

Project Summaries

1. VIDGRID- Distributed Video Analysis with Grid Technologies
Video Annotation, Distributed video analysis
2. FINGRID- Financial Information Grid-
Real time data analysis, simulation
3. Collaborative Analysis Of Offenders' Personal And Social Exclusion
Data Grid, data integration, middleware
4. SABRE in R- An OGSA Based Approach to Middleware for Statistical Modelling
Statistical analysis, Data Grid, middleware
5. Data Chronicles- Repository of Metadata
Repository of Metadata, Metadata standards
6. Learning Disabilities and Infrastructure
Data integration, metadata, data grid
7. Entangled Data- Knowledge and Community Making in E- Social Science
Community Making, Training and awareness
8. Access Grid and 'Using Access Grid in Field Research' Project
Communication and Collaboration, Practice lead Research
9. Semantic Annotation in Skills Based Learning
Video Analysis, grid, semantic annotation
10. Grid Enabled Micro Econometric Data Analysis
Data and Computational Grid, dataset integration
11. Training and Awareness Raising Resource Discovery for Researchers
Training and Awareness
12. Genealogies of Knowledge- Developing anthropological middleware to support
Grid support for fieldwork, field-based social science
13. SAMD- Seamless Access to Multiple DataSets
Workflow, Computational Grid, Data Grid, Dataset integration
14. Software Tools for Indigenous Knowledge Management
Software tools for Indigenous knowledge management
15. Contextual instantiation of indigenous domain knowledge
Interactive research, fieldwork

16. CONVERTGRID

Common Geography Conversion

17. MIT Open Course Ware

Teaching and Learning

18. BiodiversityWorld

Problem Solving Environment, Virtual Research Environment, Database and Catalogue builder

19. Discovery.net

Data Grid

20. NERC Datagrid

Data Grid, metadata

21. AXIOPE- Grid Enabled Databases and tools for Neuroinformatics

Database Building, Data integration

22. My Grid

Middleware and data grid toolkit

23. GridShib

Secure data transfer

24. RealityGrid

Middleware and datagrid toolkit

25. InThreeD

Visualisation

26. SPICE Project

Taxonomies, Metadata, Classification

27. BRIDGES- Biomedical Research and Informatics

Virtual Organisations, Virtual Research Environments, Data Grid, Middleware.

28. EDIKT

Dataset Fusion, Virtual Organisations, Community Making

29. VOTES- Virtual Organisation for Trials and Epidemiological Studies

Data sharing, Security, Virtual Organisations

30. Particle Physics Grid-

Grid Computing, Grid Toolkits

31. PsyGrid

Data Grid, community making

32. Geode <http://www.geode.stir.ac.uk/>
Data sharing, Security, Small Data Grid
33. SRBN- Scottish Research and Bioinformatics Network (AKA Scotgrid)
Small Data and Computational Grid
34. N/A16 Dataset Integration-
Dataset Integration
35. GVIZ
Grid enabled visualisation
36. GEMEPS
Data Grid, Data Searching
37. GEODISE
Grid enabled Problem Solving Environments
38. BASIS
Community making, Data Grid
39. WOSE- Workflow Optimisation Services for e-Science Applications
Workflow optimisation for e-science applications
40. DAME- distributed Aircraft Maintenance Environment
Community Making, Virtual Research Environment, Problem Solving Environment, Data Grid
41. JIGSAW- Distributed and Diagnostic Visualisation-
Visualisation
42. Fasttrack- A training Scheme to Support Understanding of the Grid
Training and awareness scheme
43. VIPIR- Visualisation in Parallel
Visualisation
44. E VIZ- High Performance Visualisation
Grid-enabled high performance Visualisation
45. GRENADE (Grid-Enabled Desktop Environments)
Grid enabled desktop environments
46. Astrogrid- A Virtual Observatory
Computational Grid, Data Grid
47. GROWL- VRE Programming and Toolkit
Virtual Research Environment Toolkit
48. An Investigation of the Disclosure Risk Issues Posed by the GRID

Security, disclosure risks assessment

49. Silchester Project

Virtual Research Environment, Data Sharing

50. BUREK- Building a Virtual Research Environment in the Humanities

Virtual Research Environments

51. VRE in the History of Political Discourse 1500-1800

Virtual Research Environment, Data Sharing

52. Wungrid- Music Collaboratory

Data Grid, Middleware, Music

53. e-Diamond Project

Data Grid, Image Analysis, Metadata, Standardisation

54. DigitalRecord

Data, Grid, Digital Record, Corpus, Qualitative Analysis

55. TextGrid

Data Grid, Multiple, Annotation, Text, Community

56. Memetic

Virtual Research Environments, Access Grid

57. ICEAGE

Training and Awareness, Grid Education, Forum

1

VIDGRID: Distributed Analysis with Grid Technologies,

<http://www.kcl.ac.uk/depsta/pse/mancen/witrg/vidgrid.html>

The Vidgrid is designed to show the potential of the grid to improve existing video analysis techniques. It also seeks to establish a set of general parameters and requirements for video analysis among distributed researchers in light of the ways in which both researchers and practitioners handle, catalogue, share, examine, present and disseminate materials to others. It has created a software demonstrator, and provides an analysis of the key difficulties associated with distributed video analysis over the Grid at different stages of the research process- including data collection and storage, data analysis and data presentation.

Summary Keywords- Video Annotation, Distributed video analysis

Category- Practice led research, Communication and Collaboration

FINGRID: Financial Information Grid, Completed Demonstrator Project

<http://www.computing.surrey.ac.uk/grid/fingrid/>

The Fingrid is a grided tool aimed at modelling trends in financial markets while taking into account real-time events. Grid middleware was used to create a 24-node grid for the managing, retrieving, and analysing of 'live' data. This involves simultaneous sharing and analysis of quantitative data (time and financial figures) and qualitative data (financial news). The news was analyzed using existing text extraction and management systems, and the quantitative data was graphically simulated with various tools. Both sets of data were then fused and correlated, and the models distributed via the grid. The simulations could also be updated in real-time with new qualitative and quantitative data (delivered through any of the nodes) while it was still running. At its root, the FINGRID provides a way of collaboratively (or individually) simulating and analysing both quantitative and qualitative data in the same software research environment. The project also developed the **Systemquirk** software tool (See tools summary).

Summary Keywords- Real time data analysis, simulation

Category- Data Capture, Data Structuring and Enhancement, analysis, communication and collaboration

Collaborative Analysis Of Offenders' Personal And Area-Based Social Exclusion

<http://www.shef.ac.uk/~scgisa/escience/home.htm>

<http://www.allhands.org.uk/2004/proceedings/papers/90.pdf>

The purpose of this project is to develop and promote the data grid through practically addressing research questions in the social sciences.

The questions the project tried to answer were-

- 'Are there measurable relationships between the characteristics of young people with difficulty, and the socio-economic and physical characteristics of the area they live in?'
- 'Are government policies addressing the multiple facets of social exclusion, including health, education, employment, and crime prevention, targeted in the areas that need them most, where young offenders live and most crimes are committed?'

This will involve combining the searching, integration, and portrayal of data across different information platforms held by distributed research institutions and governments. It aims to show how the grid can facilitate collaboration and allow interdisciplinary research to be conducted amongst experts in different disciplines who not work together otherwise, in this instance with particularly challenging sensitive and confidential data.

Summary Keywords- Data Grid, data integration, middleware

Category- Data Structuring and Enhancement, communication and collaboration

4

SABRE in R- An OGSA Component-Based Approach to Middleware for Statistical Modelling <http://www.ncess.ac.uk/research/pdp/#ogsa>

<http://sabre.lancs.ac.uk/>

SABRE is a well-known social and political science program for the statistical analysis of binary events. Such data are common in many surveys either with recurrent information collected over time or with a clustered sampling scheme. See **Sabre** in the tools section.

The aim of this particular project is to use gridded middleware for handling the large-scale statistical modelling problems dealt with by SABRE in the past. Sabre used to operate on individual machines. Now, it can now it can take advantage of Grid and attempt more computationally intensive calculations, and deal with a wider array or data-types. The project hopes to expand the use of SABRE, and create an enhanced version of it which can be freely downloaded as open source software.

Summary Keywords- Statistical analysis, Data Grid, middleware

Category- Data Structuring and Enhancement, Analysis

5

Data Chronicles- A Repository of MetaData?

http://www.nesc.ac.uk/action/projects/project_action.cfm?title=265

The project will investigate the feasibility of making a central repository of metadata for social science data sources to help social scientists make the most effective use of data held in existing archives. Too often, (as in the Arts and Humanities) social scientists require 'expert knowledge' of a field in order to locate relevant data.

This project's 'Repository' represents an advance on existing archives in three respects. Firstly, it will include administrative and survey data that is often important in social science. Secondly, users will be able to search the Repository using either generic or specialised terms to find and navigate sources. Thirdly, the metadata to be held will include references to 'key debates', - recognising that social science concepts cannot be defined 'once and for all' in a piece of metadata- a similar challenge to that faced by the arts and humanities in many areas.

The project will focus on two topic areas in the first instance: drugs and employment - and use this to test the long-term feasibility of developing a repository covering the full range of social science data sources. The results of this investigation might be of interest to projects in the Arts and Humanities that concern metadata standards.

