

ECRI 2007[★]

★ *european conference on research infrastructures*

ECRI 2007[★]

★ *european conference on research infrastructures*

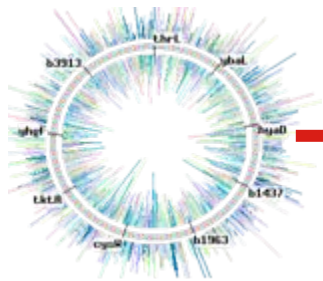
Scientific Repositories and their Management: Constructing a sustainable infrastructure for biological information in Europe

Janet Thornton

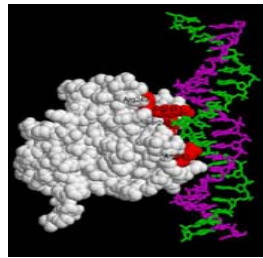
European Bioinformatics Institute

ECRI 2007, Hamburg, Date

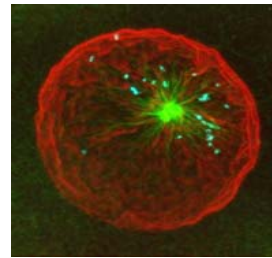
A European Infrastructure for Biological Information ELIXIR



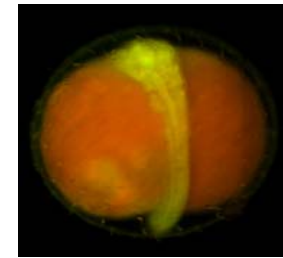
Genome



Protein



Cell



Embryo



Fruitfly



Mouse



Human Development,
Ageing, Disease

ELIXIR Mission

To construct and operate a sustainable infrastructure for biological information in Europe, to support life science research and its translation to medicine and the environment, the bio-industries and society.

This will contribute to improvements in all human endeavour associated with living systems including:

