

# BRIDGES

BioMedical Research Informatics Delivered by Grid Enabled Services

[www.brc.dcs.gla.ac.uk/projects/bridges](http://www.brc.dcs.gla.ac.uk/projects/bridges)

## Overview

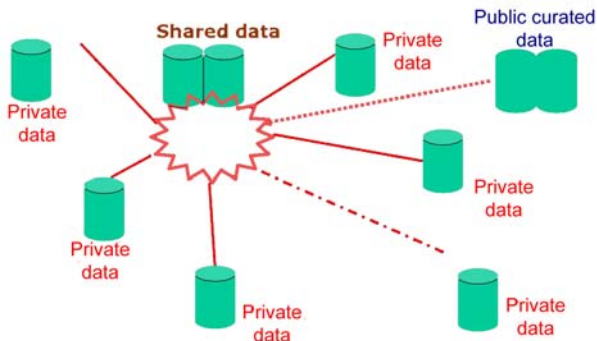
High blood pressure affects 25% of adults in western societies

£4.34M Wellcome Trust funded Cardiovascular Functional Genomics (CFG) project investigating this through physiological models of hypertension in rat

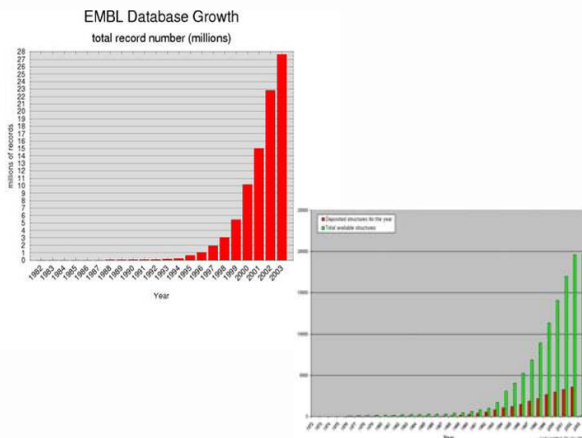
Bridges is a supporting project to CFG and will provide Grid infrastructure to facilitate scientific research

CFG project partners are distributed but need to access and integrate various software and especially data resources

Main aims of BRIDGES are to develop re-useable infrastructure to provide data federation incorporating appropriate security concerns



## Data Explosion



Data sources are growing exponentially

## Data Heterogeneity



Data comes from numerous sources in the CFG project:

- Genotype and phenotype data from rat breeding experiments
- Marker positional data from radiation hybrid and linkage analysis
- Gene expression data from microarray experiments
- Clinical data from human studies.
- Data from numerous online databases such as
  - GenBank, EMBL (nucleotide sequences),
  - SWISS-PROT (amino acid sequences),
  - Protein Data Bank (3D molecular structures),
  - Molecular Classifications (SCOP, CATH),

All of this data needs to be integrated to improve the overall research of the CFG scientists

## Problems to be addressed

BRIDGES will directly address the following problems facing the CFG biologists:

How to integrate data with multiple levels of security including public data, project only data and private data?

How to search multiple distributed databases through single optimised queries?

How to use multiple tools in a coordinated (and automated) manner, e.g. how to develop re-useable workflows for the CFG scientists?

How to deal with inconsistencies of online databases and possible "dirty data"?

How to get more "up to date" data?

## Technical Approach

BRIDGES will address these problems through Development of re-useable Grid services based upon Globus Toolkit version 3 technologies.

The virtualisation of multiple distributed data sets to provide a single virtual data set for use by the biologists – this will exploit IBM's DiscoveryLink technology.

The access to and integration of multiple distributed data sets in a Grid environment using results from the OGSA\_DAI/DAIT projects ([www.ogsadai.org](http://www.ogsadai.org))

A secure environment offering authentication and authorisation – building on results of the PERMIS security authorisation project ([www.permis.org](http://www.permis.org))

Development of XML schemas defining where to obtain online data, how often to download it and how to 'parse' it so that it may be integrated into a single database building on an XML configured application for automated download of data called 'local cache' created at Glasgow.

We will also investigate the problem of maintaining up to date parsing rules through on going work by initiatives such as "BioJava" and "BioPerl".

## Computational Infrastructure

Large scale compute power is required to perform data manipulations on a genome wide scale

The BRIDGES project will exploit resources such as the 95 node ScotGrid Beowolf cluster, and the IBM p690 Regatta Server Blue Dwarf – this will allow analysis of large SMP machines and cluster technologies for biological and life science research

## Further information

Further information on BRIDGES can be found at [www.brc.dcs.gla.ac.uk/projects/bridges](http://www.brc.dcs.gla.ac.uk/projects/bridges)

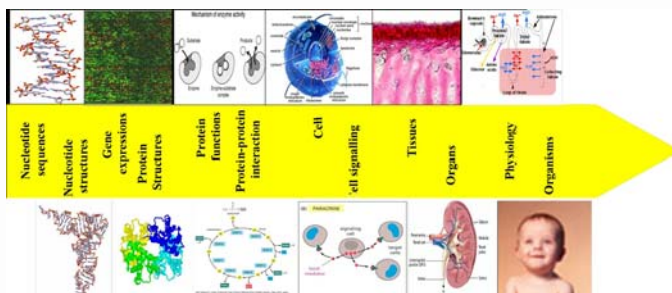
Or by contacting:

Dr Richard Sinnott ([ros@dcsgla.ac.uk](mailto:ros@dcsgla.ac.uk))  
– Grid related information

Prof David Gilbert ([drg@brc.dcs.gla.ac.uk](mailto:drg@brc.dcs.gla.ac.uk))  
– Bioinformatics related information

Dr Neil Hanlon ([hanlonn@dcsgla.ac.uk](mailto:hanlonn@dcsgla.ac.uk))  
– Biological related information

Dr David White ([david\\_white@uk.ibm.com](mailto:david_white@uk.ibm.com))  
– IBM DiscoveryLink information



## BRIDGES project members

Richard Sinnott, Neil Hanlon, Malcolm Atkinson, David White, Anna Dominiczak, David Gilbert, David Berry, Ela Hunt, Magnus Ferrier, Derek Houghton, Micha Bayer