Summary Keywords- Repository of Metadata, Metadata standards

Category- Data Structuring and Enhancement, Strategy and Project Management

6

Learning Disabilities and Infrastructure

<http://www.essex.ac.uk/hhs/research/Projects/LDgrid.htm>

This project recognises that data (in this case, Health Data) is 'often distributed amongst numerous organisations, and rarely has consistent structures and meaning'. The project's exemplar is Learning disability data – which is distributed amongst the government, voluntary sectors, and academia. This project will review the requirements for, and assess the potential of, grid related technologies, especially in terms of ontologies, web services and distributed computing – to solve these common problems,

The project will assess the user requirements and technical feasibility of e-science technologies for such research. In order to do this, it will build up a firm understanding of the user requirements both for policy and research, using interviews and workshops. It will also examine the technical feasibility of joining up disparate data sources within the service providers using Grid technologies.

Summary Keywords- Data integration, metadata, data grid

Category- Data Structuring and Enhancement, Strategy and Project Management

7

Entangled Data- Knowledge and Community Making in E- Social Science

http://www.nesc.ac.uk/action/projects/project_action.cfm?title=259

“This project will conduct a comparative study to understand how and why groups of research scientists do or do not collaborate using shared digital data sources. It will develop insights into the likely use and non-use of e-science technologies, and the social and technological innovations that may be required as e-science

It will do this by tracing the ways in which data travels across geographically distributed networks of researchers, and analysing the concepts of ‘community’ that are involved in making data travel. This promises to contribute to our wider understanding of academic knowledge production and disciplinarily. Three cases of ICT-based data sharing practices will be examined, one in the physical sciences and two in the social sciences.

One group are already using e-science technology to collaborate and share data, but are in the early stages of infrastructure and application design. The second group is longer standing, but analyses data complex enough to ask questions about both cultures of interpretation as well as grid-enabled possibilities. The final group uses a mix of technologies, and works primarily with visual data. The project will analyse the social circumstances of these three groups that make it sensible to share data (or not), by comparing discourses about sharing with actual sharing practices.

The case study results will be translated into ideas for grid engineering projects and will feed insights into the e-science community through scenarios, concepts and an ‘innovative ideation phase’.”

This particular study – which is almost anthropological in nature - finishes this year and may yield some useful insights and implications for the creation of an e-arts and humanities agenda.

Summary Keywords- Community Making, Training and awareness

Category- Communication and Collaboration

8

ACCESS GRID and related projects

<http://www.accessgrid.org/>

‘New Technologies, New Applications’

<http://www.ncess.ac.uk/research/small-grants/#Access%20Grid%20Nodes>

The Access Grid is becoming essential in the hard sciences. It allows large scale distributed meetings and collaborations via video conferencing at Access Grid nodes. Lectures and meetings can theoretically take place between geographically distributed

researchers and students as if they were in the same room. Unsurprisingly, more and more universities are investing in these nodes. There seems no reason why video conferencing and meetings of geographically distributed researchers should be any less applicable to the A&H than it is in the sciences. Although Access Grid nodes are expensive to set up, configuring a machine for access grid use is relatively simple and costs around £100. A social science project ('New Technologies, New Applications') has also investigated the potential of the Access Grid to facilitate social science fieldwork and interviews). The E-Science support centre is producing a briefing paper on the potentials and realities of using the Access Grid in the Arts and Humanities.

The Access Grid may be thought of as a prototype for "next generation video conferencing". The aim of the inventors of Access Grid was to implement a technology that could support productive meetings between remote participants that are as effective as face-to-face meetings - all using commodity hardware. This can only happen if one is able to forget the technology and concentrate on the meeting itself.

Features of the Access Grid:

- Very high quality audio
- Big display to enable full-size people shots and simultaneous viewing of all remote sites
- Multiple cameras to show groups and multiple viewpoints

Collaborative software to enable remote participants to share and interact with data

Summary Keywords- Video Conferencing

Category- Communication and Collaboration, Practice lead Research

9

Grid-enabled data collection and analysis – semantic annotation in skills-based learning (Videa Annotation)

<http://www.ncess.ac.uk/research/small-grants/#semanticannotation>

This project will demonstrate how Semantic Grid technologies can be used to provide enhanced techniques for data collection and use within a grid-enabled environment. It will look at **semantic annotation** (i.e. machine-processable annotation using Semantic Web technologies) as a case study. It will look at how it is used both in capturing and working with digital records.

It intends to produce training output in the application of grid technologies for the wider social science community, and set the agenda for future work in this area in the social sciences.

This project will examine new forms of data capture and analysis involving multiple simultaneous video and audio sources and video focus groups.

Summary Keywords- Video Analysis, grid, semantic annotation

Category- Practice led research, Communication and Collaboration

10

Grid Enabled Micro Econometric Data Analysis Dr SA Peters

University of Manchester

http://www.nesc.ac.uk/action/projects/project_action.cfm?title=249

This project is a demonstrator- looking to show how the grid can enhance existing research practices.

The seamless interfacing of distributed data and computational power via the grid has not been widely applied in economics. This pilot project aims to measure and demonstrate the effectiveness of such a system applied to micro-econometric analysis – in this case, of the welfare of ethnic minority groups in the UK. Building on a grid-enabled framework for a macro-economic problem, the project will develop methods of **fusing data** from complementary data sets in order to allow new research questions to be addressed. In its own words, the broad aims of this project are to:

- (i) “Investigate and demonstrate the feasibility of performing grid-enabled micro-econometric analysis by extending an existing e-science framework for macro-economic data analysis
- (ii) To Investigate the potential of constructing ‘matched/data-fused data’ from original, separate grid-enabled sources
- (iii) To investigate the potential gains of performing modern micro-econometric analysis on high performance computers.”

Summary Keywords- Data and Computational Grid, dataset integration

Category- Data Structuring and Enhancement

11

Training and Awareness Raising Resource Discovery for Researchers (ReDReSS) Professor Robert Crouchley University of Lancaster

<http://redress.lancs.ac.uk/>

This training and awareness project for e-science in the social sciences is employing several strategies-

- (i) It will create a 'web based resource discovery portal' aimed at introducing and explaining how Grid technologies can enhance qualitative and quantitative social science research. This will include short introductory guides, training materials and exemplars (drawn from the demonstrator projects);
- (ii) (ii) It will run a road show series across the social science community explaining how e-science can enhance the capacity to analyse complex social and economic problems
- (iii) (iii) It will organise a programme of research workshops which will discuss and debate how e-science can contribute to the development of theory, methods and collaborative working within the social sciences.

ReDReSS is an extension of the services provided by TRAMSS (Teaching Resources and Materials for Social Scientists, [tramss.data-archive.ac.uk /index.asp](http://tramss.data-archive.ac.uk/index.asp))- which aims to make e-Science relevant to substantive researchers in the social sciences. It promises to make all of its material freely available online when it has been created.

Summary Keywords- Training and Awareness

Category- Training and Awareness

12

GENEALOGIES OF KNOWLEDGE- Developing anthropological middleware to support field-based social science

<http://grok.anthropology.ac.uk/proposal.html>

The main aim of this project is to design and deploy Grid based support for key research processes in fieldwork based social science. In particular, this will involve supporting bibliographic research, interactive collection and aggregation of data within the fieldwork segment, and access to, and aggregation of, external data from the fieldsite. It will also involve the consolidation, analysis, modeling and dissemination of this data from the institutional base. The project will develop Grid resources to support the research process for fieldwork based research.

Summary Keywords- Grid support for fieldwork

Category- Practice lead research, communication and collaboration

13

SAMD- The Seamless Access to Multiple Datasets

<http://www.sve.man.ac.uk/Research/AtoZ/SAMD/>

This project was designed to demonstrate the use of Grid technologies for dataset manipulation and analyses in a social science context. Using an exemplar problem drawn from social science (the asymmetric effects of interest rate changes on the UK's GDP)- the project shows how the integration of access to both data and high performance computational resources within a single sign-on environment enables the automation of complex workflows. This might facilitate the 'scaling' up of software applications used for research in the Arts and Humanities. SAMD claims to transform an existing web-based data service into a flexible data repository "on the Grid" with a user-friendly interface.

Overall, SAMD claims to demonstrate the successful incorporation of emerging Grid technologies into an existing social science data service. It shows how the integration of access to both data and computational resources within a single sign-on environment enables the automation of complex workflows, facilitating the scaling up of social science research applications.

Summary Keywords- Workflow, Computational Grid, Data Grid, Dataset integration

Category- Data Structuring and Enhancement, Analysis, Data Publishing and Dissemination,

14

Software Tools for Indigenous Knowledge Management

http://archive.dstc.edu.au/IRM_project/software_paper/IKM_software.pdf

In the words of the authors-

"Indigenous communities are beginning to realize the potential benefits which digital technologies can offer with regard to the documentation and preservation of their histories and cultures. However they are also coming to understand the opportunities for misuse and misappropriation of their knowledge which may accompany digitization. In this paper we describe a set of open source software tools which have

been designed to enable indigenous communities to protect unique cultural knowledge and materials which have been preserved through digitization. The software tools described here enable authorized members of communities to: define and control the rights, accessibility and reuse of their digital resources; uphold traditional laws pertaining to secret/sacred knowledge or objects; prevent the misuse of indigenous heritage in culturally inappropriate or insensitive ways; ensure proper attribution to the traditional owners; and enable indigenous communities to describe their resources in their own words. Hopefully the deployment of such tools will contribute to the self-determination and empowerment of indigenous communities through the revitalization of their cultures and knowledge which have been eroded”

Summary Keywords- Software tools for Indigenous knowledge management

Category- Data Capture, Data Structuring and Enhancement, Data Publishing and Dissemination, Strategy and project management

15

Contextual instantiation of indigenous domain knowledge

David Zeitlyn, Stephen Lyon, Michael Fischer, Paul Sillitoe

<http://bateson.anthropology.ac.uk/overview/overview.html>

“Some social research is based on interaction with people, where continuous analysis influences subsequent data collection, and where the research themes themselves may change while the research is being carried out.

