Executive summary of the workshop
Data Collection in Ageing European Projects
Brussels, 10-11 June 2008

This workshop was co-organized by the Directorate General Research of the European Commission and by the European Coordination Action Link-Age (LSHM-2005-513866).

PARTICIPANTS

Database creator/ Provider/ Project leader in Bioinformatics

Benjamin Damien	benjamin.damien@bioxpr.be	BioXpr
Françoise Gofflot	Francoise.GOFFLOT@titus.u-strasbg.fr	MousePat
Laurent Vasseur	vasseur@igbmc.fr	MousePat
Axel Skytthe	ASkytthe@health.sdu.dk	Danish Twin
Patrice Godard	patrice.godard@fundp.ac.be	Senegen
Joao Magalhaes	joao@genetics.med.harvard.edu	Senescence info
John Hancock	J.Hancock@har.mrc.ac.uk	Casimir
Henning Hermjakob	henninghermjakob@googlemail.com	EMBL
Pascal Kahlem	pkahlem@ebi.ac.uk	ENFIN
Jennifer Hallinan	J.S.Hallinan@newcastle.ac.uk	Cisban
Glyn Nelson	glyn.nelson@newcastle.ac.uk	Cisban
Tom Kirkwood	Tom.Kirkwood@newcastle.ac.uk	Cisban

Representatives of European project in ageing research–database users

Alexander Bürkle	alexander.buerkle@uni-konstanz.de	Markage
Blair Henderson	blair.henderson@cemit.at	Tolerage
Marc Robinson-Rechavi	Marc.Robinson-Rechavi@unil.ch	Crescendo
Gillian Butler-Browne	butlerb@ext.jussieu.fr	
Claudio Franceschi	claudio.franceschi@unibo.it	Geha
Brian Clark	bfcc@mb.au.dk	Proteomage
Peter Kristensen	pk@mb.au.dk	Proteomage
Heinz Osiewacz	osiewacz@em.uni-frankfurt.de	Mimage
Jörg Servos	Servos@bio.uni-frankfurt.de	Mimage
Karlfierd Groebe	karlfried.groebe@proteosys.com	Mimage
Bas Zwaan	B.J.Zwaan@biology.leidenuniv.nl	Lifespan

Ethic specialist

Ninette Rothmüller	ninette.rothmueller@newcastle.ac.uk

Intellectual property specialists

Hakim Haouideg	hakim@ip-law.be
Denis Dambois	Denis.Dambois@ec.europa.eu

Organisers

Béatrice Lucaroni	Beatrice.Lucaroni@ec.europa.eu	Commission
Béatrice Rayet	beatrice.rayet@fundp.ac.be	Link-Age
Olivier Toussaint	olivier.toussaint@fundp.ac.be	Link-Age

The coordination action LINK-AGE organised a joint workshop with the European Commission Directorate-general for Research on collection, standardisation, compatibility and accessibility of the data generated by EU funded research projects in ageing.

The main objective of this meeting was to thoroughly discuss all questions that arise from research on data collection in ageing such as: What are the needs for sharing the results generated by the different European projects in ageing? How useful a centralised database for all these results would be? How to design and structure such a database? What are the needs for standardisation in this topic? What are the ethical concerns? What would be the requirements for such a project in terms of personnel, facilities and resources? What are the conflicts between the intellectual property and the public use and analysis of the data?

Since ageing research aims at studying multiple mechanisms, with multiple experimental models and human studies, researchers face a clear need for data and systems integration. Indeed, the big questions cannot be answered by a lot of disconnected and separate studies and the added value of data co-ordination justifies the effort required for data integration and sharing.

Ageing is a complex phenomenon, so we should follow the adage of Lord Rayleigh, Nobel Prize in Physics in 1904: "Neither seek nor avoid complexity in finding the appropriate solution to a problem".

Several Bioinformatics experts were invited to join some of the current European ageing project coordinators during this one day and a half workshop in order to address the questions and to thoroughly discuss three main concerns of the coordinators: i) data protection, ii) informed consent of participants in human studies and iii) the best way to structure a database and the relevance of data centralisation and modelling.

A first topic discussed was the concern of **data standardisation, structure and centralisation of the databases**.

Several ageing projects have defined a specific workpackage on bioinformatics. It has been suggested that every new proposal submitted for funding that will generate data should include a chapter explaining how partners will deal with data collection, storage and dissemination inside and outside the consortium.