health and medicine

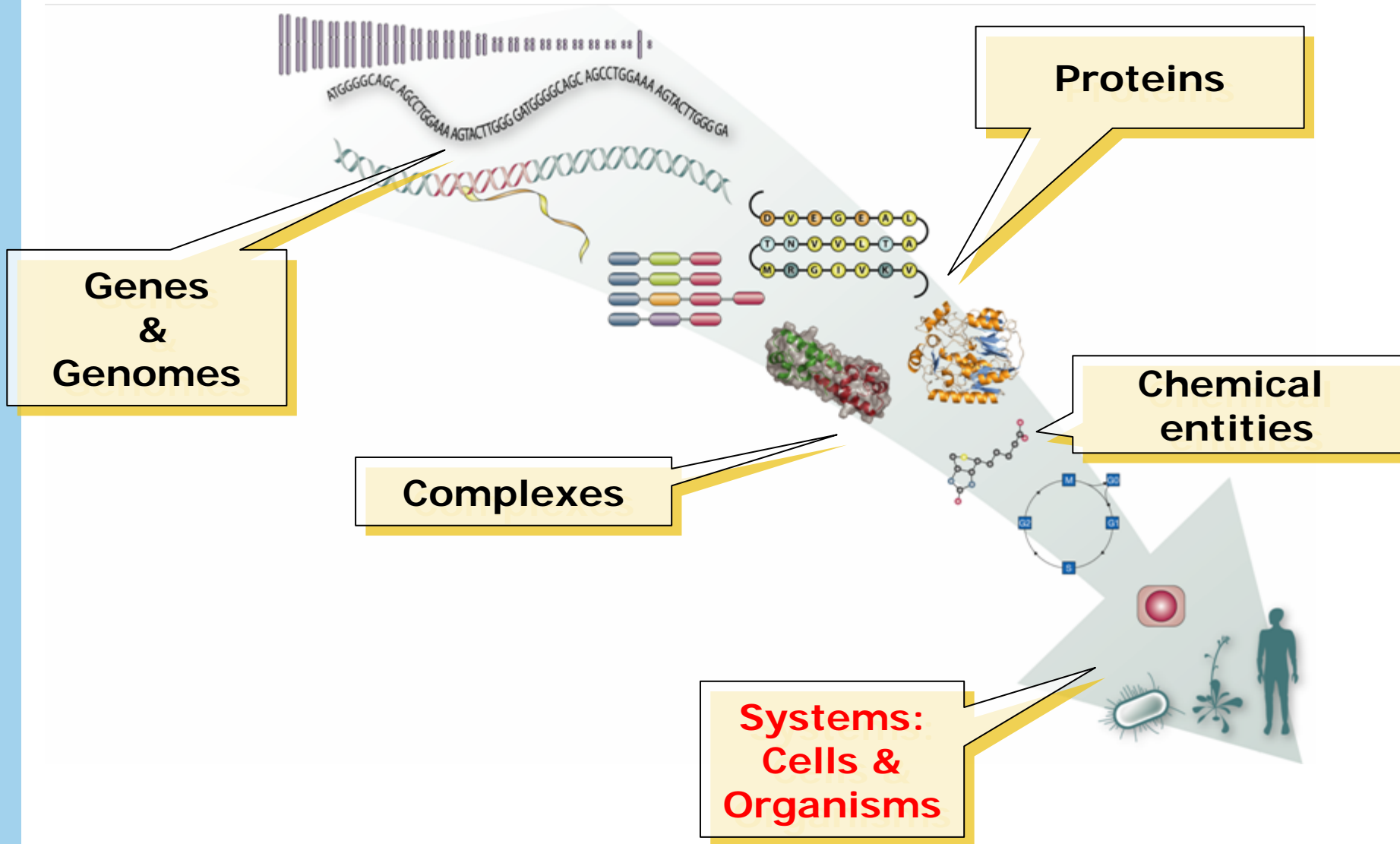
the environment

Agriculture

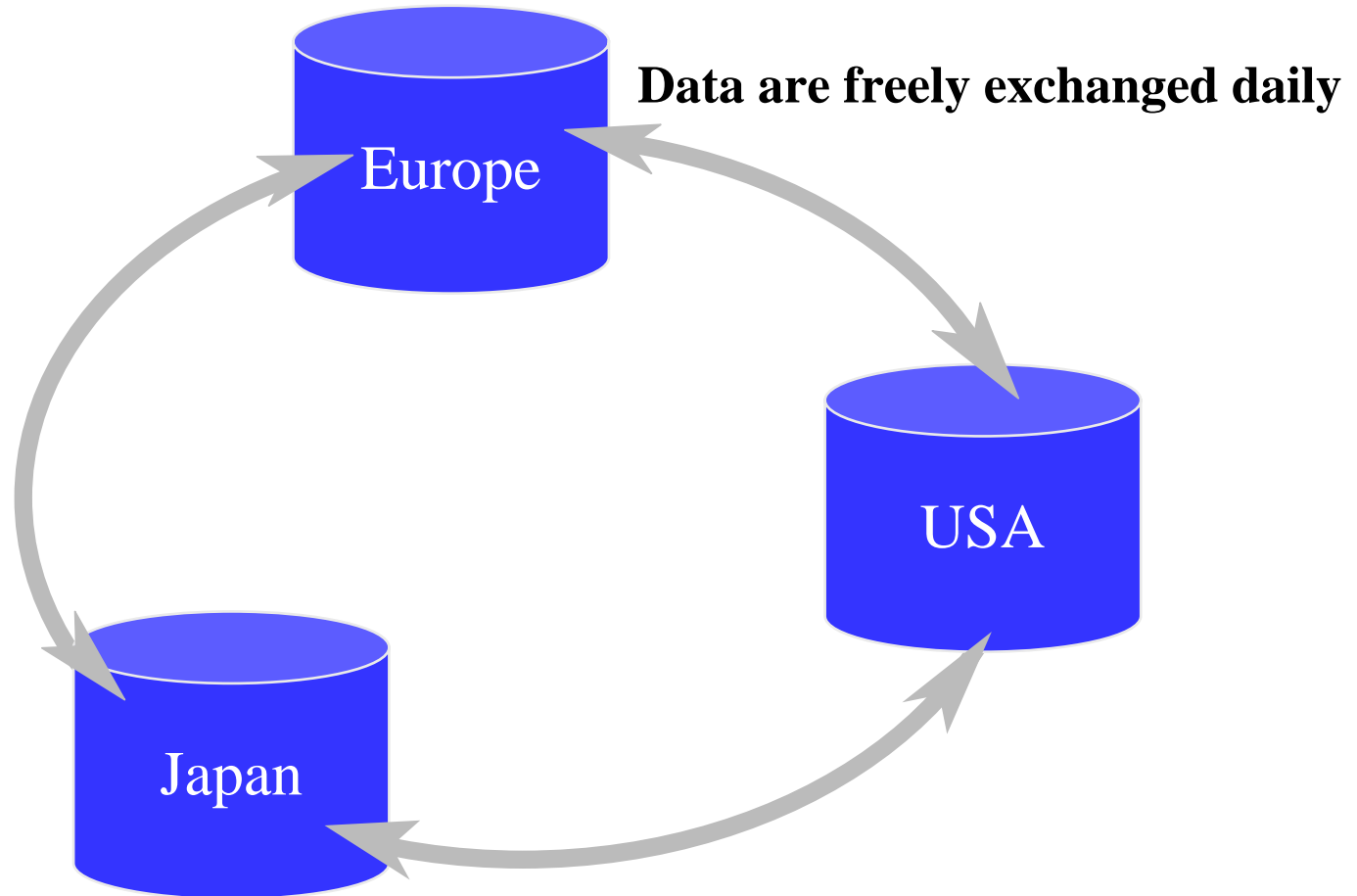
Fisheries

Forestry

Biotechnology

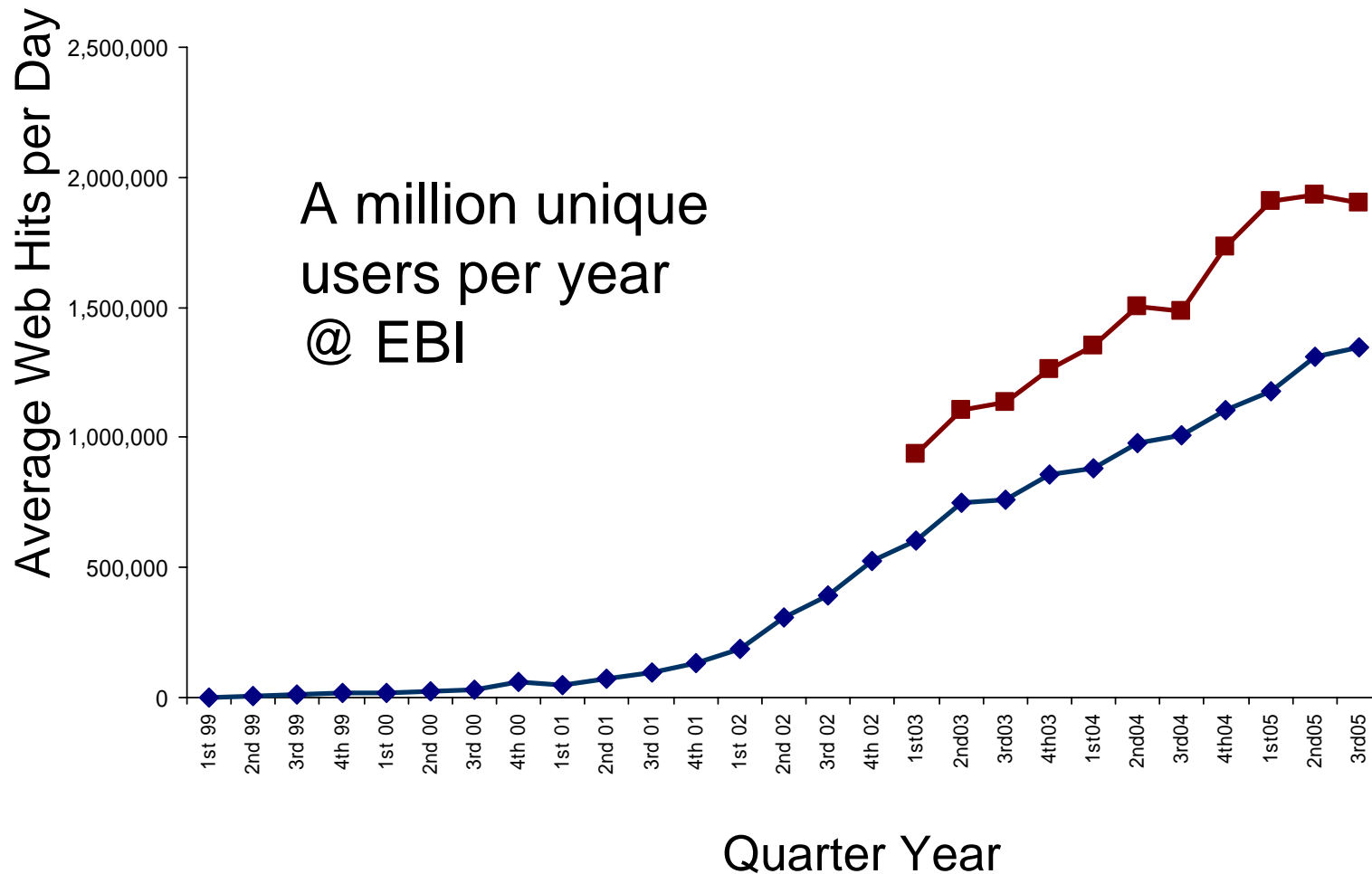


Data are freely deposited

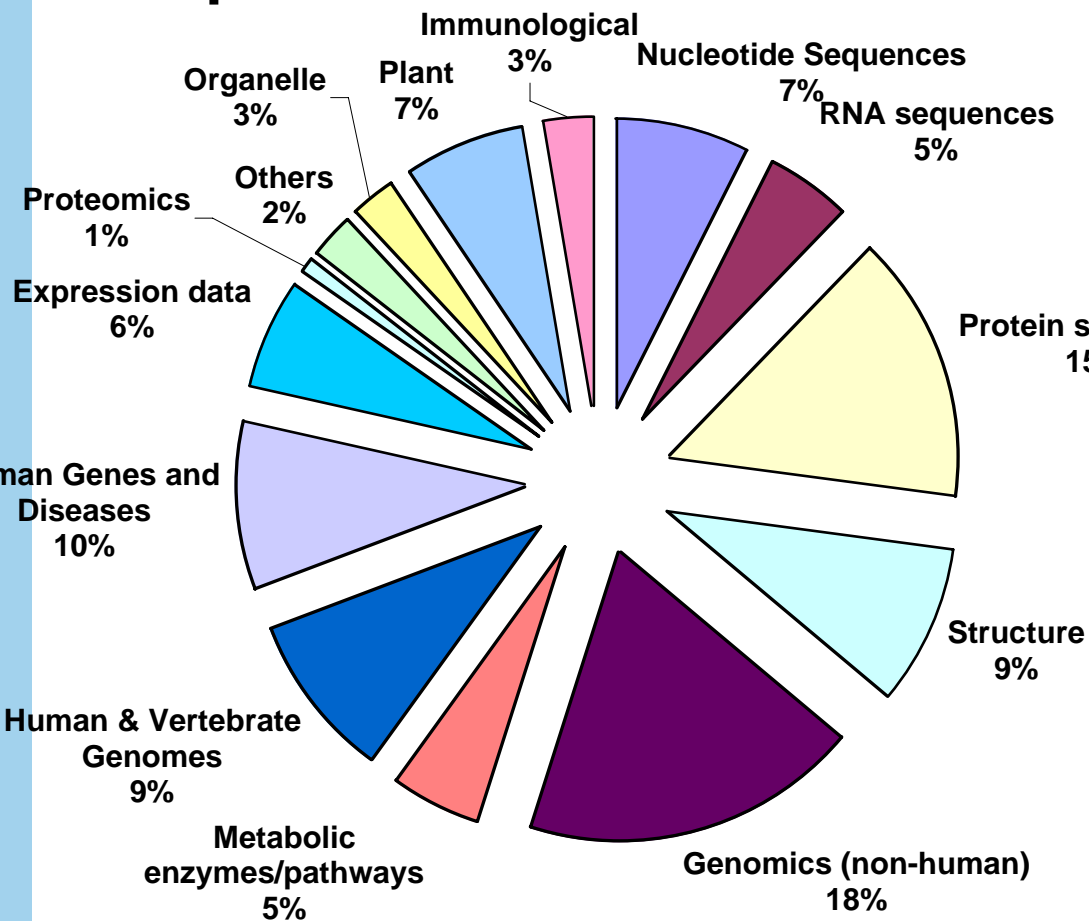


Data are made freely available to all

Large numbers of users of data resources



Specialised Molecular Data Resources



More than 700
in total

30% in Europe

(All use core
resources as
reference data)

Core databases

They are universally relevant to biomolecular science.

They have a huge user community.

They aim to be complete collections.

Completeness is assured by exchange agreements with other data centres world-wide (typically the USA, Japan and Europe at present).

The science they represent is stable enough to allow standardisation of the data structure.

Standards where available are followed.

They are actively involved in relevant standard development.

Journals insist on data deposition in these databases.

Specialist Data Resources

Data are more specialised (e.g., one species or family) and do not aim to be comprehensive.