E-Science technologies have enormous potential for advancing the quality, applicability and applications of interactive research, and directly address many of the greatest problems encountered when applying interactive methods. Interactive methods result in data sets that are difficult to code or otherwise organise, where the data are often difficult to compare, and where the management of data, and especially of transformations of the data (such as coding), can be daunting. The access grid, data grid and computational grid can all be leveraged to support these aspects of interactive research.”

This project is investigating the organisation, structure, transmission, creation and deployment of knowledge, using interactive research methods, and exploiting middleware to support toolkits that can be adapted to specialised, research driven qualitative research tools.

This demonstrator project approaches e-social science research support using a largely interactive, qualitative approach. It will develop support for qualitative data and

software components that address interactive collection and aggregation of data within the fieldwork segment, access to and aggregation of external data from the field site, and consolidation, analysis, modelling and dissemination from our institutional bases. This will be most visible as a distributed generalised Qualitative Data Analysis (QDA) framework that will support data collection and integration, layering, aggregation and collaborative analysis and dissemination within a grid framework and where possible within a conventional WWW environment, and in asynchronous mobile contexts.

This project might be good to look to as a demonstrator of the use of e-science for interactive research.

Summary Keywords- Interactive research, fieldwork

Category- Data Structuring and Enhancement, Practice lead research

16

CONVERTGRID

<http://www.sve.man.ac.uk/Research/AtoZ/ConvertGrid/>

<http://pascal.mvc.mcc.ac.uk:9080/convert>

Many researchers now wish to investigate complex research questions that require the combination of datasets from multiple sources. If multiple datasets are required, the user has to currently launch a web browser, navigate different interfaces, download the required portion of the dataset to local disk, and import it manually into whatever analysis system is appropriate for the task. However, if the datasets relate to different *geographies* the user will first have to convert the datasets to a common target geography.

This conversion process requires specialist skills and therefore represents a major barrier to systematic cross-analysis of multiple datasets. The ConvertGrid project offers a simple solution to this problem of geographical conversion. A web-based interface allows a user to select data from various data sources for a specified target geography. The resultant queries are run in a single sign-on environment against a number of databases which contain data relating to different geographies. The resultant data streams are transferred to the Grid-enabled version of the current ESRC funded Convert system which uses look-up tables derived from the All Fields Postcode Directory to automatically convert the native geographies to a common target geography.

Summary Keywords- Common Geography Conversion

Category- Data Structuring and Enhancement, Strategy and Project Management

17

MIT Open Course Ware

<http://www.gridpp.ac.uk/qcdgrid/>

<http://ocw.mit.edu/index.html>

MIT (Massachusetts Institute of Technology) in Boston have made a large number of their courses available online. The paper materials, handouts, bibliographies, and most excitingly- video and audio files of the lectures- are available through the MIT website, and special forums and wikis exist to support geographically distributed researchers and help them collaborate. Essentially the idea is to have courses in any subject available online within a virtual learning environment.

Summary Keywords- Teaching and Learning

Category- Data publication and dissemination

18

BIODIVERSITYWORLD: A Problem Solving Environment for Global Biodiversity

<http://www.bdworld.org/index.php>

<http://www.trianacode.org/docs/index.html#>

Biodiversity World is a good example of a problem-solving environment. Various heterogeneous data was merged to form a database, and customised analysis tools can then be run on it. BDWorld's database handling is characterised by the diverse types of database it draws upon (many from the web) and the heterogeneity of the data, (with respect to both the data structures and standards). The power behind BDWorld comes from a software tool called **Triana**. It provides an environment that allows even the technologically illiterate researcher to create programmes to interrogate and manipulate data by compiling workflows via a graphical system. The programs can be used on a wide variety of data: numerical data, audio data, images, and text files. Essentially, when a researcher is confronted with a monotonous research process- or large or diverse multiple datasets- the **Triana tool** can be used to create and execute custom queries of all of the data through a central graphical interface- which theoretically creates exciting opportunities for the simultaneous analysis of multiple datasets. See description of Triana on the tools list.

Summary Keywords- Problem Solving Environment, Virtual Research Environment, Database and Catalogue builder

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration

19

DISCOVERY.NET- A Data Grid, Prof. Yike Guo

<http://www.discovery-on-the.net/>

Discovery Net is a multidisciplinary project serving application scientists from various fields including biology, combinatorial chemistry, renewable energy and geology. It provides a **service-oriented computing model for knowledge discovery**, allowing users to connect to and use data analysis software as well as data sources that are made available online by third parties. Essentially, the project is all about establishing standards for e-knowledge discovery; it is looking at how best to structure a data grid. It does this by *defining* the standards, architecture and tools that enable a Data Grid to function effectively. These are the standards, architecture and tools that-

- Allow researchers to plan, manage, share and execute complex knowledge discovery and data analysis procedures available as remote services.
- Allow service providers to publish and make available data mining and data analysis software components as services to be used in knowledge discovery procedures.
- Allow data owners to provide interfaces and access to databases, data stores, sensors and experimental results as services so that they can be integrated in knowledge discovery processes.

Summary Keywords- Data Grid

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration

20

NERC Datagrid- <http://www.e-science.clrc.ac.uk/web/projects/nercdatagrid>

The NERC Datagrid is similar to the Discovery.net project insofar as it is an intellectual exercise aimed at pioneering an effective model for a Datagrid.

They contend that the effective 'scientific use' of data is compounded by the difficulty of 'getting started' because data is sometimes not discovered; many researchers relying on word of mouth to find out about new or re-processed datasets. Currently, users can find out general information about collections of data (datasets), and then must either manually find and locate specific files or databases of interest. To evaluate the usefulness of the data, they then have to do a great deal of work- often of a quite trivial kind- such as examining the metadata standards used, database schema, and the types of digital file formats. Therefore, a lot of time is wasted as researchers 'reinvent the wheel', and many projects do not end up happening because of the unwieldiness of many datasets.

There is a considerable amount of poorly described and catalogued data held by research groups and institutes, some of which is held in locally managed archives, and some of which is simply stored in files whose contents and location are known only to the originators. Much of this data- in the sciences - is amenable to *more sophisticated automated management* that would both make life easier for all.

The proposed new NERC DataGrid claims it will deal with these issues, making it easier to discover and utilise data both from the managed archives of the designated data centres and from the semi-managed archives of the larger research programmes.

This proposed DataGrid will be designed so that it will eventually underlay geospatially referenced data delivery and discovery in the UK across the earth science disciplines, but will initially be developed to support activities in the atmospheric and oceanographic sciences. However, data is a generic commodity used as much by Arts and Humanities Researchers as Scientists- so this datagrid could almost certainly be exported outside the sciences.

“The underlying concept is that documentation, model data and observational data should all be amenable to easy discovery and manipulation by both software agents and user-driven Graphical User Interfaces (GUIs). While users will inevitably require extra information about data characteristics and quality for detailed research, the proposed technology will make it easier to *locate and use data for inter-comparison* and to confront models with reality (a major requirement for progress). All tools and software developed will be publicly available, and every effort will be made to integrate the requirements of the wider NERC community in the development cycle.

To enable this, a metadata and a data model are being implemented. More detail can be found in the [metadata](#) and [data virtualisation](#) papers presented at the 2003 UK e-Science All Hands Meeting. PowerPoint presentations outlining the [metadata](#) and [data](#) models are also available. “

Summary Keywords- Data Grid, metadata

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration

AXIOPE: Biomedical Data Projects , Completed

<http://www.axiope.com/>

<http://www.allhands.org.uk/submissions/papers/219.pdf>

Various databases – including one on Neuroinformatics- have been built up with a tool called ‘Catalyzer’ developed by Axiope. Distributed databases were merged, and represented with a universal management system. Catalyser is a user-friendly database builder, and its outputs can be grid-enabled. See the description of Catalyser on the tools list.

Summary Keywords- Database Building, Data integration

Category- Data Capture, Data Structuring and Enhancement, Data Publishing and Dissemination,

My Grid

<http://www.mygrid.org.uk/>

The myGrid project was primarily designed to employ a Grid-enabled middleware framework to support data intensive tasks in silico experiments in biology. It consists of a comprehensive set of loosely coupled middleware components, which are specifically tasked to manage the complex, scientific processes within this bioinformatics domain. The technology developed by the myGrid project is applicable for managing many aspects of research by adopting its technology for storing data, sequencing tasks in a specific order, informing users of target events or changes in experiments, resource discovery, and provenance management.

The myGrid toolkit can be freely downloaded and has been used for building discovery workflows for investigations into Williams-Beuren Syndrome and Grave’s Disease by collaborating Life Scientists.

The myGrid toolkit consists of core components for forming, executing, managing and sharing discovery experiments. These components are chosen and invoked by developers and experiment designers to produce the desired end result.

MyGrid provides a series of supplementary services to aid a scientist, which aim to cover fundamental issues such as reproducibility of analyses, publishing results, and being informed of changes in experiments. In many ways, myGrid functions as a

virtual lab book, helping scientists to track and write up experiments in a readable format, and subsequently distributing and running it elsewhere

Bioinformaticians and service providers develop and run experiments via the Taverna workbench.