Consensus for standards in ageing research is an absolute requirement if we want to pursue the goal of more integration and comparison of experimental data.
The importance of data standardisation has been well documented by the Bioinformatics experts present at the workshop. Henning Hermjakob presented the work done by several working groups that are trying to define the best standards to use in different molecular disciplines. The use of the right standard also stands for the vocabulary or ontology allowing for example data comparison between different species as it was stressed out by John Hancock, Laurent Vasseur/ Françoise Gofflot and Bas Zwaan.
A relevant standardisation of the data also means having a correct definition of the experiment to perform. A common language has to be used by all. It has also to be point out that before starting an experiment, we should know exactly what information we want and what standard to use. A database can be seen like a lab book. Moreover, too often, the experimenter will go to the statistician at the end of the study, when it is too late. An experiment that will generate

data, for which statistical analysis will be required, should be designed in advance, with the help of a statistician.
The collection of the data is of course a main step in the database implementation. Since data increase exponentially but not the researchers, there is a risk to loose information. The design of a database is a crucial step. It should respond to a specific problem asked by the researchers. The format and standards used should be well defined. To date, the most important thing is maybe to be able to find the information in the numerous existing databases, by centralising them on the web and trying to have as much compatibility as possible between them. The idea to construct a database that will integrate all types of data is probably not currently the best solution as we may not be ready for it.

The centralisation of the databases is a major issue. A centralised way to access the data does not necessarily mean the same physical place. However, for each kind of data, there is a need for one place where to get the information instead of 20 differently web located places. In order to be able to compare between data or to perform integration and modelling, the format has to be well defined.
Several initiatives exist, that aim at centralising maximum information on specific themes making them available to the research community. For example the CASIMIR project presented by John Hancock, the ENCORE platform introduced by Pascal Kahlem and the work done at the European Bioinformatics Institute at EMBL presented by Henning Hermjakob.
In ageing research, the CISBAN group in Newcastle directed by Tom Kirkwood is modelling and integrating experimental ageing-related data in an effort to perform systems biology studies.

A second issue discussed was **the ethical concerns in databases and the definition of the ideal informed consent**

With the extension of human data collection and their use for research purposes, ethical concerns are more and more at the forefront. Major advances have been reached in European research and projects need to address ethical issues seriously. The importance to define correctly informed consent was stressed.

How should the informed consent form look like?
We need to have a clear idea of what informed consent is and how the concept of informed consent can be adhered to during and after a project. There are different models of consent that are used. The specific form of consent that is implemented for use in a project must be carefully chosen. For example, if a consent form allows the donors to withdraw their consent, this must be possible. For some, this may cause a problem, due to uncertainty about how long we will be able to use the data. However, the consent form used determines what can be done with the sample. Biobanks exist that cannot be used anymore because the permission to use the samples or data for future uses was not obtained. New models of consent make it possible for participants to decide whether they would like to agree to future and on-going use of the sample, or be recontacted if the possibility of new use arises. It is very important that the public trusts the database users. Transparency is required.
The form of consent used should follow international level guidelines instead of national level ones to facilitate collaboration. Training workshops are needed.

The Commission is well aware of ethical concerns and has published a checklist of ethical issues for the attention of the public. A two stages ethical evaluation procedure is in place for

proposals submitted to the Framework Program in which ethical issues have been identified. For example, if human beings are involved, the proposed research project will automatically proceed to a review panel composed of ethics professionals. This is an eligibility criterion for all submitted projects to identify and explain the ethical issues involved in their Research & Development.
Until now, there is no harmonised consent form due to the high number of member states, local ethics committees and competent bodies. Some research themes, such as research aimed at human cloning for reproductive purposes or the creation of human embryos for research purposes alone, are completely banned from FP.

An open consultation has been launched by the OECD for the guidelines on the use of genetic databases of human and on the ethic and informed consent: these consultations are still accessible on the webpage (see the useful links section below).

Last, but not least, the third subject treated was **the protection versus the opening of the data**

We sometimes face a contradictory situation. Everyone agrees that sharing data is an important prerequisite to go forward towards data integration and systems biology. On the other hand, partners of European projects have to protect data in order to have the time to fully benefit from data they have generated in their projects. The accessibility of the data may also be restricted by the consortium agreement.
On this, we had two interesting presentations on the legal protection of databases in FP6. We learned that the originality to structure the data falls under the Copyright protection and that investments to collect the data are regulated by the « sui generis » right.