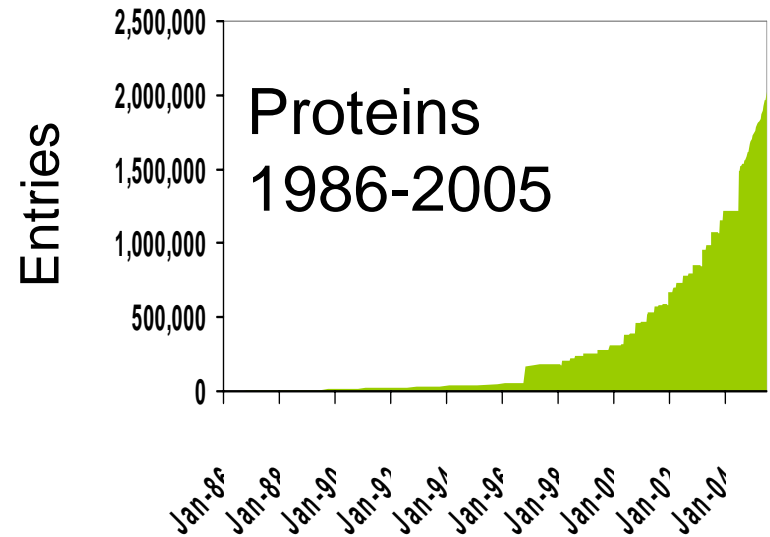
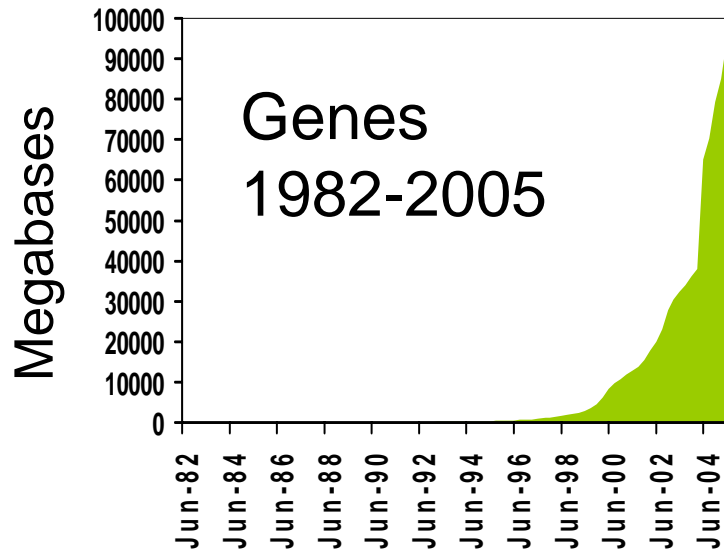
They are investigator-led products of research groups with content which reflects the research interests of their provider.

Many are derivative or ‘summarising’ databases which combine and organise data from a range of other databases.

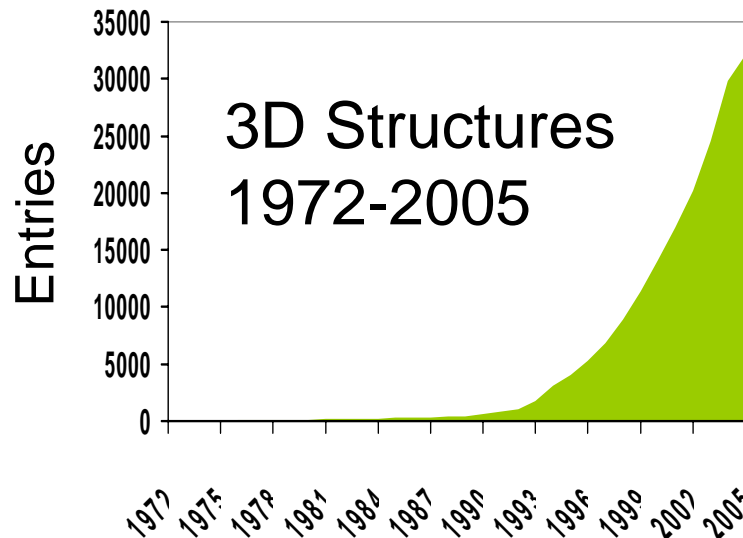
Most offer a more limited service, and may be less stable and designed only for experts.

Some may be candidates for core

All Biomolecular Data Resources are growing rapidly



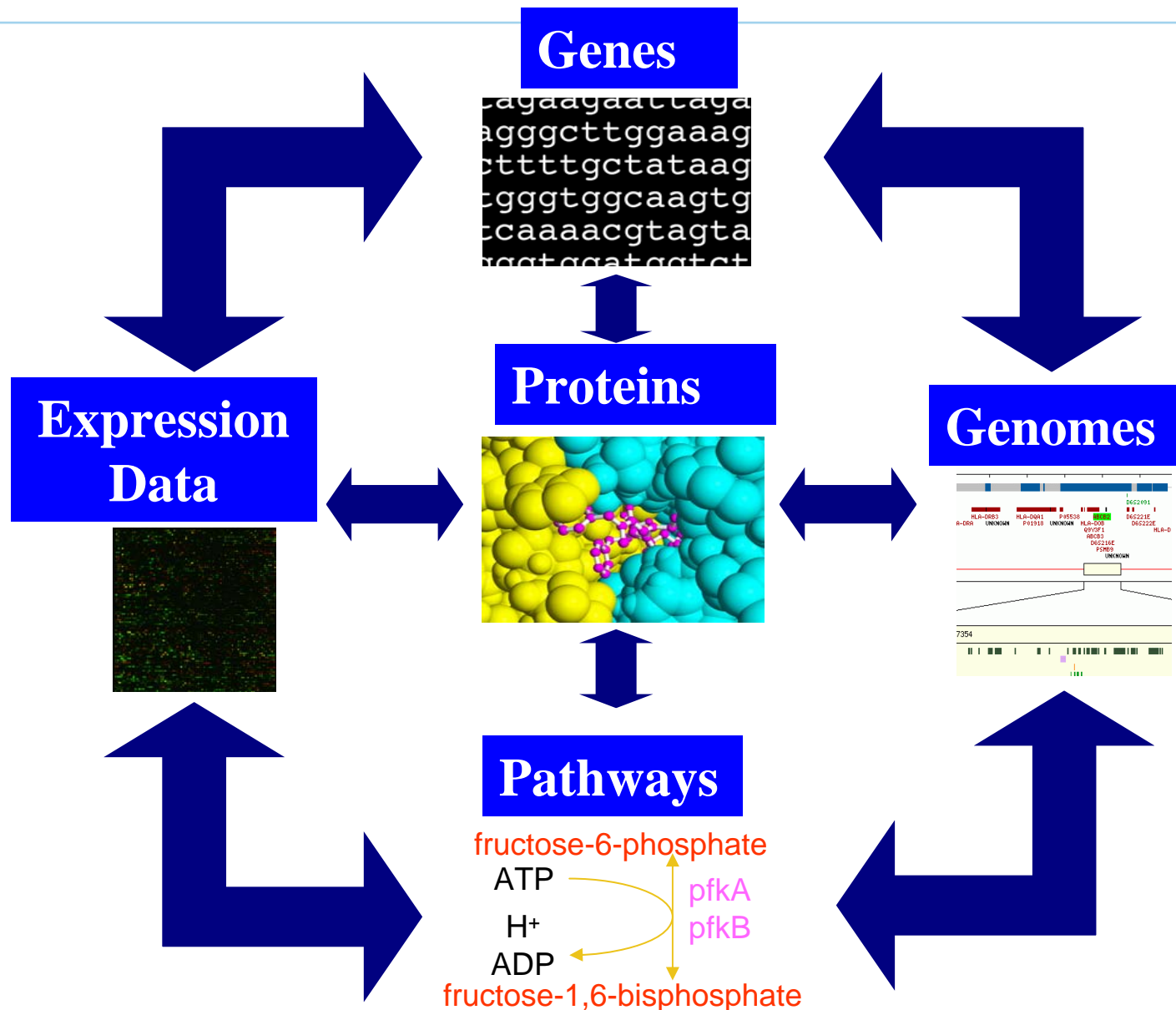
**> 6 million
protein
sequences
recently
deposited!**