Summary Keywords- Middleware and data grid toolkit

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration

23

GridShib

<http://grid.ncsa.uiuc.edu/GridShib/>

GridShib, a joint project of NCSA, University of Chicago, and Argonne National Laboratory, is an integration of Globus Toolkit (GT) and Shibboleth. The complete software package consists of two plugins, one for Globus Toolkit and another for Shibboleth. With both plugins installed and configured, a GT Grid Service Provider (SP) may securely request user attributes from a Shibboleth Identity Provider. As such, GridShib allows Grid virtual organizations and higher-educational institutions in a given federation, to securely exchange attributes about each other such as names, resources, and roles. This further helps in developing and implementing specific profiles for each organization or institution in that federation.

Among the benefits of GridShib are:

- Builds on existing technologies, which primarily include Shibboleth and the Globus Toolkit.
- It's based on established security standards and these include SAML, X.509 and XACML.

Summary Keywords- Secure data transfer

Category- Data Publishing and Dissemination

24

RealityGrid

<http://www.realitygrid.org/>

<http://www.sve.man.ac.uk/Research/AtoZ/RealityGrid/>

RealityGrid's conception of Grid Computing is what they call 'an ambitious and exciting global effort to develop an environment in which individual users can access computers and databases simply and transparently, without having to consider where those facilities are located' They are particularly concerned with computationally intensive visualisations. It is a multi-million pound EPSRC 'steering' project, which is looking to break into new ground in computationally intensive grided visualisation.

RealityGrid has thus provided a highly flexible and robust computing infrastructure for supporting the visual modelling of scientific matter. It is currently creating grid middleware to make itself more widely useable. Also, it is creating a flexible, and customisable 'Problem Solving Environment' package through which the power of the RealityGrid might be harnessed. This might offer a gateway for the Arts and Humanities to engage in computationally intensive visualisation through one of the most advanced scientific visualisation services.

Summary Keywords- Middleware and data grid toolkit

Category- Data Capture, Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration, Analysis, Strategy and project management

25

InThreeD- <http://www.dur.ac.uk/inThreeD/>, Nick Holliman, Durham University

InThreeD is a project that is investigating software tools aiming to make 3D displays as easy to use as 2D displays. The first software demonstrator allows users to easily create high quality stereoscopic images.

The project is researching novel approaches to mapping real or virtual **scene depth** captured by a stereo camera pair onto a defined range of perceived stereoscopic depth on a 3D display. Our new approach allows the image creator to have control over this depth mapping such that a region of interest in the scene can be given preferential depth representation compared to the rest of the scene. This provides a solution to the problem that 3D displays can only comfortably reproduce a limited range of perceived depth. The inThreeD prototype will be a plug-in for the widely used 3DS Max package.

The project acknowledges that anywhere where 3D information is used and stored could benefit from visual representation on-screen with an extra dimension.

Summary Keywords- Visualisation

Category- Data Structuring and Enhancement

**SPICE Project <http://www.sp2000.nies.go.jp/>
(Species 2000- <http://www.sp2000.org/>)
'Investigating a federated architecture for creating a catalogue of life'**

This project is essentially an intellectual exercise. It draws upon a number of discrete databases maintained by experts- (concerning species types in biology). The objective is to create a federated architecture which will enable heterogeneous data to be made available under the same set of standards, classifications, and metadata. The project aims to employ wide taxonomic coverage which can only be achieved through wide agreement. It investigates the problems of creating a system of more common classifications and standards in a field where disagreements between researchers is rife. Essentially, it is an investigation of whether a master-system can be used to harmonise data - making it more conducive to sharing, data base building, storage, and many other things.

It aims to-

Create a set of standard biological data fields and associated functions that will permit many databases to contribute taxonomically authoritative (species) entries to a seamless catalogue of information about all known species. It aims to create a 'canonical knowledge model' which can function effectively with heterogeneous databases.

Create a model of taxonomic and professional practice which permits federated cooperation both between expert biological communities and between biological organisations;

The taxonomic databases produced by the project are available in the Species 2000 database.

Summary Keywords- Taxonomies, Metadata, Classification

Category- Data Structuring and Enhancement, Data Publishing and Dissemination.

BRIDGES- Biomedical Research and Informatics-
<http://www.brc.dcs.gla.ac.uk/projects/bridges/>

The BRIDGES study is a good example of the impact of a relatively small grid on the informatics of a particular academic subject- in this case, Biomedical Research.

BRIDGES (Biomedical Research Informatics Delivered by Grid Enabled Services) is developing and exploring database integration over six geographically distributed research sites in Cardiovascular Genomics. Three classes of integration are being developed to support a sophisticated bioinformatics infrastructure supporting: 1.) data sources (both public and project generated), 2.) bioinformatics analysis and visualisation tools, 3.) and research activities combining shared and private data. Security is also a high priority as the project concerns personal medical data.

Both [OGSA-DAI](#) and IBM [Information Integrator](#) technology are being employed as middleware and the project will access their usefulness.

Project Goals

The project promises to deliver the following results:

1. An effective environment for biomedical bioinformatics. This will include federated access to data, analysis and visualisation across at least the UK centres with appropriate authorisation and privacy.
2. An improved understanding of the requirements for the support of academic biomedical research virtual organisations. This will be published as a final project report and exemplified with publicly available re-usable data access and integration components.
3. An evaluation of the utility of various existing and emerging federation tools (e.g. replication tools such as GIGGLE, query tools such as DiscoveryLink and platforms such as OGSA-DAI) in this class of application. A particular issue is whether Grid-based technology can assist with the management of the bioinformatics infrastructure and processes.

Summary Keywords- Virtual Organisations, Virtual Research Environments, Data Grid, Middleware.

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

28

EDIKT- <http://www.edikt.org/>

EDIKT– stands for e-Science Data, Information and Knowledge Transformation. It is an initiative at the National e-Science Centre to construct **data management** and **interpretation** software tools. These will underpin the seamless linking, management

and interpretation of vast amounts of data. **EDIKT** will do this by enabling researchers to harvest the knowledge hidden in the acres of data with which leading researchers work.

The authors acknowledge that “Bigger datasets allow us to see a more complete picture, but extracting knowledge from them is like finding a needle in a haystack– and sometimes we don’t actually know what the needle looks like! **EDIKT** will develop tools – based on state-of-the-art computer science research – that will integrate and search these enormous datasets.

EDIKT is initially investigating the use of new database techniques in astronomy, bioinformatics, and particle physics. It is also examining database techniques that might be used in creating virtual global organisations using the new Open Grid Services Architecture.

The Project has also produced two downloadable ‘spin off’ tools – Eldas and BinX - Enterprise Grid applications which deal with enabling Virtual Organisations. See tools list for further descriptions.

Summary Keywords- Dataset Fusion, Virtual Organisations, Community Making
Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration

29

VOTES- Virtual Organisation for Trials and Epidemiological Studies

<http://labserv.nesc.gla.ac.uk/projects/votes/index.html>

The Virtual Organisations for Trials and Epidemiological Studies (VOTES) project will investigate the application of Grid technologies to the clinical trials domain. Specifically, the project will focus upon the areas of:

- Patient recruitment
- Data collection
- Study management aspects of clinical trials

By bringing Grid solutions to these large-scale problems, the VOTES project hopes that the clinical domain will benefit from the resulting advances in statistical analysis and compilation.

Context and Challenges

The clinical trials domain is a particularly challenging one in that potentially overlapping, dynamically varying, geographically distributed, heterogeneous groups

of people, resources and data sets need to come together for given trials. These might be described as Virtual Organisations (VOS). The VOs for clinical trials obviously have very strict security requirements on the data sets, the data resources, who is allowed to see, use them, in what context etc.

The main challenges involved in this project will be to federate data from many sources, crossing domains that do not necessarily entirely trust each other. Also, the issues surrounding data classification will need to be addressed, with implementations of data schemata, meta-data and ontologies. Security is also very important.

Project Plan

Using Grid applications that are in widespread use today, a portal will be constructed that provides distributed access to several databases- modelling the schema and structure on those used by health organisations. The Grid technologies that will be used include: GridSphere, the Globus Toolkit (v4.0) and OGSA-DAI.

Summary Keywords- Data sharing, Security, Virtual Organisations

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

30

Particle Physics Grid- <http://www.gridpp.ac.uk/news/>, **Professor Tony Doyle**
(project Leader)

This is a £35 million collaborative project between 20 universities and CERN- the European organisation for nuclear research. It's aim is to deliver a robust and efficient computational grid for the UK physics community- who often have need of large computational processing power. The project is not just about infrastructure- its also concerns the development of software applications which will make the grid accessible and understandable to those that use it. It is probably the largest and most expensive project of its kind in the UK.

Summary Keywords- Grid Computing, Grid toolkits

Category- Data Structuring and Enhancement , Analysis, Communication and Collaboration.

31

PsyGrid- <http://www.psygrid.org/>

Psygrid is a Datagrid that deals with continually changing and geographically distributed large datasets of extremely sensitive medical data from people with first episode psychosis.

Psygrid has two philosophical aims. The first is to develop the Mental Health Research Network (MHRN) into a functioning “e-community” in order to pull together partners and strengthen Research and Design capacity. The second aim is to build a secure electronic database to hold anonymised clinical and demographic data about patients. These data sets will be integrated with other existing data sources within the NHS and elsewhere- such as the Mental Health Minimum Dataset.

In addition, web/grid-based methods of enhancing data collection in multisite clinical trials will be developed. Thus, the information system will track a large, representative cohort of individuals with first episode psychosis to further research ends.

