesday 12 and Thursday13 of October 2011

The Hotel Mercure Poznań

Andrew Lyall PhD

ELIXIR Project Manager

EMBL-EBI

BioMedBridges

- First European consortium coordinated by ELIXIR
- FP7 Call 8 2.3.2: Award €10.6M, 4 years, 21 organisations, 12 WP
- To “build bridges” between the ESFRI BMS Research Infrastructures
- Participants are ESFRI BMS RI & European e-Infrastructures
- Deliverables are infrastructure components that will provide computational ‘data and service’ bridges between the ESFRI BMS RI, clustering them together and linking basic biological research data to data in the other domains.
- Also includes: GÉANT, DANTE, EGI.eu, PRACE & CERN

- Providing secure, robust and ethical access to data for a wide range of users will be the major effort of the project...

BioMedBridges: Participating Organisations

1	EMBL	ELIXIR	European Molecular Biology Laboratory	Germany
2	UOXF	INSTRUCT	University of Oxford	UK
3	KI	BBMRI	Karolinska Institutet	Sweden
4	STFC	INSTRUCT	Science and Technology Facilities Council	UK
5	UDUS	ECRIN	Heinrich Heine University Düsseldorf	Germany
6	FVB	EU-Openscreen	Forchungsverbund Berlin EV	Germany
7	TUM-MED	BBMRI	Technische Universität München	Germany
8	SZN	EMBRC	Stazione Zoologica Anton Dohrn	Italy
9	ErasmusMC	EuroBioImaging	Erasmus University Medical Center Rotterdam	Netherlands
10	TMF	EU-Openscreen	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.	Germany
11	HMGU	Infrafrontier	Helmholtz Zentrum Munich	Germany
12	MUG	BBMRI	University of Graz	Austria
13	VUMC	EATRIS	Vrije Universitaeit Amsterdam Medical Centre	Netherlands
14	Inserm	ECRIN/ERINHA	Institut National de la Santé et de la Recherche Médicale	France
15	UCPH	ELIXIR	University of Copenhagen	Denmark
16	UH	EATRIS	University of Helsinki, Institute for Molecular Medicine Finland	Finland
17	EGI	e-Infrastructure	European Grid Infrastructure	Netherlands
18	CSC	e-Infrastructure	Centre for Scientific Computing Finland	Finland
19	UMCG	BBMRI	University Medical Centre Groningen	Netherlands
20	CIRMMP	INSTRUCT	Consorzio Interuniversitario di Risonanze Magnetiche di Metalloproteine	Italy
21	DANTE	e-Infrastructure	Delivery of Advanced Network Technology to Europe	UK

ELIXIR: A first generation ESFRI RI*

Project – FP7, €4.5M, 4 Year, 32 Participants, Coordinated by EMBL-EBI

Aim – To build a ***sustainable*** European infrastructure for biological information, supporting life science research and its translation to medicine, the environment, the bio-industries and society.

Significance – Identified by ESFRI as being one the few RI that is of global significance for Europe

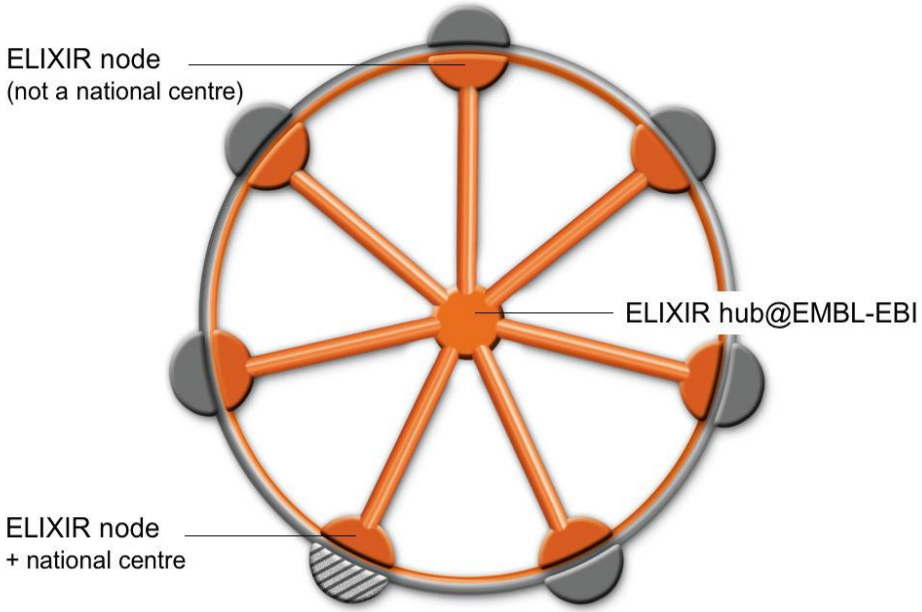
*ELIXIR is also an e-Infrastructure…

ELIXIR: Services and structure

Services:

- Management of Europe's growing volume of biological data
- Services for ESFRI RI in medicine, agriculture and environment (cf. BioMedBridges)
- Biological domain expertise
- Computer Tools Infrastructure
- Computational infrastructure
- Training centres for users of ELIXIR.
- Industry translational services
- 3 million users growing to 10 million in 2020
- Petabytes now growing to exabytes in 2020

Hub and Nodes:



European Bioinformatics Institute

- Outstation of the European Molecular Biology Laboratory
- International organisation created by treaty (cf CERN, ESA)
- 20 year history of service provision and scientific excellence
- Sited at the Wellcome Trust Genome Campus, Hinxton, Cambridge

