

**BioMedBridges** is a joint effort of ten biomedical sciences research infrastructures on the roadmap of the European Strategy Forum for Research Infrastructures (ESFRI). Together, the partners strengthen biomedical resources by developing technical bridges to allow **interoperability between data and services** to answer scientific questions and address challenges in the **biological, medical, translational and clinical domains**.

## Overcoming practical obstacles to data sharing

Data and information are currently scattered in different databases and resources that are often only accessible to the scientific community that produced them. In addition, the communities frequently use different ways of storing and accessing the data. However, **access to these resources to a wider community across disciplines and national borders** and its interoperability is critical to **answering a wide range of complex scientific questions and address current societal challenges**.

Using the tools provided by BioMedBridges, each European research infrastructure will be able to deliver real and sustained improvement in data services, which will benefit the other infrastructures and the European and global research communities.



### Heterogeneous data

#### Types of data

atome biomacromolecular biomarker biosamples **cancer** cell  
**chemicalcompound** diseases dna dnacopynumber  
 elements **gene** genomicrearrangements genotype icd10 indels interaction  
 metabolite morphologicalimage mutation organismpart particle **pathways**  
**phenotype** **protein** rna samples shape  
 specie structure symptom **targetdrug**  
**transcript**

#### Data standards

**aim** amplitudes biomart bridg cdash **cdisc** clinical commercial  
**dicom** dictionaries emdb envs file  
 data  
 finnish genomic standards consortium gff gsc h7 icd10 kegg kogg meddra miame  
**mixs** **mmcif** mp mp ontology national nmr-star observ-om odm omim  
**ontology** open geospatial consortium pcom protein 3d coordinates  
 protocol sdfm send services snomed ssh structural structural amplitudes transfer web >gap  
 >xml

#### Data formats

**aim** arb bam bedgraph biomart **cdisc** cif **CSV** **dcm** **dicom**  
 embi emdb **fasta** **fastq** files formatvcf godml genbank **gff** hkl jpeg  
 images ins json lab loinc miame mysql mzdata mzml mzxml odm openclinica templates  
 pathology pdb pdf proprietary format sca scanner-specific sdfm **tabular**  
 tiffjpg tsv txt vcf **xml** xsd

## BioMedBridges will:

- ✓ Add value to existing data by making it discoverable and linking it: creating links between data that were not linked before will hugely increase the potential for new discoveries
- ✓ Bring together different communities in the biological and medical sciences by creating a common understanding of and approach to data (standards, formats, conventions, provenance models, ontologies)
- ✓ Enable interoperability of vast life science data resources across national borders
- ✓ Turbo-charge the participating ten new biomedical sciences research infrastructures as they build up new resources

**Training** is central to the BioMedBridges activities. With an emphasis on internal training, operators as well as users at the ten participating research infrastructures are trained to use the tools and resources necessary to achieve data interoperability.

The project is accompanied by a **technology watch panel** that captures the expertise from different e-infrastructure providers (STFC, DANTE, GEANT, CERN, CSC, EGI) to ensure that the solutions adopt the state of the art.

## Five use cases

### Interoperability of large-scale image data sets

Extensive image data sets representing different biological scales spanning biomolecules, cells, tissues and organisms will be linked, enabling drug-target and biomarker discovery for human disease.

### Improving the link between data from mouse disease models and humans

The mouse is an important model organism for studying human disease. Harmonising ontological descriptions of phenotype in mouse and human and improving links between mouse model data and human data, using diabetes and obesity as examples, will increase the relevance of data that is generated in mouse studies for clinical studies.

### Personalised medicine

Support for the development of targeted treatment options for patients with terminal acute myeloid leukemia by secure integration of patient information with data generated by the different research infrastructures in an ethical, robust and sustainable manner.

### Integrating structural data

Structural data on biomolecules will be linked with clinical data, maximising its value by enabling its use in studies of important biological and medical problems.

### Integrating disease-related data and terminology

Integrating information from biobanks and molecular resources makes samples more easily discoverable and will assist the identification of underlying causes of disease.

## Data interoperability across many different disciplines