LINKS to LITERATURE

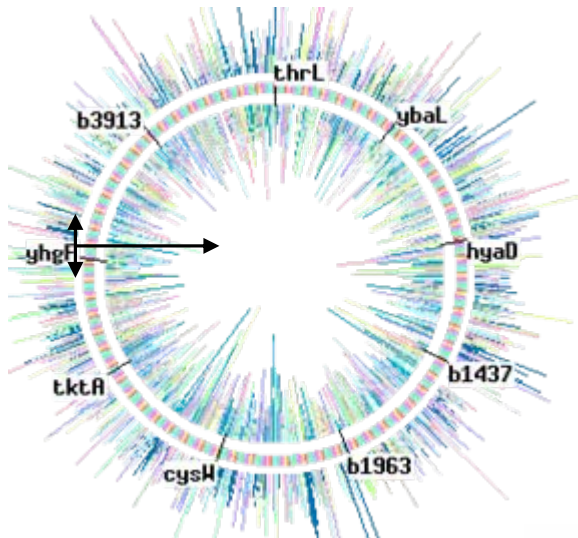
Literature	Articles
Chinese Biological Abstracts	132,763
CiteSeer	493,423
Patents	1,146,819
PubMed/Medline	16,731,346

Data Integration is vital

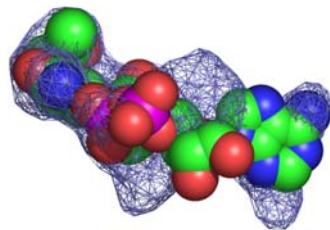
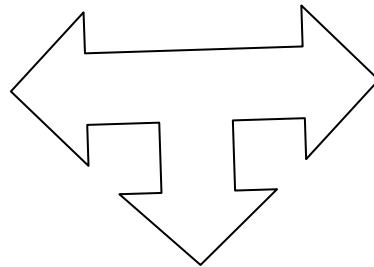