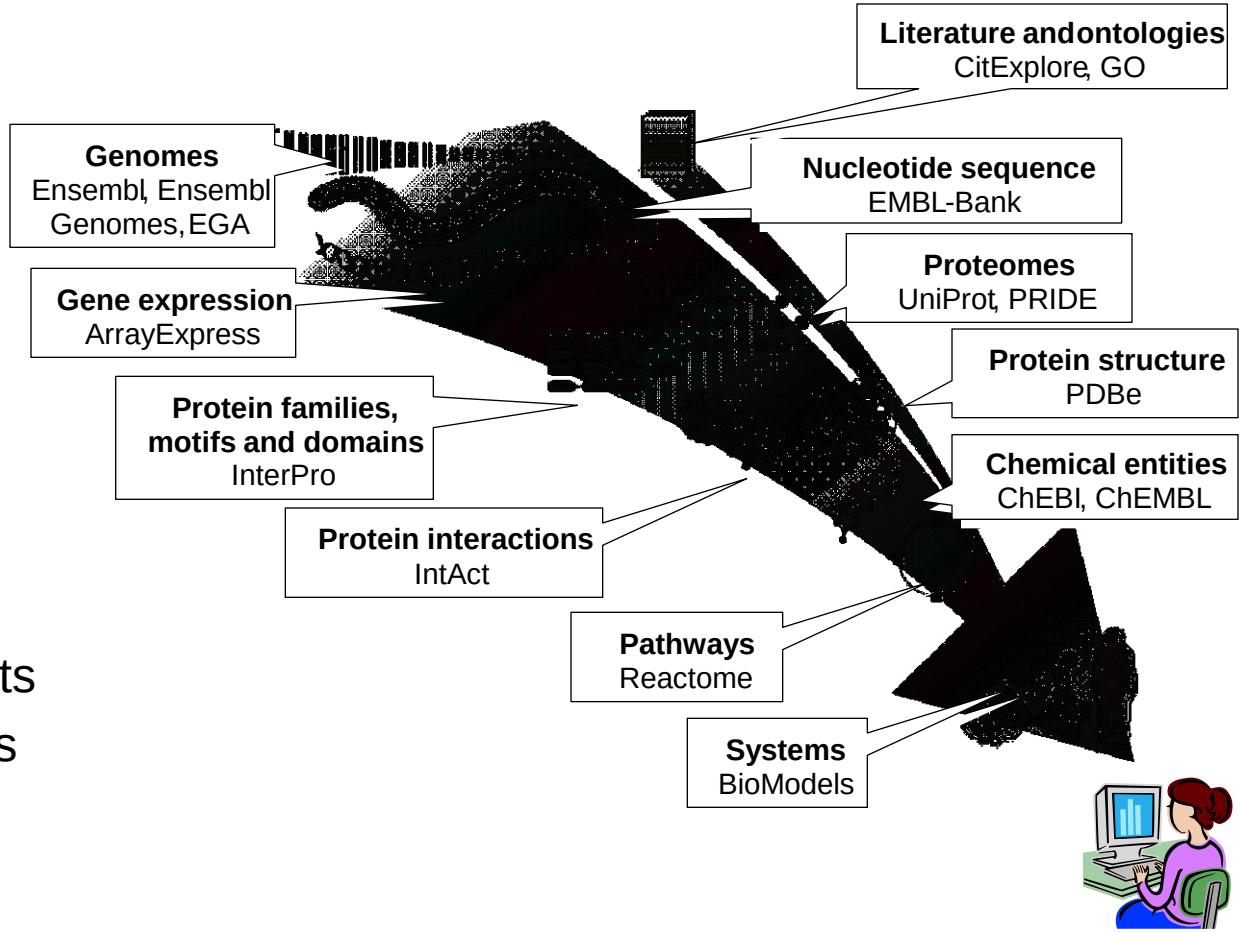
- €40 Million Budget
- > 500 Staff
- Several Million Users
- 15 Petabytes of data
- 10,000 Processors

EMBL-EBI Mission

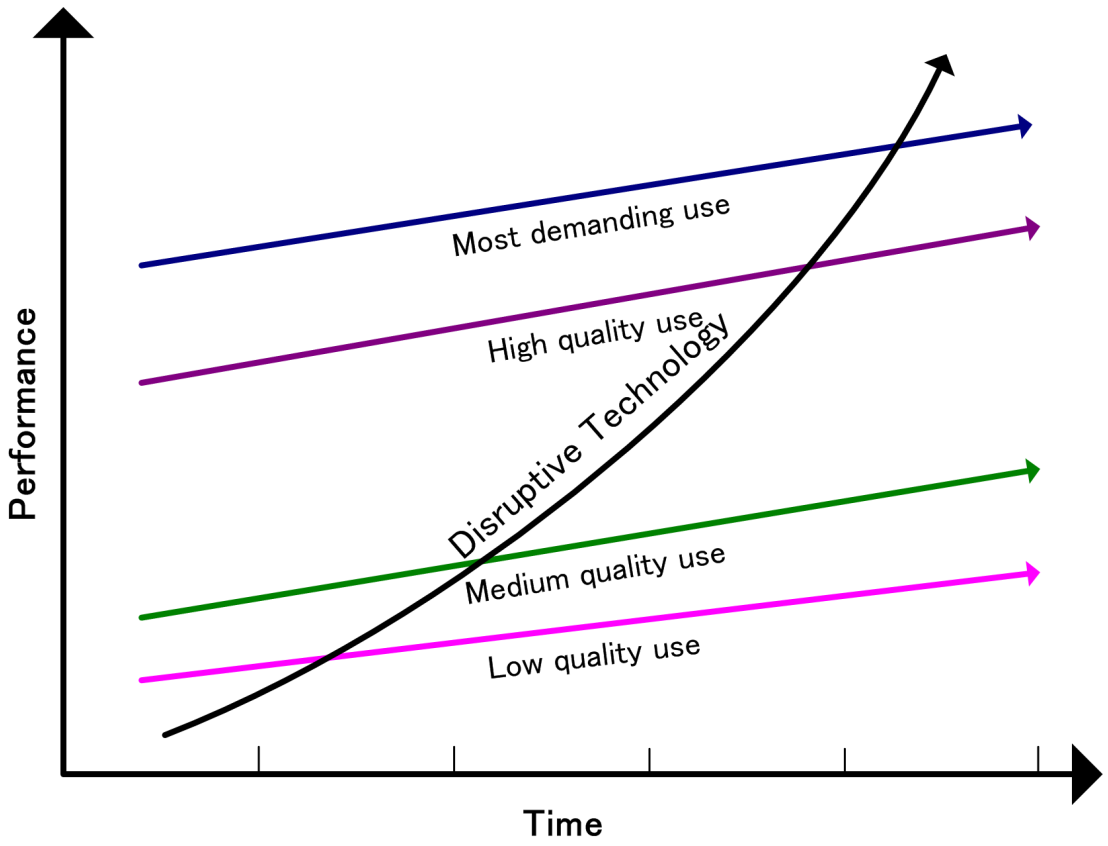
- To provide freely available **data** and bioinformatics **services** to all facets of the scientific community in ways that promote scientific progress
- To contribute to the advancement of biology through basic investigator-driven **research** in bioinformatics
- To provide advanced bioinformatics **training** to scientists at all levels, from PhD students to independent investigators
- To help disseminate cutting-edge technologies to **industry**

Comprehensive, universal, integrated…

- Life sciences
- Medicine
- Agriculture
- Pharmaceuticals
- Biotechnology
- Environment
- Bio-fuels
- Cosmaceuticals
- Neutraceuticals
- Consumer products
- Personal genomes
- Etc…