Yet, the non-legal guidance from the OECD and the Commission is to go towards more accessibility. The European Bioinformatics Institute at EMBL and other data integration initiatives are pushing the scientific community towards a better and more efficient sharing of their data.

A few presentations gave examples of databases that are either fully accessible like the MousePat or the Human Ageing Genomic Resources or databases that are restricted to research purposes and kept anonymous (for evident ethic reasons) like is the case of the Danish Twin Registry.

Altogether the presentations allowed the participants to better understand the problem. It is now time to go forward in the discussion to find consensus on standards and databases centralisation. One opportunity to do so will be the LINK-AGE Topic Research Group on data collection that will be held in Newcastle before the end of the year.

USEFUL LINKS

Ageing European project: the databases are/ will be available only to the partners.

www.geha.unibo.it: three databases implemented.

www.proteomage.eu: database implemented (BioXpr).

www.mark-age.eu: the database is not yet implemented.

www.lifespannetwork.nl/: database implemented.

www.crescendoip.org: database implemented.

www.mimage.uni-frankfurt.de: a database is not yet implemented.

TOLERAGE: The web site of the coordinator institute is www.i-med.ac.at/mypoint/ and the web site of the project management is www.cemit.at/ which is holding the project web site: http://www.cemit.at/folgeseite.cfm?id=227.

Walloon region research project (in French):http://recherche-technologie.wallonie.be/projets/

Senegen: no web site. Database implemented by BioXpr.

Databases publicly accessible:

www.mitomap.org: hosts the first mitochondrial DNA complete sequences. It has the capacity to assign haplogroups automatically, to recognize amino acid variations automatically and to obtain all the references related to a polymorphism automatically.
Each partner can use this database to store and analyze data from different project on mitochondria. This allows a future merging of all stored data, thus creating one of the largest mitochondrial DNA published database in the world.

www.lifecompetence.eu/: FP6 SSA project on building a online database - competence platform containing 680 health related projects from FP6.

http://genomics.senescence.info/: the Human ageing resources web page.

http://www-mci.u-strasbg.fr/mousepat/organ: the collection of nuclear receptors expression in the mouse brain.
http://www.eumodic.org/: EUMODIC undertakes a primary phenotype assessment of up to 650 mouse mutant lines derived from ES cells developed in the EUCOMM (http://www.eucomm.org/) project.

http://empress.har.mrc.ac.uk/: EMPReSS is a database of SOPs (Standard Operating Procedures), developed by the EUMORPHIA (http://www.eumorphia.org/) consortium that can be used to describe the phenotype of a mouse.

http://bgee.unil.ch: a dataBase for GeneExpression Evolution.

http://www.ebi.ac.uk/intact: The IntAct Molecular Interaction Database.

http://www.ebi.ac.uk/pride/: The PRoteomics IDEntifications database.

www.interphenome.org: the integration of mouse phenome data resources.

http://www.p3gobservatory.org/: Public Population Project in Genomics. The Observatory is a central Internet repository of scientific information and tools aimed at facilitating the development, realization and harmonization of research projects.

http://reactome.org/: a curated knowledgebase of biological pathways.

http://www.bbmri.eu/: Biobanking and Biomolecular Resources Research Infrastructure FP7 project (Research Infrastructures)

http://www.phoebe-eu.org/eway/default.aspx: PHOEBE (Promoting Harmonisation of Epidemiological Biobanks in Europe) - ongoing FP6 Coordination action

http://www.danubianbiobank.de/DanubPublic/misc/mscHome.jsp: DANUBIOBANK (Danubian Biobank Initiative - Towards Information-based Medicine) - ongoing FP6 Specific Support Action.

http://www.oecd.org/document/12/0,3343,en_2649_34537_40302092_1_1_1_1,00.html: OECD Guidelines for Human Biobanks and Genetic Research Databases (draft for public consultation attached)

Bioinformatics tools:

http://mibbi.sourceforge.net/: The minimum information for biological and biomedical investigations.

http://fuge.sf.net: The Functional Genomic experiment.

http://www.hupo.org/: The Human Proteome Organisation.

http://imex.sf.net: The International Molecular Exchange Consortium is a collaboration between the BIND, DIP, MPACT(MIPS), MINT and IntAct databases.

http://www.biomart.org/: a query-oriented data management system.

http://www.molgenis.org/: The Molecular Genetics Information System (MOLGENIS) is an biological database infrastructure that can be customized for specific experiments like large scale transcriptomics, metabolomics, proteomics and genetical genomics.

http://taverna.sourceforge.net/index.php: Data gathering and processing using workflows. The Taverna workbench is a free software tool for designing and executing workflows.

http://bioinf.ncl.ac.uk/comparagrid/: Federated, service-oriented semantic data integration architecture.

http://www.calibayes.ncl.ac.uk/: Integration of GRID-based post-genomic data resources through Bayesian calibration of biological simulators.

http://www.basis.ncl.ac.uk/: Web-based services for quantitative study of the biology of ageing.