Integration with Other Data is increasingly important

e.g. Linking from Molecules to Medicine & Agriculture



Genomes



Metabolites

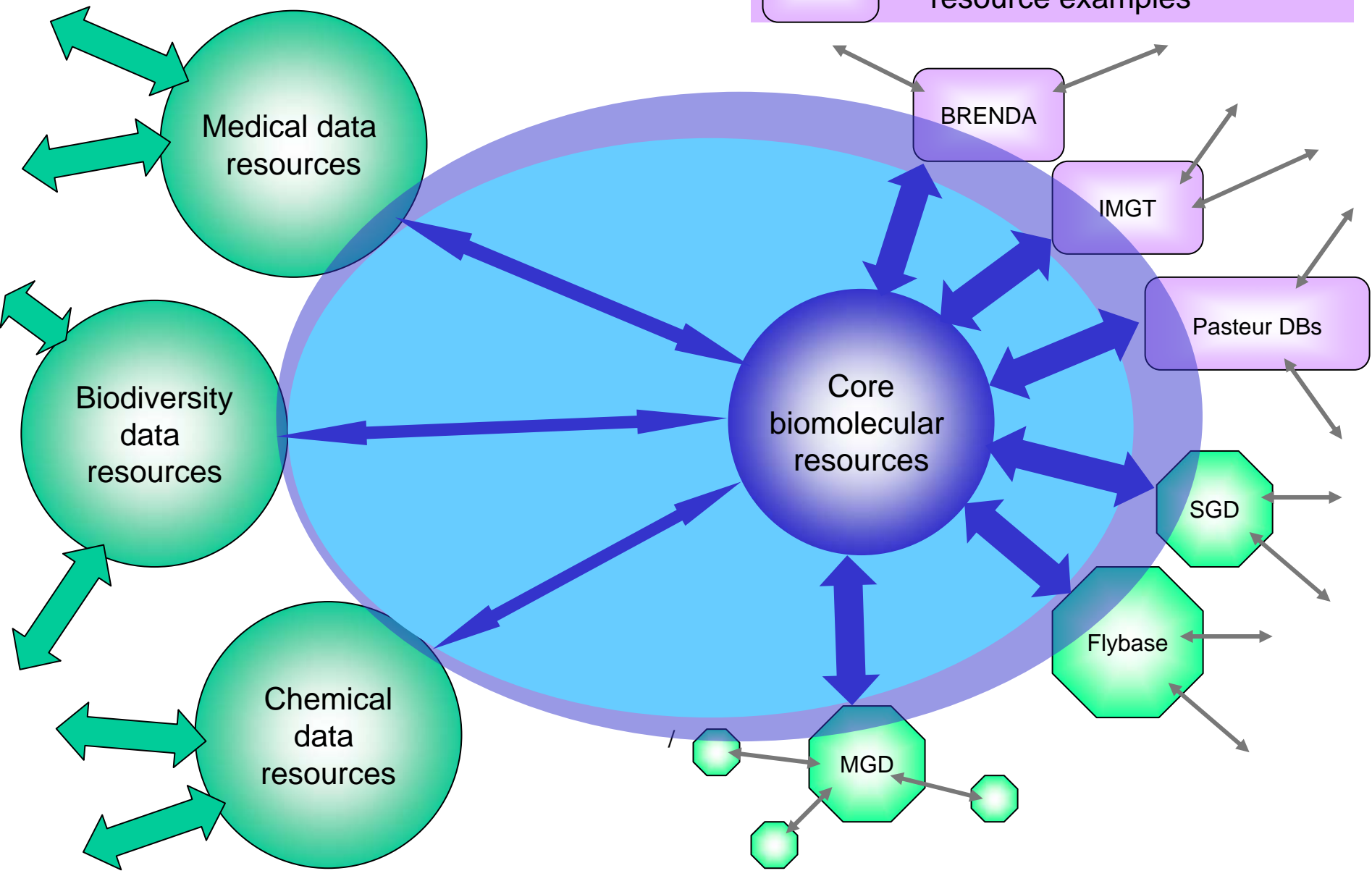


Proteins



Large resources in related disciplines

Specialist biomolecular data resource examples



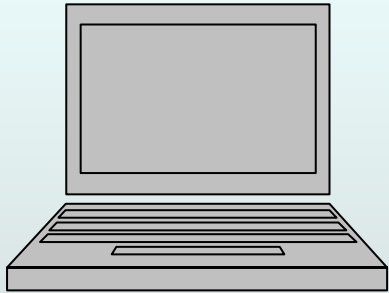
Model organism resource examples

EMBRACE EU Network of Excellence

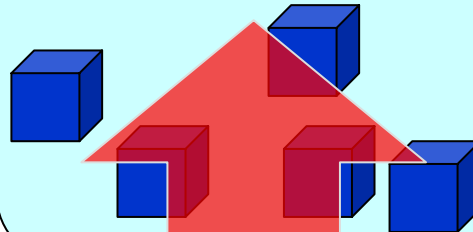
Making tools available easily to all

Web Services

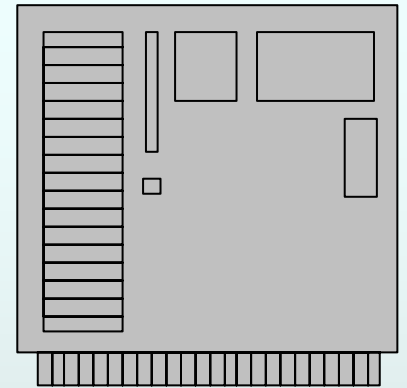
Test problems



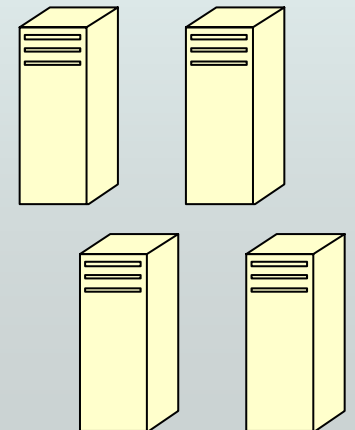
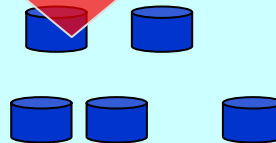
Tools