A computer science team at University of Manchester will build the Psygrid platform and database. An ethics, security and privacy workstream will consult and advise the programme. There will be liaison with the NHS Information Authority and the NHS National Programme for IT (NPfIT).

From late 2005, clinical data from consenting people with first episode psychosis will be entered around the MHRN hubs on to a secure web-based database.

The PsyGrid DataGrid itself will provide resources to clinicians, researchers and those interested in first episode psychosis. It will consist of databases, e-Learning resources, electronic records and communications systems which will all be linked together by middleware and available through the internet

Summary Keywords- Data Grid, Community Making

Category- Data Publishing and Dissemination, Communication and Collaboration,

32

GEODE- Grid Enabled Occupational Data Environment, Prof. Paul Lambert, Stirling <http://www.geode.stir.ac.uk/>

The GEODE project is concerned with the technologies behind the distribution of ‘occupational information’ within the social science research community. Occupational information refers to summary statistics and data that are associated with particular occupational positions. Such information is used by many social research projects, but is not always available in an easily accessible form. It collates this data from several sources.

The GEODE project is an attempt to use e-Science and Grid technologies to facilitate access to occupational information. The project has two main aims. The first is to develop an **online 'portal'**, which will allow social science users to securely submit their datasets (such as survey data with occupational

records), and have the data returned with appropriate occupational information (such as social class classifications) matched to the relevant cases. The second aim is to develop an **online environment to act as a 'depository' for occupational information**, for use by a wide range of national and international social science researchers.

Naturally, such a project can help foster collaboration between researchers, considerably widen access to data, and save a lot of spadework (see above).

Overall, this is a quite a small and simple (yet effective) Data Grid which was cheap to set up (£45,000), which could act as an aid to study where geographically distributed data and researchers are concerned.

Summary Keywords- Data sharing, Security, Small Data Grid

Category- Data Publishing and Dissemination, Communication and Collaboration.

33

SBRN: Scottish Bioinformatics Research Network AKA SCOTGRID

<http://labserv.nesc.gla.ac.uk/projects/sbrn/index.html>

<http://www.scotgrid.ac.uk/>

This is a small data and computing grid, which involves 3 Scottish Universities (Dundee, Edinburgh, and Glasgow). They aim to share resources and information, and collaborate in ways they did not before through using this Network. Although this is nowhere near as large or conceptionally innovative as some of the big multi-million pound Scientific Grids (such as the Particle Physics Grid), it is a small and relatively contained example of how information and computing grids can enhance the research process - that might be more appropriate to A&H scholars and departments with more limited technical experience and less financial resources.

Summary Keywords- Small Data and Computational Grid

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

34

N/A16 Dataset Integration-

http://www.nesc.ac.uk/action/projects/project_action.cfm?Title=65

Dr A Bateman, Dr C Orengo, Dr A Murzin, Dr T Hubbard, Dr R Apweiler

This is a major dataset integration project. Through distributed computing technologies and grid middleware it plans to co-ordinate and integrate information from five existing gene databases allowing researchers from the UK and around the world to increase the pace of their basic research. There is a substantial degree of overlap between the information in the gene databases, and dataset integration will enable a pooling of knowledge.

Summary Keywords- Dataset integration

Category- Data Structuring and Enhancement

35

gViz Project - Visualization Middleware for Grid Users

<http://www.comp.leeds.ac.uk/vis/gviz/>

This project finished in July 2004, and was funded under the UK e-Science Core Programme. It studied the middleware required to support visualization systems in a Grid computing environment

The gViz project is about visualisation through the grid. Its goal has been to investigate the middleware required to support visualization systems in a Grid environment. They describe their approach as pragmatic: first to evolve two existing visualization systems to work in a secure distributed fashion, providing tangible benefits to the e-science community within the lifetime of the project; second, from that initial basis to seek abstractions that have greater generality and will contribute to the long-term development of visualization for e-science.

Software (the gViz Library- latest version 1.1) has been developed to act as the middleware 'glue' between simulations running on the Grid and a visualization system running on the desktop. An XML language (skML) to describe visualization pipelines has also been developed, together with a proposal for handling visualization data formats. Demonstrators with application to environmental science and computational biology were created. The latest software is **gViz Library 1.1**. The published results of the project can be found at <http://www.comp.leeds.ac.uk/vis/gviz/publications.html>

Summary Keywords- Grid enabled Visualisation

Category- Data Structuring and Enhancement, Data Publishing and Dissemination.

36

GEMEPS- Grid Enabled Microarray Experiment Profile Search-
<http://labserv.nesc.gla.ac.uk/projects/gemeps/index.html>

The aim of this project is to develop Grid based solutions that facilitate the discovery and re-use of previously undertaken microarray (Genes and DNA) experiments in a secure environment. It is essentially a Data Repository.

It's most important goal is to use the capability of the grid to identify previously undertaken microarray experiments, and allow the integration of these data sets with local microarray experiments. This methodology can help prevent the same work being done twice. To this end, the project is developing tools to enable the comparison of different biological experiments- with an eye to identifying and drawing out the common factors. At the moment, there is no way of retrieving experimental data which enables biological scientists to discover whether an experiment similar to their own has already been undertaken. Such a system as GEMEPS which can help identify and share common experimental data is potentially a revolutionary labour saving device.

Challenges

The project concedes that its users generally want to keep close control over who accesses their pre-publication microarray data- a challenge which naturally also applies in the Arts and Humanities. Also, centralised data repositories are costly to set up and maintain, and depend on the researchers handing over their sensitive data. Such a data repository also relies on being kept up to date in a rapidly changing academic area for it work properly. Finally, it acknowledges that different systems in different labs create inconsistencies in the presentation of the data which must be dealt with by grid middleware.

Approach

GEMEPS will use grid technology to allow users to securely and selectively share sensitive, pre-publication microarray data. Comparison of experiments will be by means of ranked lists of expression profiles, allowing users to compare experiments across platforms. Public data will also be used to enhance this search facility. GEMEPS will also utilise technology used successfully in previous projects for authenticating and differentially authorising users (web portals, **PERMIS authorisation software**)

Summary Keywords- Data Grid, Data Searching

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration, Strategy and Project Management

GEODISE - Geodise: Grid enabled optimisation and design search system

<http://www.geodise.org/>

<http://e-Science.ox.ac.uk/public/eprojects/geodise/index.xml?style=printable>

GEODISE is one of a new breed of EPSRC funded research programmes involving multi-disciplinary teams working to build a state of the art design tool demonstrator.

Intelligent design search tools will one day become a vital component of all engineering design systems and will steer the user through the process of setting up, executing and post-processing design search and optimisation activities. Such systems require large-scale distributed simulations to be coupled with tools to describe and modify designs, using information from databases and knowledge bases. These tools are physically distributed and under the control of multiple elements in the supply chain. Whilst evaluation of a single design may require the analysis of gigabytes of data, to improve the process of design can require assimilation of terabytes of distributed data.

The Geodise e-Science project is a collaboration between academia and industry. The project's aim is to develop a Grid enabled Problem Solving Environment (PSE) providing access to computational fluid dynamics (CFD) analysis packages and specialised design search and optimisation tools, which are used for designing products as diverse as vacuum cleaners and aerospace systems. The design of such complex products involves realistic simulation of their reactions under different conditions. This requires access to considerable local computational resources and the analysis of very large data sets. Often only developers in large companies have access to such systems but, with the development of Geodise, many more engineers and designers will have access to high-powered computers over the Grid.

The Geodise toolbox is a suite of Grid-enabled design optimisation and search tools within the Matlab* environment. The tools provide functions allowing access to distributed computational and database resources, a suite of design optimisation and search algorithms, CFD solvers and pre- and post-processing. An integrated knowledge base provides context sensitive design advice to the user about the use of these tools. Users may also be notified of job completion or failure by email or wireless Short Message Service (SMS) text messages to devices such as mobile phones. Given commands in a high level interpretive language it is straightforward for the engineer to exploit available Grid-enabled resources to tackle computationally and data intensive tasks.

In the foreseeable future intelligent, Grid-enabled systems like Geodise, are set to become the work horses of engineering design systems. They will aid the user through the process of design optimisation from setting the system up, performing computations, and finally through post-processing activities.

*The Matlab package provides an interpretive generation language for numerical computation, built-in math and graphics functions and numerous specialised toolboxes for advanced mathematics, signal processing and control design. The Matlab product is widely used in academia and industry to prototype algorithms, and to visualise and analyse data.

Summary Keywords- Grid enabled Problem Solving Environments

Category- Analysis, Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

38

BASIS- Biology of ageing e-science- <http://www.basis.ncl.ac.uk/>

The goal of the BASIS project is to deliver a web-based system that will serve the 'biology-of-ageing' research community by helping to integrate data and hypotheses from diverse biological sources. It aims to establish a community-based web of activity to foster collaboration in this area

BASIS is one of 20 pilot projects funded under the UK e-science initiative to help develop UK grid application. The models it creates will be made available. The project will comprise of geographically distributed components (simulators, software, databases) but will rely on the Grid to enable access to multiple resources within a secure environment.

Summary Keywords- Community Making, Data Grid

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

39

WOSE- Workflow Optimisation Services for e-Science Applications

<http://esc.dl.ac.uk/WOSE/>

Rob Allen, Asif Akram- Daresbury Laboratory and Florian Urmetzer- University of Reading

The WOSE project claims that little attention has been paid to the composition of services to develop applications to enable service discovery and optimisation of e-science applications.