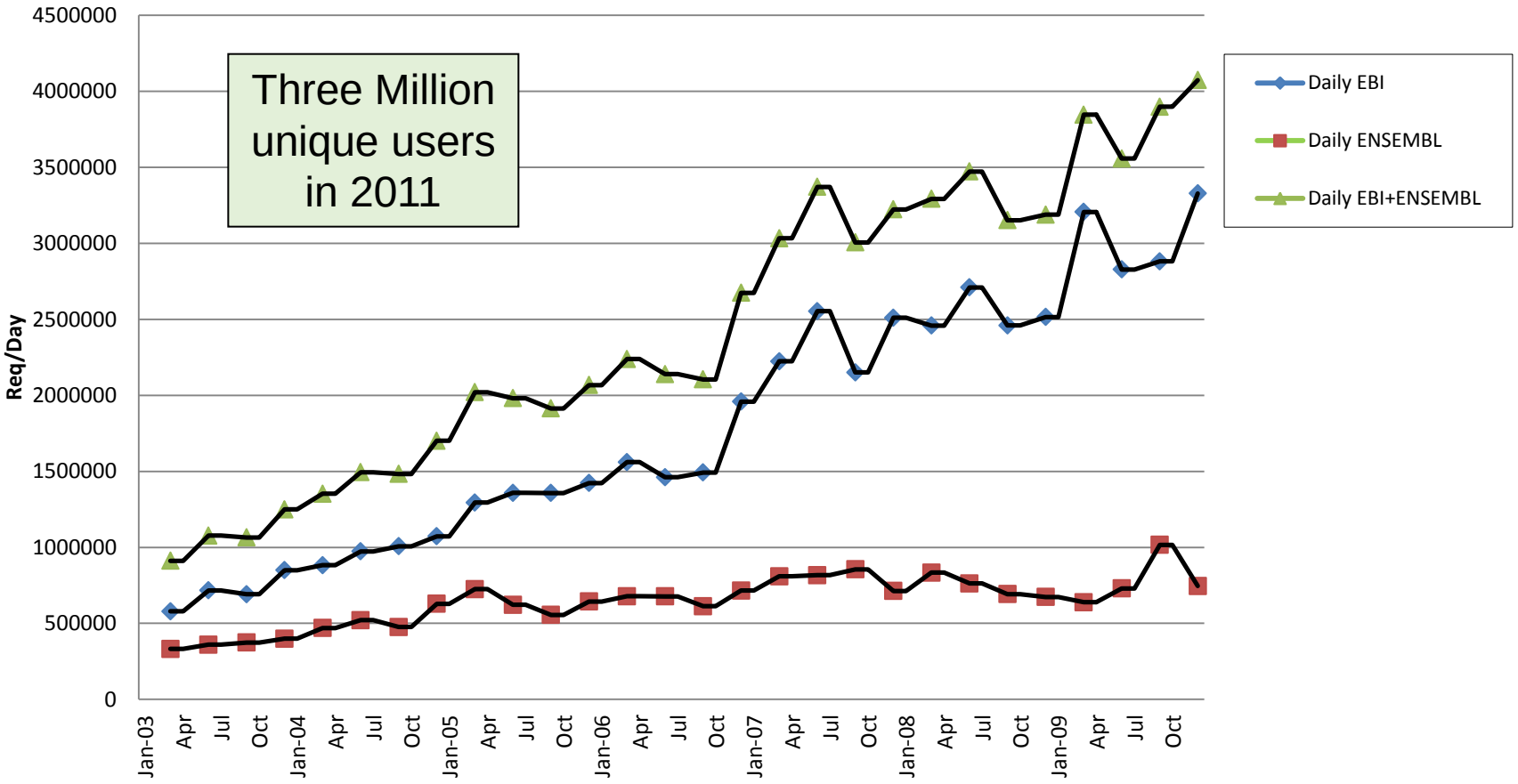
Disruptive technologies



"A technology becomes disruptive when the rate at which it improves exceeds the rate at which users can adapt to the new performance."

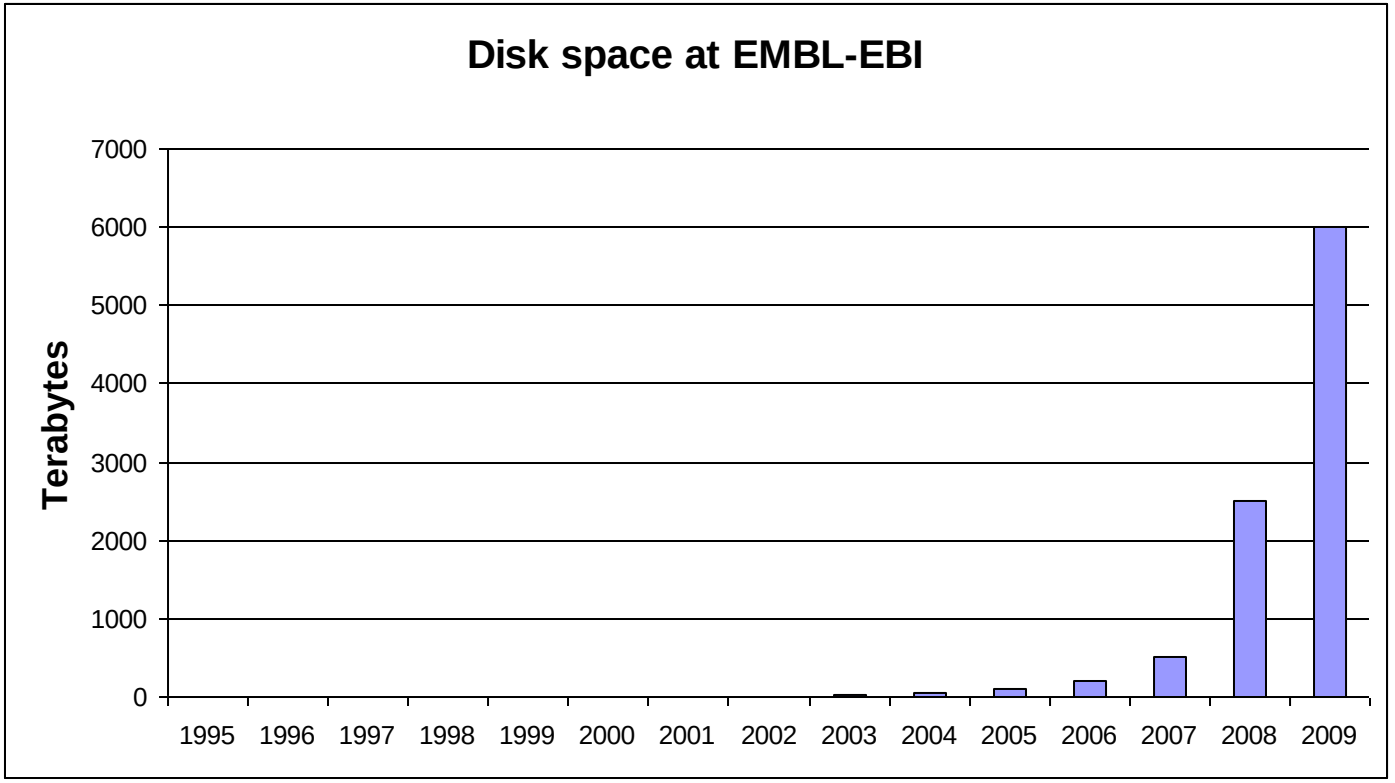
The Innovator's Dilemma. Clayton M. Christensen. Harvard Press. 1997