Intellectual property rights:

http://ec.europa.eu/research/fp6/model-contract/pdf/fp6-iprguidelines_en.pdf: IPR Guidelines (FP6).

http://eur-lex.europa.eu/LexUriServ/site/en/oj/2002/l_355/l_35520021230en00230034.pdf: Rules of participation (FP6).

http://www.ipr-helpdesk.org: Intellectual Property Rights-Helpdesk.

http://ec.europa.eu/research/fp6/model-contract/pdf/checklist_en.pdf: C.A. checklist (FP6).

http://ec.europa.eu/invest-in-research/policy/ipr_en.htm: Relevant links and reports on IPR, tech transfer, etc.

http://ec.europa.eu/internal_market/copyright/prot-databases/prot-databases_en.htm: Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

http://www.oecd.org/dataoecd/9/61/38500813.pdf: OECD guidelines on access to research data

http://www.allenovery.com: an international legal practice in 28 major cities worldwide.

http://www.oecd.org/document/12/0,3343,fr_2649_34537_40302092_1_1_1_1,00.html: The consultation period for the OECD draft Guidelines for Human Biobanks and Genetic Research Databases has been extended to 15 July 2008. OECD Principles and Guidelines for Access to Research Data from Public Funding

http://ec.europa.eu/invest-in-research/pdf/download_en/patents_for_researchers.pdf: "Why researchers should care about patents"

http://ec.europa.eu/invest-in-research/policy/ipr_en.htm: Intellectual property and technology transfer:

http://ec.europa.eu/research/science-society/document_library/pdf_06/council-conclusions97236_en.pdf: Council Conclusions on scientific information in the digital age: access, dissemination and preservation

http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf: COM(2007)56 final - Communication on scientific information in the digital age: access, dissemination and preservation

European Bioinformatics Projects:

http://www.casimir.org.uk/

http://www.enfin.org

National Bioinformatics project in ageing:

http://www.cisban.ac.uk/

Bioinformatics service providers:

http://www.genomatix.de/

http://www.bioxpr.com

REFERENCES FROM THE LITERATURE

ETHIC
Genomics and equal opportunity ethics. A W Cappelen, O F Norheim and B Tungodden. J Med Ethics 2008;34:361-364.

DATABASE
de Magalhaes et al. 2005. *Nucleic Acids Research*, **33**:D537-D543

Interphenome: Mammalian Genome 18, 157-163 (2007)

Kerrien et al. IntAct--open source resource for molecular interaction data. Nucleic Acids Res 35 (2007) D561-565
Orchard et al. Submit your data the IMEx way. Practical Proteomics (2007)

MIAPE parent document plus technology-specific modules:
Taylor, C. et al: The minimum information about a proteomics experiment (MIAPE) Nat Biotechnol. 2007 Aug;25(8):887-93.
Orchard, S. et al: The minimum information required for reporting a molecular interaction experiment (MIMIx) Nat Biotechnol. 2007 Aug;25(8):894-8

The HUPO PSI Molecular Interaction Format - A community standard for the representation of protein interaction data. Henning Hermjakob et al. Nature Biotechnology 2004, 22, 176-183.
Kerrien et al. Broadening the Horizon - Level 2.5 of the HUPO-PSI Format for Molecular Interactions. BMC Biol 5 (2007) 44
Data deposition recommendations:
[Editorial] Democratizing proteomics data. Nat Biotechnol. 2007 Mar;25(3):262.
[Editorial] Time for leadership.Nat Biotechnol. 2007 Aug;25(8):821.
[Editorial] Mind the technology gap. Nat Methods 2007 Oct; 4(10):765.
[Editorial] Thou shalt share your data. Nat Methods 2008 5(3):209.

Presentation of the HUPO in:
Nature Vol 437 Issue N°7056 2005.