Compute Power

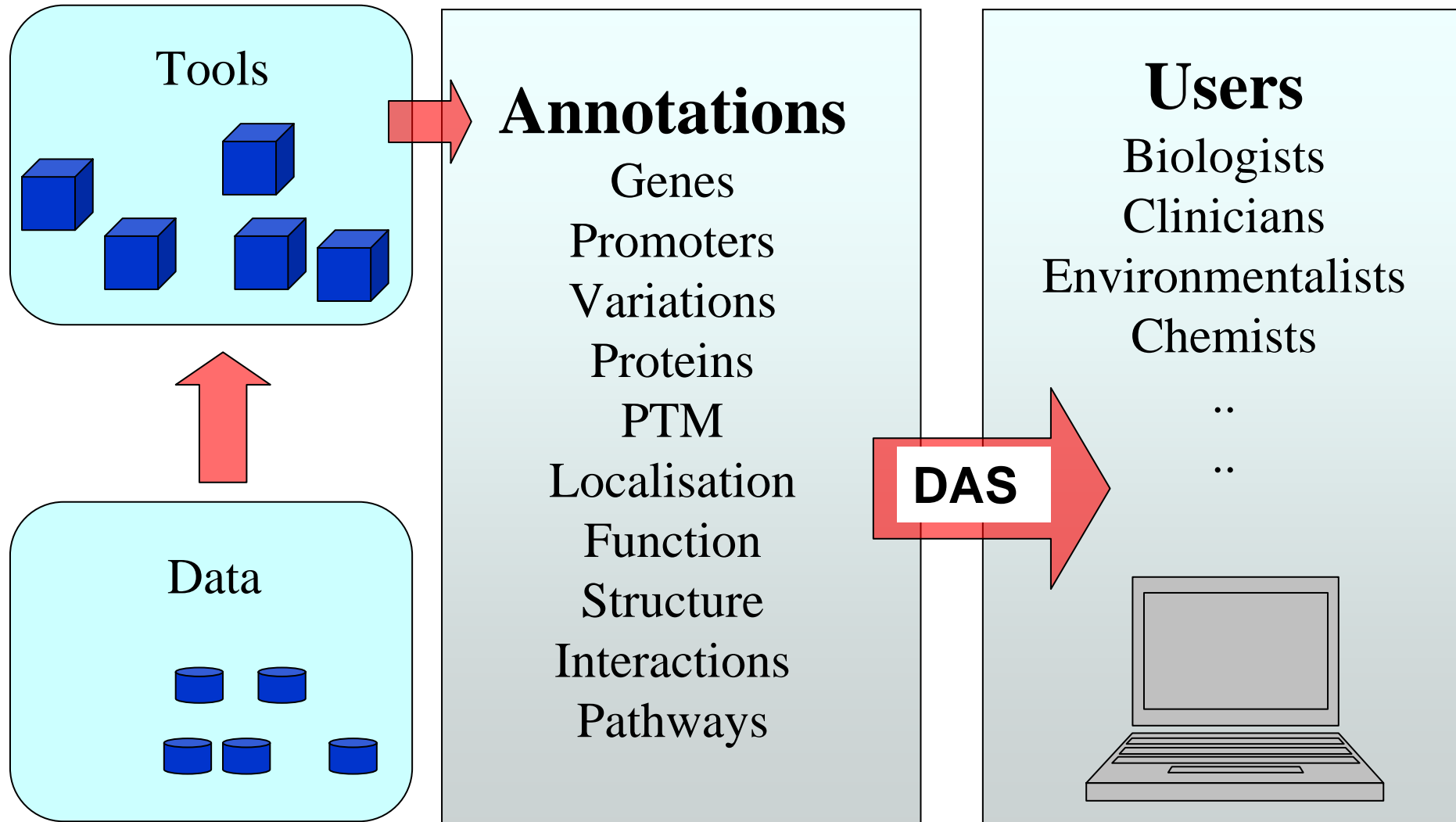


Data



BioSapiens EU Network of Excellence

Making annotations available



Rationale for ELIXIR

Optimal Data Management

Coordinated Data Resources with improved access

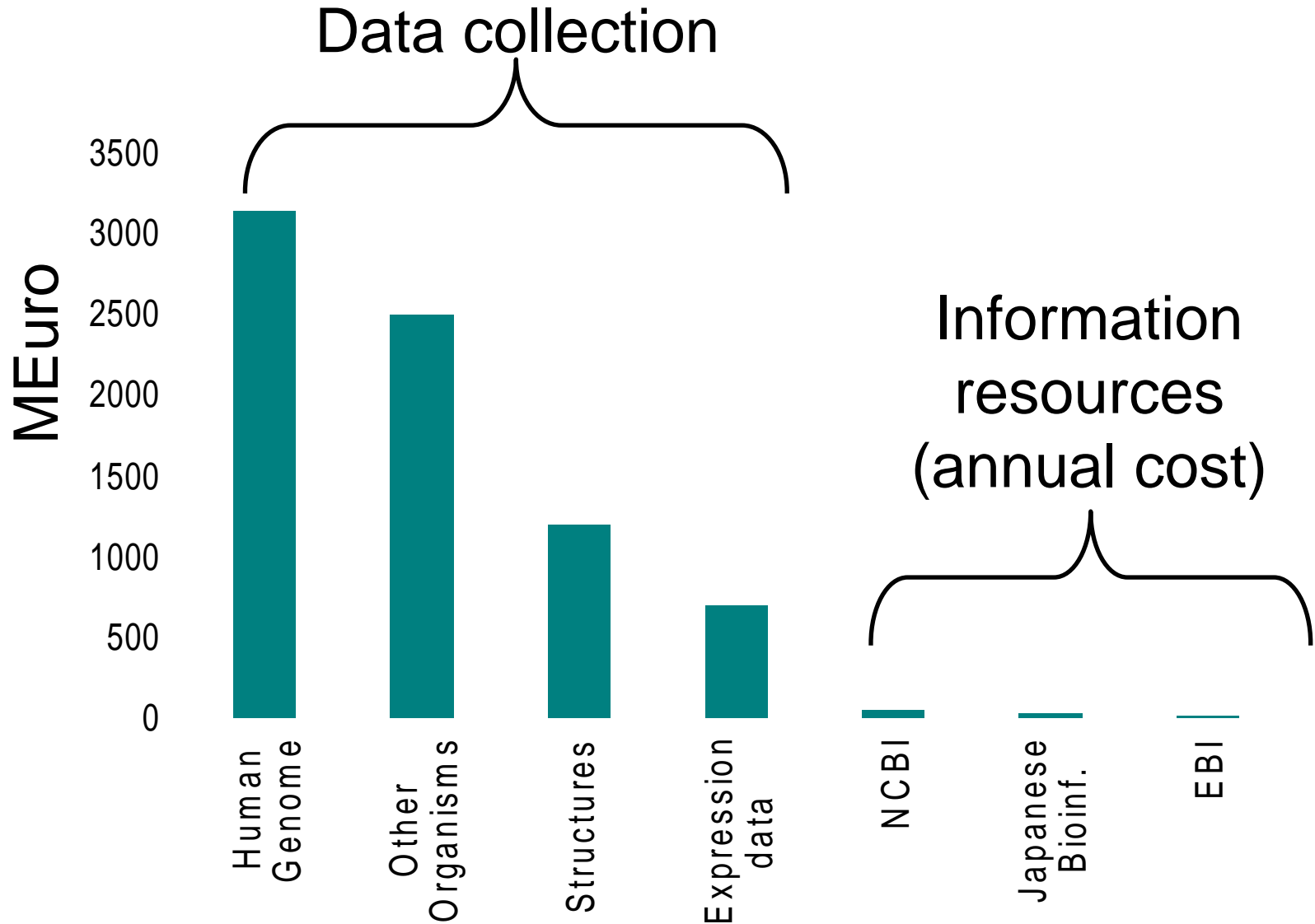
Integration and interoperability of diverse heterogeneous data