Very large user community



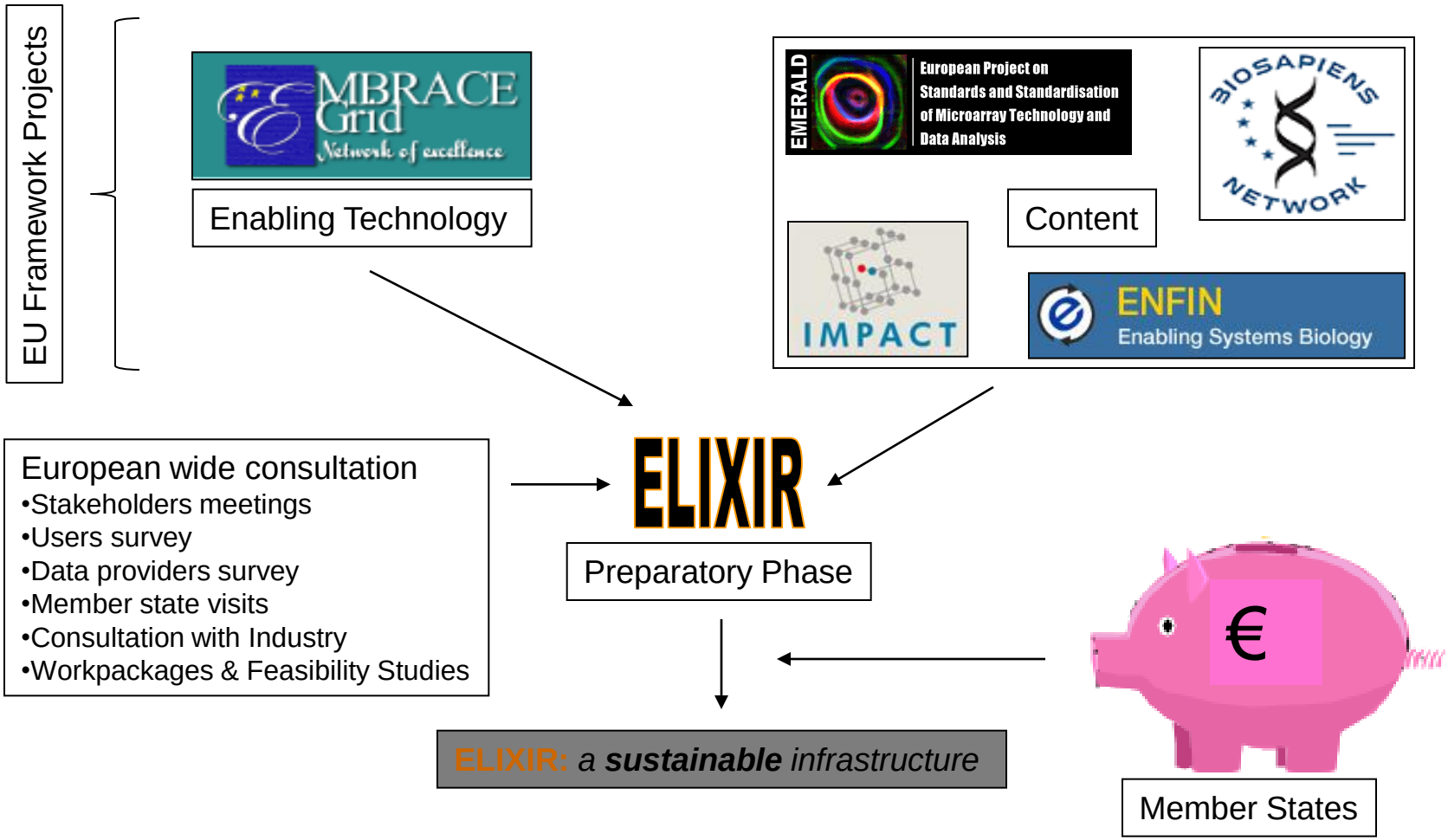
EMBL-EBI

Increasing amounts of data



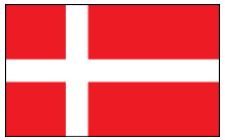
Ten petabytes at May 2010

ELIXIR: Scientific and technical history



ELIXIR: Status Summer 2011

UK: €12 million (ELIXIR Hub)



Denmark: €5 million

Sweden: €1.7 million

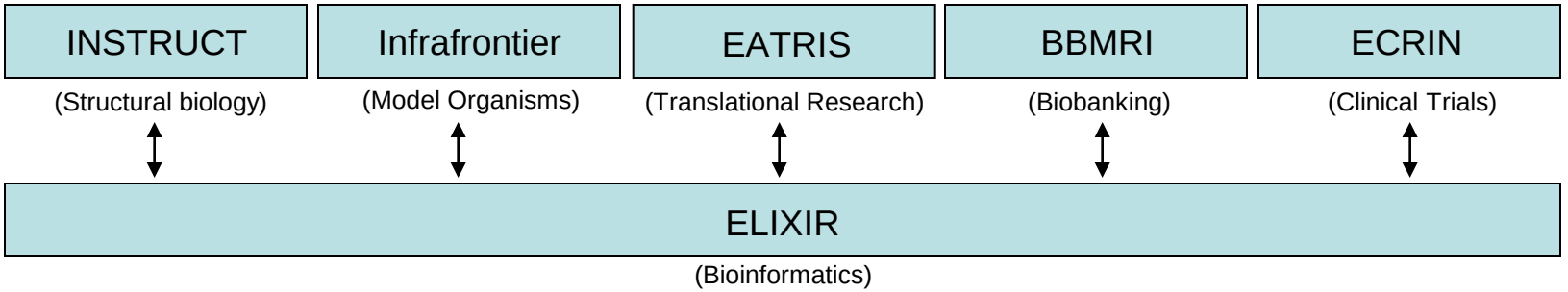
Finland: € 6.85 million (ELIXIR + BBMRI + EATRIS)