The WOSE project will allow users to develop applications by composing existing services that make use of distributed resources. This will be done through a problem-solving environment (PSE) that automatically defines service relationships and dependencies in the form of an annotated workflow graph. Thus, creating applications becomes somewhat clearer and easier. The WOSE PSE will also help users make services available via the Grid. Overall, this research aimed at dynamically selecting the best schedule of services to execute a composite service-based application according to a set of user specified criteria.

The WOSE PSE provides a useful method to model a diverse range of possible Grid resources (e.g., 'computational servers, data repositories, software libraries and scientific instruments'). It's target audience is naturally researchers considering setting up an e-science project, who need guidance in how to go about it. It also might enable workflow enhancement for existing services

Summary Keywords- Workflow optimisation for e-science applications

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, training and awareness.

40

DAME- distributed Aircraft Maintenance Environment

<http://www.cs.york.ac.uk/dame/>

http://www.cs.york.ac.uk/dame/SJ4205-Leaflets7_DAME1.pdf

DAME will develop 'generic test bed' for Distributed diagnostics that will be built upon grid-enabled technologies and web services. This project is a demonstrator for this generic framework- the exemplar concerns the maintenance applications for civil aerospace engines. This project will draw together a number of advanced core technologies, within an integrated web services system:

DAME is particularly concerned with addressing a number of problems associated with the design and implementation of on-line decision support systems. The most significant of these are- 1.) access to remote resources (experts, computing, knowledge bases etc), 2.) communications between key personnel and actors in the

system, 3.) control of information flow and data quality, and 4.) the integration of data from diverse global sources within a strategic decision support system.

The new web services model for information brokerage on the Internet - offers an inherently pragmatic framework within which to address these issues. DAME will exploit emerging open standard web service architectures over a GRID network to demonstrate how the data management aspects of maintenance support systems can be handled within a 'unified knowledge broker model'.

Summary of DAME's Aims which might be relevant to the Arts and Humanities-

- To develop a Distributed Diagnostic Grid Test-bed.
- To design and build a system architecture for distributed diagnostic support that takes advantage of Grid middleware, focusing particularly on the management of the data within the Grid.
- To study, and increase understanding of, the performance issues associated with data grids (e.g. data replication).
- To design a distributed data store for unstructured, non-indexed data

Summary Keywords- Community Making, Virtual Research Environment, Problem Solving Environment, Data Grid

Category- Data Publishing and Dissemination, training and awareness.

41

JIGSAW- Distributed and Diagnostic Visualisation-

http://www.nesc.ac.uk/action/projects/project_action.cfm?title=122

Jigsaw is attempting to find ways in which visualisations can operate successfully in a distributed manner when it has to deal with (possibly) millions of data items that may be, for example, incomplete, anomalous, or dynamically changing. They acknowledge that the problem of visualising large data sets is an active and well-documented research area.

However, the problem of **uncertainty** in data that is to be visualised has received little attention. They term this as 'uncertainty visualisation' and claim that 'What is missing at the moment is a way of incrementally building a visualisation from a stream data source, where the contents of that data source can only be guessed'. Jigsaw's solution lies in increasingly flexible generation algorithms and metaphors (These are used to fuel the visualisation process). These new algorithms and metaphors should be better equipped to handle uncertainties in the data stream without disrupting the visualisation.

Summary Keywords- Visualisation

Category- Data Structuring and Enhancement

42

FASTTRACK- A training scheme to support Social Scientist's understanding of Grid Technologies- Completed

http://www.jisc.ac.uk/index.cfm?name=project_fasttrack&src=alpha

Fasttrack is essentially a training scheme for social scientists wishing to understand the potential and realities of using Grid technologies to enhance their research. Social scientists are certainly more similar to A&H researchers than are 'proper' scientists, so this project may yield valuable lessons. This project has produced training materials tailored to its findings- such as a guide to Video conferencing through the Access Grid). Its outputs and findings might be useful for Arts and Humanities Scholars wishing to make assess the potential of the Grid. It might also provide a basis for future training and awareness schemes initiated in the Arts and Humanities.

Summary Keywords- Training and awareness scheme

Category- Training and Awareness.

43

VIPAR: Visualization in Parallel

<http://www.csar.cfs.ac.uk/about/csarfocus/focus7/vipar.pdf>

VIPAR can speed up visualization. The project's original aim was to remove performance bottlenecks from the visualization pipeline. At that time several other projects had implemented solutions that were tied to specific parallel support libraries, hardware or visualization systems. These all had poor portability, VIPAR was developed to have better portability. Existing Visualization tools (e.g.AVS/Express, Iris Explorer, IBM Data Explorer and Khoros) provide access to large visualization libraries via a highly developed user interface which is easy for new or non-technical personnel to operate. The user drags and drops modules from the systems libraries and connects them to produce a network which in turn produces an associated application and visualization.

The VIPAR libraries are a suite of routines that provide an interface between the visualization system and the message passing system.

The three libraries are, VPRvsi - a visualization library to provide an interface to the visualization system, VPRidd - an intermediary library that performs the necessary calculations and provides the interface between the two other libraries, and VPRdd -

the lower level message passing and parallel process control library. These libraries are designed to sit on top of one another and provide increasing levels of complexity. VIPAR is also a construction kit for producing parallel visualization modules quickly and easily. However this tool, the DDTTool, was only ever produced as a prototype.

Summary Keywords- Visualisation

Category- Data Structuring and Enhancement, Data Publishing and Dissemination.

44

E-VIZ – High Performance Visualisation

<http://www.eviz.org/>

This ongoing project is about High Performance Visualisation (HPV). This is characterized by high-quality graphics, large datasets, and computationally-intensive tasks. A typical HPV task is a complex process (involving data collection, visualization design, task parallelisation, immersive visual display, and simulation engines). This project is addressing the conceptual models that encapsulate HPV information and tasks, and will formulate a generic model for a networked supercomputing environment (compatible with the Grid). Other key questions the project will deal with include:

- How does data flow into, within, and out of an HPV environment?
- What are the commonly used protocols governing the interactions among users, visualization and simulation systems, computational and visualization results, and local and remote resources for HPV?

Can a general purpose software tool be built to manage HPV tasks automatically, and if so, can such a tool be customised to suit HPV environments of various configurations?

Summary Keywords- Grid-enabled high performance Visualisation

Category- Data Structuring and Enhancement, Data Publishing and Dissemination,

45

GRENADE (Grid-Enabled Desktop Environments)

<http://www.sve.man.ac.uk/Research/AtoZ/GRENADE>

The Grenade project argues that if the Grid is ever to deliver its potential, it must be made accessible to a much wider community than technologists. Key to this are appropriate, **user-friendly interfaces**, accessible in the user's natural environment. The Grid-Enabled Desktop Environments (GRENADE) project is exploring the potential of Grid computing through the 'tight integration' of the Grid's capabilities

within the user's desktop environment. Through this, GRENADE hopes to improve the usability and visibility of the Grid.

The GRENADE project is completing the development of a prototype Grid-Enabled Desktop Environment based on [Globus](#), [KDE](#), and [Qt](#), in order to prime what they hope will be a self-sustaining open source project.

Two GRENADE components have been released as version 1.1 of GRENADE. See [Tools summaries](#)

Summary Keywords- Grid-Enabled Desktop Environments

Category- Data Structuring and Enhancement, Data Publishing and Dissemination.

46

ASTROGRID- A Virtual Observatory

<http://www.astrogrid.org/>

Astrogrid is an innovative multi-million pound project that is designed to offer a virtual observatory. In essence, it enables the sharing of very large multimedia files produced by various telescopes around the world. It also contains various tools with which to manipulate the astronomical images- such as a movie maker. The Grid can be accessed in two ways- via a web based portal (where a user can 'login') and via a downloadable Workbench. See [Astrogrid Workbench](#).

The Astrogrid is essentially a grid-enabled virtual research environment - it enables the rapid sharing of geographically distributed large files, and their analysis and sharing within the same overall environment.

Summary Keywords- Computational Grid, Data Grid,

Category- Data Capture, Data Structuring and Enhancement, Analysis, Data Publishing and Dissemination, Communication and Collaboration.

47

GROWL- VRE Programming and Toolkit

http://www.jisc.ac.uk/index.cfm?name=vre_growl

<http://tyne.dl.ac.uk/GROWL/>

Virtual Research Environments (VREs) are becoming important in the Arts and Humanities. The GROWL project focuses on issues which may hinder or facilitate end-user application development. In simple terms, they are trying to make it easier for a researcher to set up a VRE, and create a user-friendly client interface or software application through which to access it. See our entry in the tools section.

Summary Keywords- Virtual Research Environment Toolkit

Category- Data Structuring and Enhancement, Data Publishing and Dissemination.

48

An Investigation of the Disclosure Risk Issues Posed by the GRID, Mark Elliot,

Steven Pickles, University of Manchester

<http://www.ccsr.ac.uk/research/griddisclosure.htm>

An obvious concern regarding the Data Grid is the risk of disclosing sensitive or valuable data. This is perhaps even more true in the Social Sciences, where medical data, and politically sensitive information, is made available over the Grid. This project agrees that “GRID technology opens up a range of opportunities to enhance existing data sources and data quality to inform research, policy and service delivery. Possibilities at present include: linking data and data sets online; distributed data storage and data processing for large-scale data sets and complex analyses; data mining across different data sets; real time data updates”

This Project aims to-

- Examine the current availability of personal data over the GRID, and that projected in the near future.
- To examine the compatibility of confidentiality agreements across key anonymised data sets that could be used as part of the GRID data resources.
- To design methods of disclosure risk assessment which take account of the additional disclosure risks posed by the GRID and GRID available data.
- To implement pilot software demonstrating the disclosure risk assessment methods.