Forge Links to data in other related domains

A single European voice to influence global decisions and maintain open access

Enhance European competitiveness in bioscience industries

Address need for Increased Funding & its Coordination



Criteria for 'assessing' data resources

Usage

Cost & Value for Money

Stability

Standards

Size

International Status

Stakeholders

Funders of Infrastructure

EMBL; EU; National Government Funding Bodies; Charities; Industry

Data Resource Providers

Core Resources

Specialist (Many investigators - distributed)

Data Providers

Experimentalists

Tool Providers

Bioinformatics Groups

Users

Application for Preparatory Phase submitted May 1 2007

Partners: 30

Funding Bodies: 16

Institutes: 13

Representative of Industry: 1

Letter of Support from 35 other organisations

Plus numerous individual scientists

Key Components of ELIXIR

ELIXIR will provide:

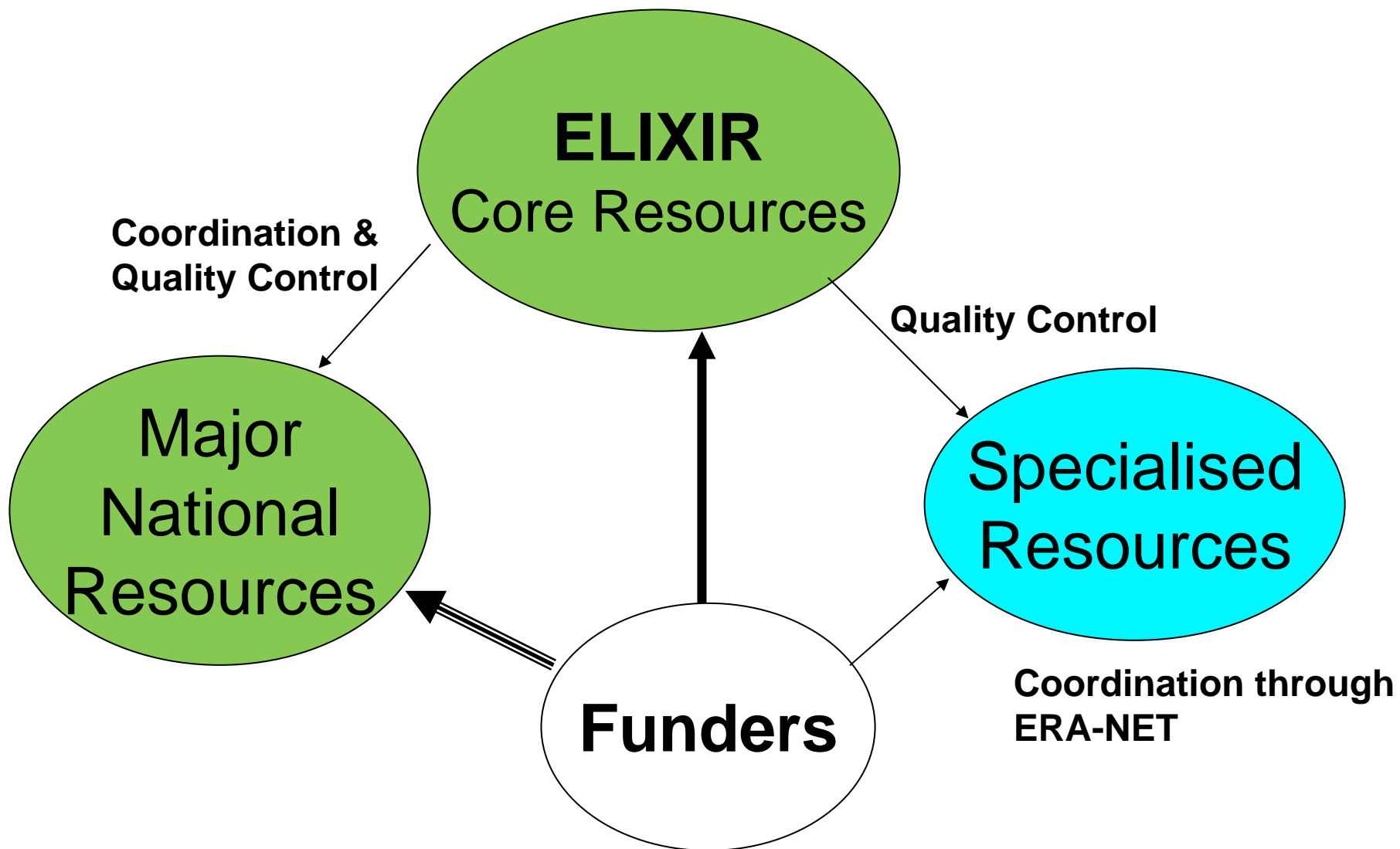
A major upgrade of the current infrastructure, including construction of a European Biomolecular Data Centre.

A trans-national infrastructure for biological information and service providers

Infrastructure for biological information in the new accession states.

The legal and financial framework for the construction and sustainable operation of this infrastructure.

Coordination of Training and Outreach to optimise use of information



Biomedical & Life Sciences

EATRIS – Centres for Translational Research

Bio-banking

INFRAFRONTIER – Mouse phenome & archive

Clinical Trials & Biotherapy Facilities

Integrated Structural Biology Centres

Upgrade to European Bioinformatics Infrastructure

LifeWatch – infrastructure for monitoring European Biodiversity

E-Science (Data Grid) and HPC computing

Key Challenges

• **Technical challenges**

- Integrating data
- Fast access needed
- Distributed data
- Pace of change in biology/medicine/genomics
- New data e.g. all human genomes; epigenetic data
- Rate of increase of data
- Development of standards

• **Sociological Challenges**

- Funding Model
 - Mixture of centralised and distributed
 - Many partners
- International context
- Increase need for storage and CPU power
- Integration with related fields e.g. medicine

Benefits of ELIXIR

Optimising access and exploitation of life-science data.

Ensuring longevity of the data and protecting investments already made in research which collected the data,

Increasing the competence and size of the already-large user community by strengthening national efforts in training and outreach.

Enhance the global success and influence of Europe in life-science research and industry.