Spain: € 1.7 million p. a. for 3 years

- Poland, Norway, Cyprus & Latvia: funding applications in process
- France, Italy, Luxembourg, Ireland: discussions in progress

First Generation ESFRI BMS RI

INSTRUCT	Integrated Structural Biology Infrastructure	€300M	€25M	2007	www.strubi.ox.ac.uk
Infrafrontier	Mouse models of human-disease archive and clinic	€320M	€36M	2007	www.emma.rm.cnr.it
EATRIS	The European Advanced Translational-Research Infrastructure	€255M	€50M	2010	www.eatris.eu
BBMRI	European Biobanking And Biomolecular Resources	€170M	€15M	2009	www.biobanks.eu
ECRIN	Infrastructures For Clinical Trials & Biotherapy	€36M	€5M	2007	www.ecrin.org
ELIXIR	Upgrade Of European Bioinformatics Infrastructure	€550M	€7M	2007	www.ebi.ac.uk
		€1631M	€138M		

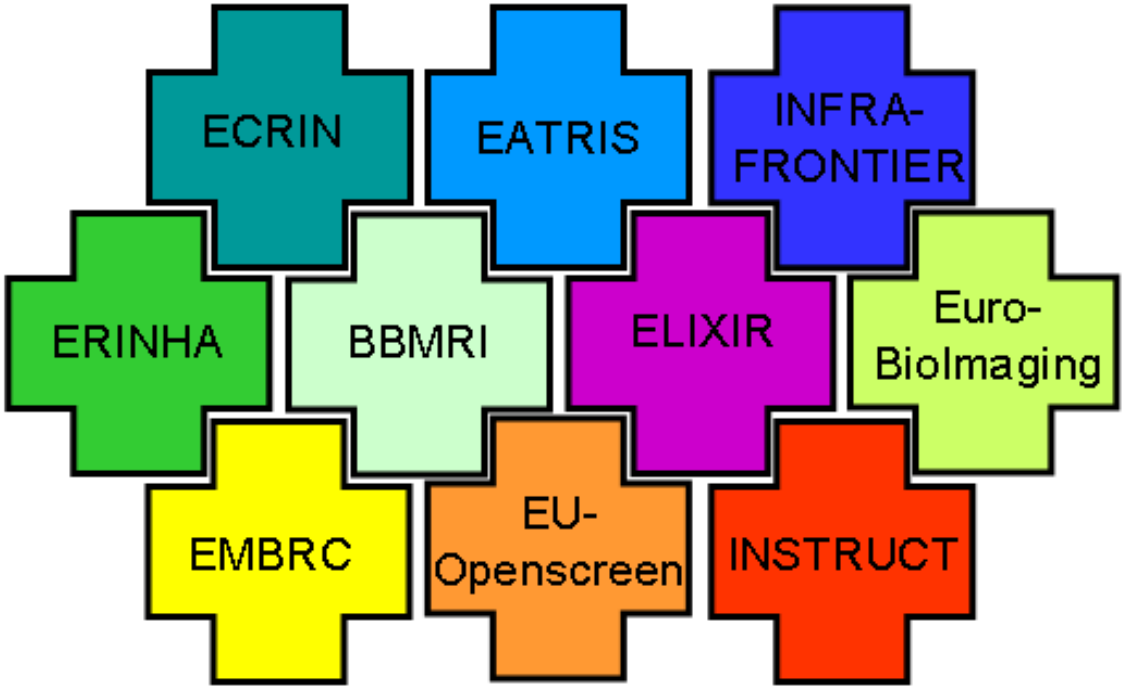


EMBL-EBI

First & Second Generation ESFRI BMS

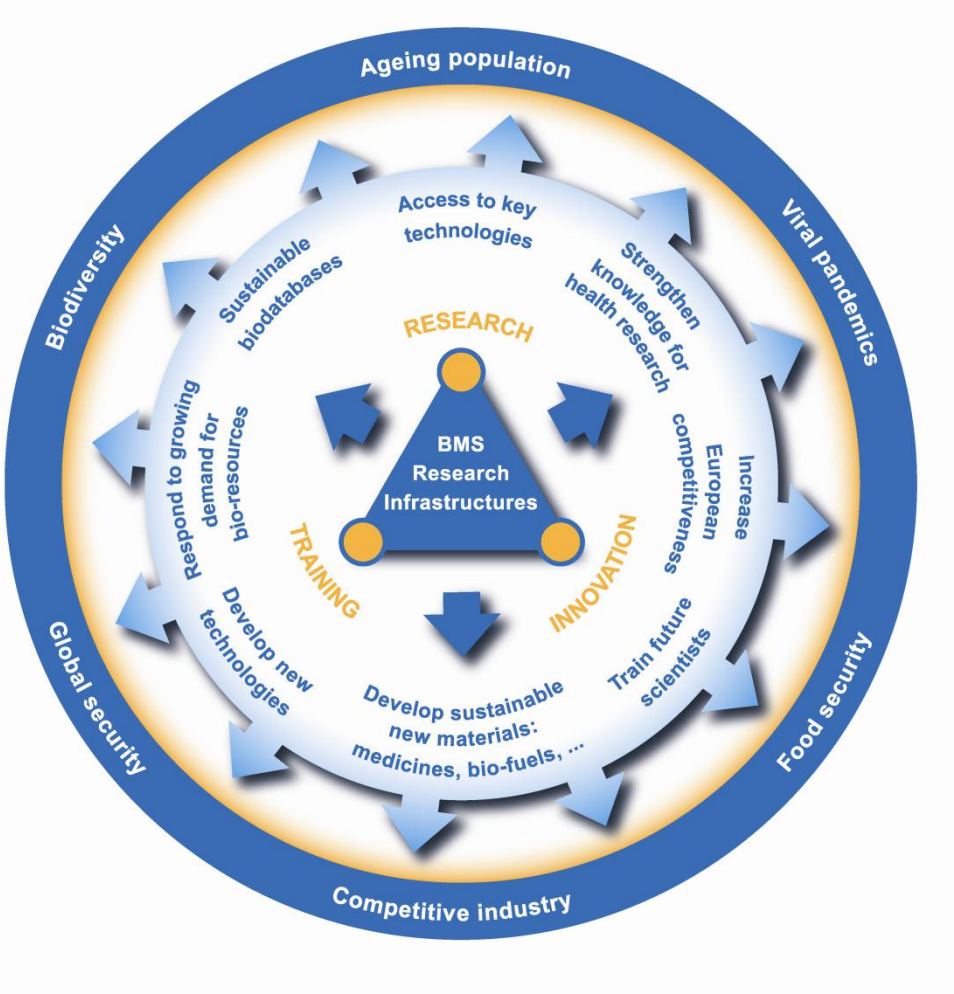
BBMRI	Biobanking and Biomolecular Resources Research Infrastructure
EATRIS	European Advanced Translational Research Infrastructure in Medicine
ECRIN	European Clinical Research Infrastructures Network
ELIXIR	European Life Science Infrastructure for Biological Information
EMBRC	European Marine Biological Resource Centre
EU-OPENSCREEN	European Infrastructure of Open Screening Platforms for Chemical Biology
Euro-BioImaging	European Biomedical Imaging Infrastructure
ERINHA	European Research Infrastructure on Highly Pathogenic Agents
Infrafrontier	European Infrastructure for Phenotyping and Archiving of Model Mammalian Genomes
INSTRUCT	An Integrated Structural Biology Infrastructure for Europe