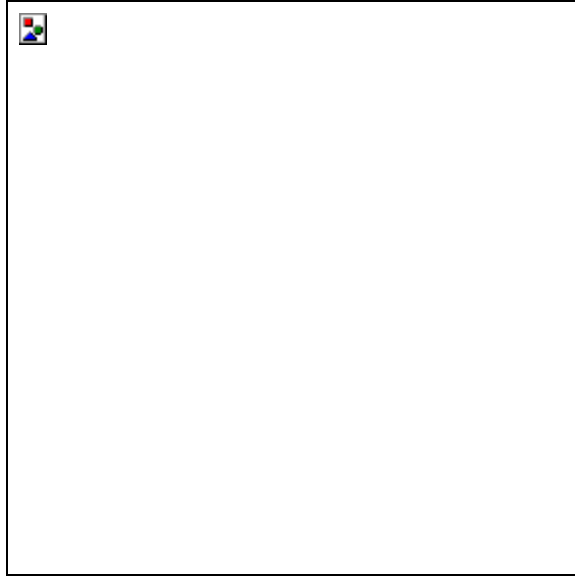
Summary Keywords- Security, disclosure risks assessment

Category- Training and Awareness.

49

Silchester Roman Town: A Virtual Research Environment for Archaeology

<http://www.silchester.reading.ac.uk/vre>



Context

The Silchester Roman Town archaeological excavation is one of the largest Roman sites in the UK. Silchester's principal data collection are contained within the York Archaeological Trust's (YAT) Integrated Archaeological Database (IADB) which is held on a server in Reading, removed from the data collection process in Hampshire and relevant experts at the Universities of Cardiff, Reading, Southampton and Oxford. The types of data involved are extremely diffuse, including numismatics, spatial site plans, matrix information, as well as the full range of data on find types and distribution. The annual excavation period at Silchester comprises eight weeks of intensive data gathering in a tight timeframe and with limited personnel. The site's excavation team has recognised the need for a more fluid and accessible network of human expertise. Many of the experts are geographically dispersed, both from each other, from the site, and from relevant information held elsewhere. This makes the exchange of ideas and interpretations, which are critical to the research process, very cumbersome; a major problem long recognised in field archaeology. This project will provide the basis for a virtual solution to it.

Aims and Objectives

The overall aim of the project is to develop a system to facilitate rapidly developing and iterative archaeological research by synchronising the three processes of gathering information, co-ordinating expertise, and managing the resulting body of data.

The specific objectives are to:

- Establish full interoperability between different IADB datasets, wherever they physically reside.

- Improve real time on-site data gathering including an assessment of the usefulness of handheld PDA devices using wireless networking (WiFi) and/or mobile phone (GPRS) technologies.
- Develop a structured mechanism for classifying thematic *Research Domains* or *Views*, which will operate across servers and across projects.
- Develop a framework for the creation of real-time online conferences involving both on- and off-site project workers and remote specialists.

Project Methodology

To explore and evaluate the use of real time on-site digital data capture, the project will run pilot projects during the 2005 excavation season utilising PDAs, smart phones and tablet PCs in combination with both wireless networking (WiFi) and the GPRS data network. The pilots will be run concurrently with the Silchester project's established recording systems. A report on the pilots will be prepared following the 2005 excavation season.

Development of the VRE will be integrated within the IADB building on existing concepts, protocols and data structures. For example, the *Research Domains* or *Views*, which will form the core of the VRE, can be seen as logical and direct developments of the IADB's *Object* data structure.

To extend the applicability of the VRE beyond Silchester, the IADB will be developed to provide interoperability, particularly with regard to *Research Domains* and *Views*, across different IADB projects and servers.

Implications/ Deliverables

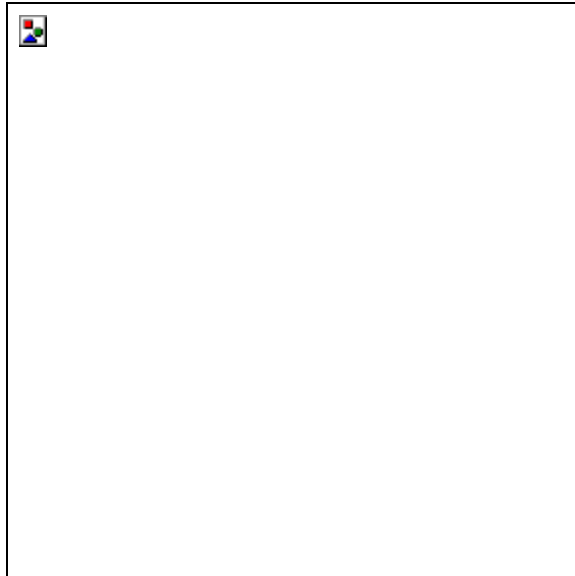
The project findings will be cycled back to all stakeholders within the Silchester project including excavation staff, post-excavation researchers and specialists by means of project seminars and practical trials during the two excavation seasons covered by the project. They will also be disseminated more broadly to the wider archaeological community through a series of reports and conference papers. The project will yield recommendations for the use of hand-held digital data capture devices on future excavations as well as a framework for a VRE for archaeology and a software implementation of such a VRE built around the IADB. Ultimately, the project will meet both the immediate needs of the Silchester excavation project, and will address a long-standing issue in field archaeology.

Summary Keywords- Virtual Research Environment, Data Sharing

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

Building a Virtual Research Environment for the Humanities (BVREH)

<http://bvreh.humanities.ox.ac.uk/>



Context

Research activities in Humanities subject fields have traditionally centred on the work of individual scholars and small groups. Their needs are now increasingly being served by the availability of networked resources which in some fields have already transformed the way in which research is undertaken. In general, however, while the scale and intensity of collaborative research activity in other disciplines have grown rapidly in step with developments in ICT, progress and awareness of similar possibilities in the Humanities have been less marked. A preliminary survey of ICT use in research projects within the Humanities Division in the University of Oxford carried out in summer 2004 has identified a number of areas of need and overlapping interest which provide a context for the creation and deployment of VRE tools for Humanities researchers. The survey revealed strong support for such an initiative and suggests that the Humanities research resources at Oxford provide an appropriate testbed for a project to investigate and identify the potential benefits of a VRE for the Humanities research community in general.

Aims and Objectives

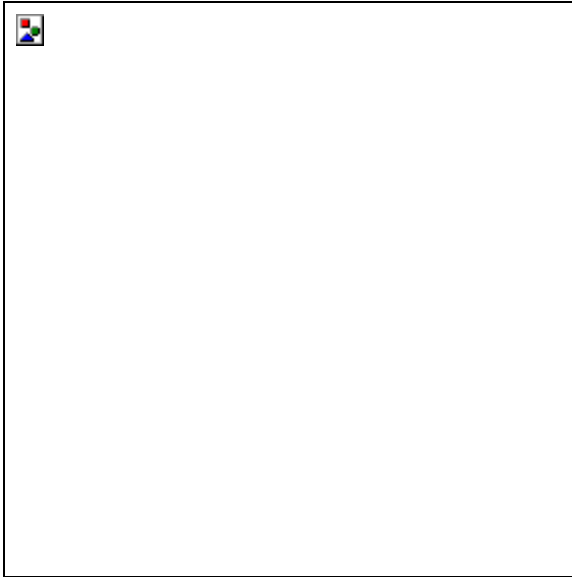
The programme aims to extend the range and depth of the preliminary survey of ICT use within and outside Oxford in order to identify specific areas in which the benefits of deploying VRE tools are likely to be substantial (for example: communication, project management and design, efficiency, access to resources) and to define those where the benefits may be more marginal; on this basis to identify specific tools appropriate for deployment as demonstrators; and, in the final cycle of project activity, to implement the demonstrators and to assess their effectiveness.

The work will be supervised and co-ordinated with a steering group in which other institutions are strongly represented. The project intends to collaborate where appropriate with the AHRC-funded ICT Methods Network and the JISC-funded ICT Awareness and Training Programme for the Arts and Humanities as well as other national Humanities services such as the AHDS, Humbul and Artifact. The proposal is designed so that the outcomes will be relevant for VRE deployments being proposed both within Oxford and nationally. In particular, the project will collaborate with the JISC-funded Integrative Biology VRE. It is proposed that applications developed as part of the demonstrator workpackage will be available locally in the first instance and, if appropriate, exposed for integration into national or inter-institutional VREs

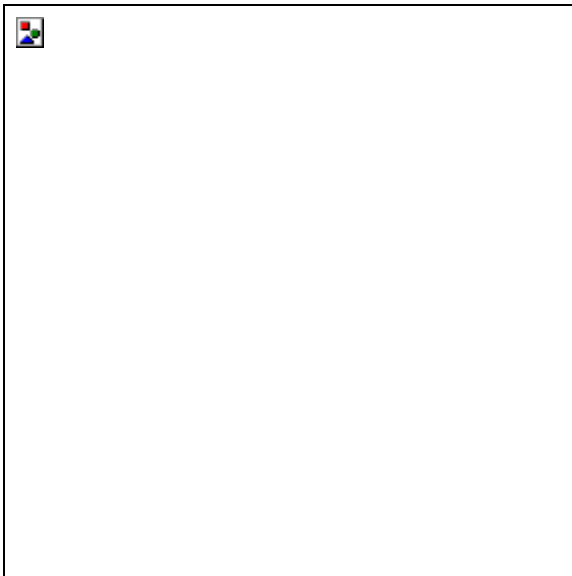
Summary Keywords- Virtual Research Environments

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration, training and awareness.