The ESFRI BMS RI depend on each other



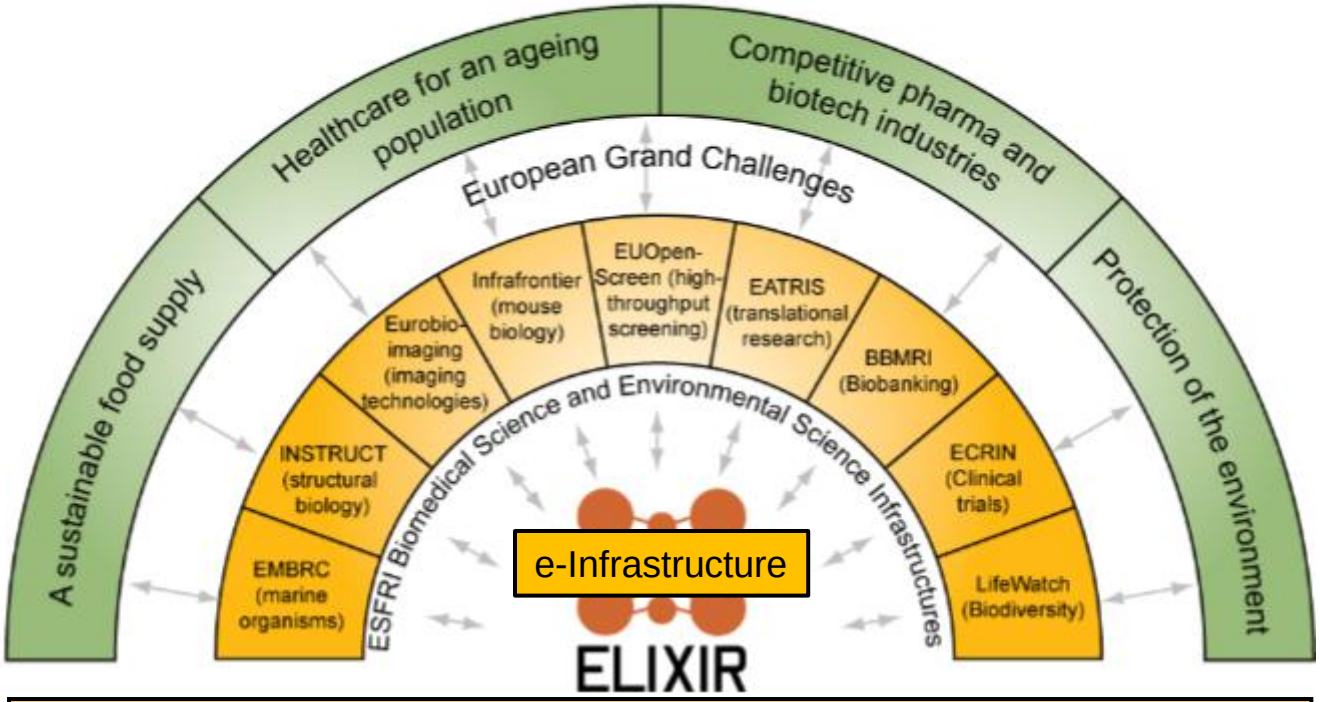
ESFRI BMS RI and Europe 2020

- Europe 2020 is the name of the EU strategy for economic growth
- Demographics, a sustainable food supply and environmental protection have been identified as particular challenges
- Emerging pandemics and bioterrorism also present threats
- European industry is facing unprecedented competition from emerging economies
- ICT and life-sciences research have been identified as providing particular opportunities for growth
- The ESFRI BMS RI will enable solutions to these challenges



EMBL-EBI

ELIXIR is an e-Infrastructure



GÉANT, DANTE, EGI.eu, PRACE, etc

BioMedBridges Objectives

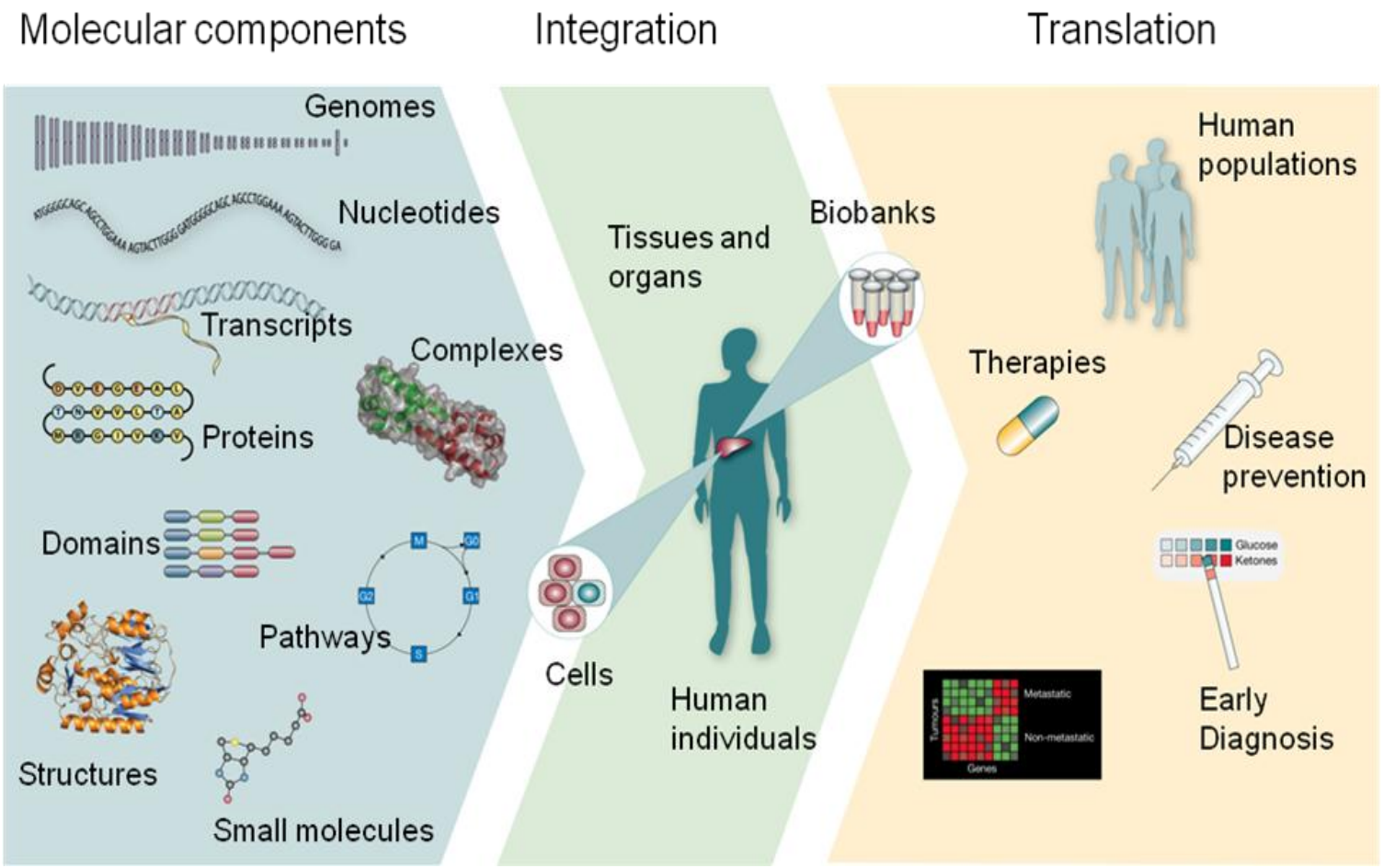
- BioMedBridges will construct the e-infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains.

- It will provide the computational ‘data and service’ bridges between the individual biological and medical sciences (BMS) research infrastructures (RIs), clustering them together and linking the basic biological research and data to the clinical research and associated data.

Building bridges that cross different…

- spatial scales
  - from molecules through cells and organs to humans and the environment
- species
  - from bacteria, through model organisms (eg mouse) to man
- temporal scales
  - from nanoseconds of molecular motions, to seconds of a heartbeat, to years of a human life and the aeons of evolution
- technologies and the heterogeneous data they generate
  - from the nanotechnology of sequencing through the spectroscopy of cellular and whole organism imaging to synchrotrons for structure determination
- research communities who have no tradition of working together
  - from basic scientists to clinicians and environmentalists.

From Molecules to Medicine…



EMBL-EBI

BioMedBridges: Data protection

- Access to much of the data in the ESFRI BMS domains has Ethical, Legal or Societal Implications (ELSI)
- It includes Personally Identifiable Information (PII)
- Working with these across organisational boundaries
- Working with these across national boundaries
- Linking these to other data such as scientific data (cf. reidentification, inferential attack)