Virtual Research Environment for the History of Political Discourse 1500-1800



<http://www.uea.ac.uk/his/research/projects/vre/>



Context

Forty years ago the history of political discourse focused on the development of ‘key ideas’ in political philosophy and built on the intensive study of a limited number of canonical texts. In the mean time the subject has been, and continues to be, transformed. Amongst the developments here are:

- supplementation of the narrowly philosophical/legal approach through the study of linguistic and cultural contexts
- expansion of the range of material under consideration to include not only formal treatises but manuscript as well as printed texts, petitions, legal depositions and rulings, speeches, broadsides, satires, newsbooks, bodily gestures and iconography

- a new historical sociology of the political and administrative structures of the early modern state and its development
- fresh approaches to the discursive resources of early modern society including scribal, print, written, oral and translations
- increasing recognition of the importance of the trans-national transmission of ideas and patterns of discourse

The growth of the subject has been paralleled by the evolution in technologies. A significant body of material is either created, stored or accessed electronically. The partners in this project have made significant steps in enhancing the use of both printed and manuscript resources particularly in the teaching area through JISC funded projects such as Virtual Norfolk, Artworld and other initiatives in the humanities such as the development of the Archive HUB.

The integration of the vast range of electronic resources is being undertaken from the nationally available resources as well as from the electronic offerings of the publishers.

Consequently the range of expertise and materials for the adequate and appropriate study of the phenomena has expanded and made collaboration in research and research skills development essential and beyond the resources of any single HE department in the field.

The current proposal is to launch a continuing collaboration, initially between two institutions but with provision for the rapid expansion of participating partners.

The VRE will also provide:

- A range of collaboration opportunities, building on the recently released SAKAI framework.
- A framework for the integration of some of the resource discovery tools available in the HE community.
- Links to video-conferencing facilities
- Testing of the relevance of Access Grid technology in a humanities environment
- Seminars and publication facilities both face to face and on-line

Aims and Objectives

The project aims to develop a virtual research and research skills development environment, capable of expansion and of facilitating multiple participation in the rapidly evolving field of the history of political discourse. The environment includes the technical and cultural support necessary to deliver substantial and transferable value from the project. The project offers an opportunity to bring into being an exemplar for the humanities approximating to the more advanced collaboration models available in the sciences.

Over the three years of funding, the project will:

- Set up and pilot the postgraduate research and teaching programme (see below), enhancing the value of on-line resources and maximising the potential of dispersed academic expertise and student demand.
- Explore the potential of the ‘Access Grid’, as a potential platform for collaboration.
- Ensure that open and transferable technical platforms are in place for access to resources, virtual research environments, academic interchanges and publications.
- Deliver a set of software tools where required, guidelines and support tools to the community on an open source basis.
- Deliver a series of virtual and actual seminars, symposia and conferences led by both academics and research students
- Roll out the programme to a number of other HE institutions in the UK by the end of the programme.
- Produce reports analysing the technical and pedagogic outcomes and their implications for other VRE projects especially in the humanities.

Summary Keywords- Virtual Research Environment, Data Sharing

Category- Data Structuring and Enhancement, Data Publishing and Dissemination, Communication and Collaboration.

52

WUN Grid- Music Collaboratory

<http://wungrid.org/music.html>).

WUN Grid stands for Worldwide University Network (WUN) Grid. The network’s aim is to collaborate interdisciplinarily. WUN Grid has a particular Humanities component in its music demonstrator (<http://wungrid.org/music.html>). It offers a distributed infrastructure to explore music.

‘Music in its different forms can be processed in many ways, to the great benefit of the community of users engaged in search, analysis, research, composition, and performance. WUNgrid enables a distributed community to collaborate and share both the musical content and the services for processing it.

Grid applications in music are equally as challenging as scientific applications, making particular demands on the collaborative infrastructure. Musical information comes in a rich variety of formats, from score representation through to digital recordings, and these can be converted automatically using Grid services, and thus be available for storage, cataloguing and retrieval using Grid infrastructure. ‘

Summary Keywords- Data Grid, Middleware, Music

Category- Data Structuring and Enhancement, Communication and Collaboration.

e-Diamond Project

<http://www.ediamond.ox.ac.uk/>

eDiaMoND is a research project which seeks to show the benefits of grid technology to eHealth- in this case, for Breast Imaging in the UK.

The eDiaMoND project aims to build a national database of mammographic images. This will be achieved by building a grid community between Oxford e-Science Centre, the MIAS IRC, IBM and Mirada Solutions. By utilising distributed expertise, eDiaMoND will use Grid technologies to develop tools that will allow this database to be used in clinical diagnoses, epidemiological studies, and in training and education of radiologists and clinicians.

Of particular interest is the project's use of technology that allows for the standardisation of images as they enter the database, which helps overcome the problems that the inconsistent tagging and marking-up of images by different equipment and centres can bring. eDiamond brings together world-class medical image analysis expertise, computer science expertise, and clinical expertise within one grided environment for a collaborative and popular cause.

Summary Keywords- Data Grid, Image Analysis, Metadata, Standardisation

Category- Data Structuring and Enhancement, Communication and Collaboration.

DigitalRecord- Understanding New Forms of Digital Record for e-Social Science

<http://www.ncess.ac.uk/nodes/digitalrecord/>

This project seeks to establish a node which seeks to explore how new forms of digital record may emerge from and for e-social science and examine how Grid based technologies can be extended to provide new processes and services through which social science information may be collected, collated, and distributed. Social scientists will collaborate with computer scientists to produce 3 'driver projects'. These are:

1. A Grid-based Assembly of Qualitative Records - 'This project will develop support for social scientists undertaking social studies of technology. Its primary focus will be on the assembly of qualitative records that marry conventional data sources with emerging digital resources to better understand the social shaping of technology in interaction.'
2. Grid-based Structuring of Assembled Records - 'This project will develop support for social scientists undertaking corpus-based studies of natural language use. Its primary focus will be on structuring a Grid-enabled multi-modal corpus that combines conventional text-based representations with visual media.'

3. Grid-based Coupling of Qualitative and Quantitative Analysis This 'will be driven by social scientists undertaking studies of learning to develop techniques that allow researchers to generate, manage, and track transitions between qualitative and quantitative representations of online teaching episodes and learning outcomes.'

Their work is driven by what they call an 'iterative user-centred prototyping approach where social scientists work in close partnership with computer scientists to develop e-social science applications that demonstrate the salience of new forms of digital record to the future of social science research.'

Summary Keywords- Data, Grid, digital record, corpus, qualitative analysis

Category- Data Structuring and Enhancement, Communication and Collaboration.

55

TextGrid: Part of the German D-Grid initiative

http://www.textgrid.de/index.php?id=ueber_textgrid&L=5

The Textgrid project is about establishing a community grid. It recognises that despite modern information technology 'text scientists still mostly work in local systems and project-oriented applications'

TextGrid aims to integrate multiple distributed corpora, and include tools that satisfy the specific requirements of text sciences- within a community network. It hopes fostering this community will transform the way academics 'process, analyse, annotate, edit and publish text data'.

The installation of a grid-enabled architecture is a prudent choice for TextGrid for two reasons. Firstly, past and current initiatives for digitising and accessing texts already accrued a considerable data volume. Secondly, the dispersed expert community might benefit from the collaborative opportunities a data grid presents.

One undertaking of particular interest to humanities scholars in particular is the distributed annotation being pioneered by Text Grid. A text is shared among the community, and is subject to multiple annotations by different experts- who can add their own interpretations, and debate with fellow scholars within the text itself. After a period of time has elapsed, the some time annotated texts will be made available to the scholarly community.

Although TextGrid is a German project, effort is being made to extend it overseas, and to provide an English translation of its documentation.

Summary Keywords- Data Grid, Multiple, Annotation, Text, Community

Category- Analysis, Data Structuring and Enhancement, Communication and Collaboration.

56

Memetic

<http://www.memetic-vre.net/>

This project is funded under the JISC VRE Programme. The specific aim of Memetic is to support asynchronous collaboration on the Access Grid. It intends to-

- Provide extensions to the Compendium and Meeting Replay tools to make installation and administration simpler.
- Develop a 'robust and functionally-enhanced Access Grid recording tool
- Enable the use of Meeting Replay within an Access Grid context by integrating the above tool.
- Develop new means of automatic annotation - to provide easy and flexible navigation of AG meeting recordings.

See Compendium tool in the tools summaries

Summary Keywords- Virtual Research Environments, Access Grid

Category- Communication and Collaboration.

57

ICEAGE

<http://www.iceage-eu.org/PDA/index.htm>

ICEAGE is a large initiative designed for Grid Education. This includes all aspects of, and issues that relate to, Grid technologies. It intends to create a global forum of expertise on this matter.

ICEAGE will enable students and educators to obtain and develop Grid Education via sustained, large-scale, multi-purpose e-Infrastructures. ICEAGE will provide demonstrator projects using grids in several areas, including the social sciences.

ICEAGE will organise educational events- in practice there will be programmes of events, shared strategies, information and materials. The forum will look to attract and retain intellectual investment from a strong membership, stimulate and lead that group to work on strategic issues in grid education, ensure Their insights are well

recorded, and establish 'a critical mass 'of members, material and activities as soon as possible.

Summary Keywords- Training and Awareness, Grid Education, Forum

Category- Training and Awareness, Communication and Collaboration