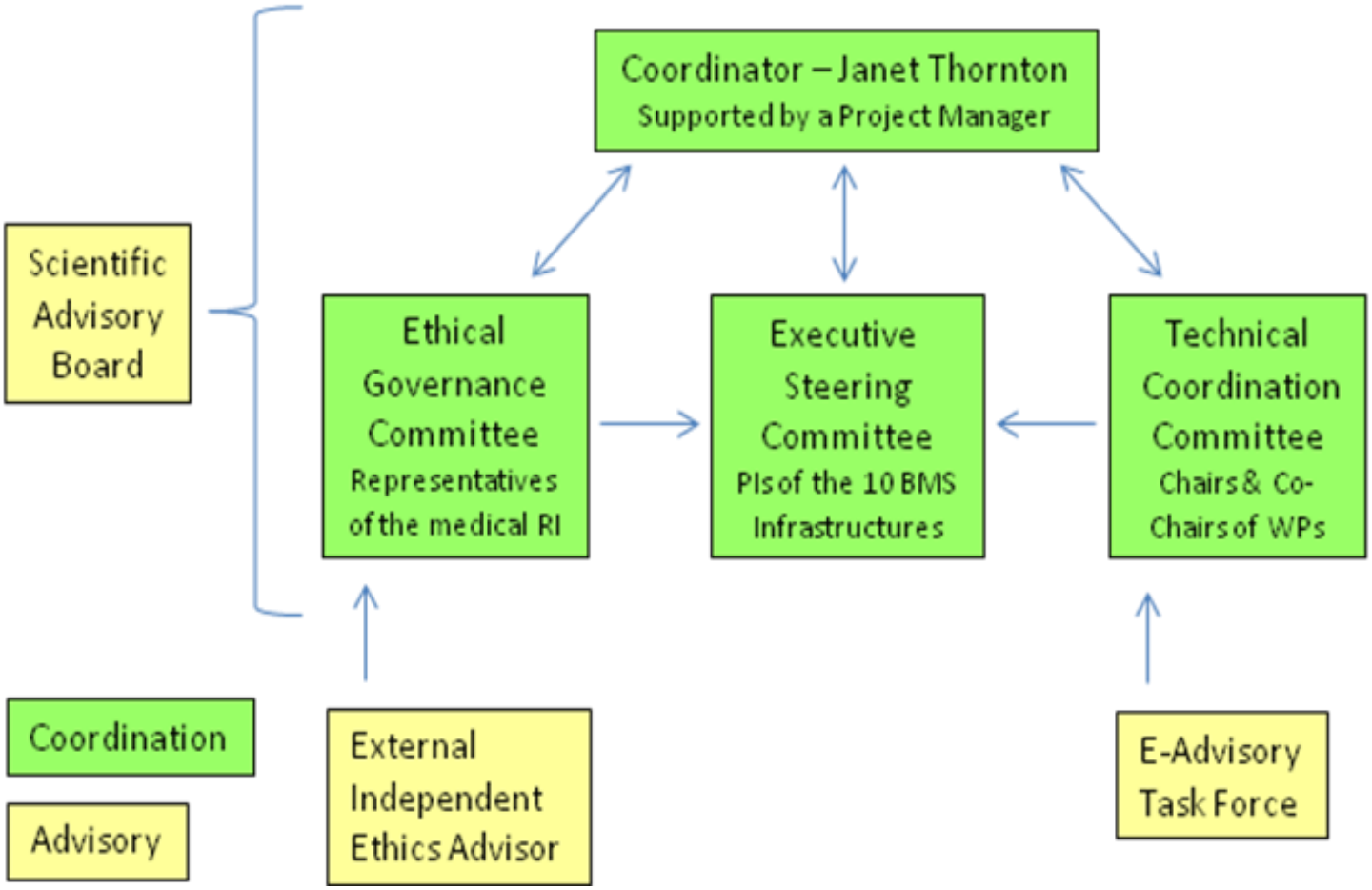
ELSI data in the life-sciences

- Clinical research
- Translational research
- Biological sample management
- Publishing, attribution and intellectual property
- Environmental protection
- Others...

ELSI data and the ESFRI BMS & ENV RI

- Clinical research
  - **ECRIN**, EuroBioImage, EU-OpenScreen, EHRINA, ELIXIR
- Translational research
  - **EATRIS**, EuroBioImage, EU-OpenScreen, EHRINA, ELIXIR, InfraFrontier
- Biological sample management
  - **BBMRI** and all the others.
- Publishing, attribution and intellectual property
  - All of them
- Environmental protection
  - **LifeWatch**, BBMRI, ELIXIR

BioMedBridges Management



EMBL-EBI

BioMedBridges Work Packages

- WP1 Management
- WP2 Outreach and inreach
- WP3 ESFRI BMS Standards Description and Harmonization
- WP4 Technical integration
- WP5 Secure access
- Five Use Cases WP6 – WP12
  - WP6 Interoperability of large scale image data sets from different biological scales
  - WP7 PhenoBridge - crossing the species bridge between mouse and human
  - WP8 Personalized Medicine - integrating complex data sets to understand disease pathogenesis and improve biomarker and treatment selection
  - WP9 From cells to molecules - integrating structural data
  - WP10 Integrating disease related data and terminology from samples of different types
- WP11 Technology Watch
- WP12 Training

WP3: Standards description & harmonization

- It will only be possible to exchange and link data between the different ESFRI BMS domains if they use common identifiers, and harmonised content, syntax and semantics. In order to achieve this, this use case will create:
  - An ESFRI BMS Online Dictionary of common molecular identifiers
  - A mapping and registry of ESFRI BMS standards
  - A report reviewing national sample safety regulations and an XML schema supporting safety assessment
  - An ESFRI BMS Meta Service Registry
  - An harmonisation semantic standards

WP4: Technical integration

- Implement shared standards from WP3 to allow for integration across the BioMedBridges project
- Expose the integration via use of REST based Web-Services interfaces optimised for browsing
- Expose the integration via use of REST based Web-Services interfaces optimised for programmatic access
- Expose appropriate meta-data information via use of Semantic Web Technologies
- Pilot the use of semantic web technologies in high-data scale biological environments

WP5: Secure access

Create a security framework that will address the ethical, legal and regulatory issues resulting from sharing data and providing access to biomaterials in order to ensure that the infrastructure components developed are compliant with national and European regulations, privacy rules and access requirements.

1. Document the regulations and the privacy and security requirements including intellectual property rights that must be observed when accessing and sharing data and bio-samples
2. Create a tool for assessing regulatory and ethical requirements
3. Define the security architecture and framework, based on security requirements and risks identified
4. Implement the security framework

WP6: Use case 1

Title: “Interoperability of large scale image data sets from different biological scales”

- The use case will demonstrate the utility of the interoperability of large scale image data sets from different biological scales (cell – tissue – organism) to enable drug target and biomarker discovery for human disease with cancer as an example.

WP7: Use case 2

Title: “PhenoBridge – crossing the species bridge between mouse and human”

- The mouse is an important model organism for human disease. This use case will harmonise ontological descriptions of phenotype in both mouse and human using diabetes & obesity as examples:
  1. Identify and develop annotations, terminologies, and mappings between terminologies for human and mouse models of diabetes and obesity
  2. Identify and group related interacting parameters in human and mouse which determine complex clinical and molecular phenotypes
  3. Formalise rules for phenotypic annotation in human and mouse to work towards automation of phenotypic discovery
  4. Deploy a service which builds on the rules, terminologies and annotations generated in this use case at the EMBL-EBI

WP8: Use case 3

Title: “Personalized Medicine – integrating complex data sets to understand disease pathogenesis and improve biomarker and treatment selection”

- Personalized medicine (PM) is starting to have a major impact on the treatment of cancer but there is as yet no systematic effort to make PM data available to scientists and clinicians in an ethical, robust and sustainable manner. This use case will:
  1. Develop a process for sharing & access PM data in a secure and ethical manner
  2. Define types of PM data being generate by the different ESFRI Projects
  3. Develop mechanisms of interoperability for PM data types

WP9: Use case 4

Title: “From cells to molecules – integrating structural data”

- Modern structural biology is producing data about biological molecules at a far wider range of resolutions that previously. If the value of these data is to be maximised then tools are needed that will allow data at multiple resolutions to be deployed against important biological and medical problems. This use case will develop software, database and web-based services to do this using the components created by the other work packages.

WP10: Use case 5

Title: “Integrating disease related data and terminology from samples of different types”

- Modern clinical research needs large collections of high-quality well-documented samples from humans and model organisms. This use case will
  - Mapping between data elements in EMBL-EBI’s BioSample Database and selected resources from BBMRI
  - A prototype linking ICD10/SNOMED CT concepts to Ensembl gene identifiers
  - A prototype federated query interface.

BioMedBridges WP11: Technology Watch

- Comprises representatives of GÉANT, DANTE, EGI.eu, PRACE & CERN as well as technical experts from the ESFRI BMS RIs
  - Bring together the technical experts of the BioMedBridges partners European ICT & e-Infrastructures to monitor and report on developments and provide advice to the project
  - Facilitate adoption of e-Infrastructure technologies by the BioMedBridges Work Packages and the ESFRI BMS RI
  - Communicate advice from the ICT Infrastructures and the e-Infrastructures to the BioMedBridges partners
  - Will produce annual reports on status of e-Infrastructures relevant to the progress of BioMedBridges including requirements for use cases & recommendations for adoption of new technologies and standards.

BioMedBridges starts 1 January 2012…

Thank you for your attention…