

eArts and eHumanities - eScience technologies and methodologies in Arts and Humanities research

Workshop programme

This workshop brought together researchers who are exploring the use of e-Science technologies in the arts and humanities. It comprised two sessions. The first was a Birds of a Feather session - presenting the work of TextGrid in Germany, the Arts and Humanities e-Science Initiative in the UK, and related projects in the US. The second session continued with presentations from the community as well as discussed how to cooperate better on emerging standards and tools for eHumanities and eArts.

Contributions came from:

- David de Roure (Southampton) spoke about the usage of Semantic Grid technologies in musicology research support
- Andreas Aschenbrenner (University of Goettingen, TextGrid) introduced the concept of e-Humanities and the new European research infrastructure to support arts and humanities research (<http://www.dariah.eu>)
- Peter Gietz (TextGrid) presented the work at TextGrid (<http://www.textgrid.de>), a virtual workbench to support research in textual studies
- Alex Voss (NCeSS) will present on collaboration support for virtual research communities
- Mark Hedges (AHDS) reported on the work being done on grid-enabling humanities data sets at the AHDS (<http://www.ahds.ac.uk>).
- Sally-Jane Norman (Newcastle) introduced her project Associated Motion Capture User Categories (<http://www.ncl.ac.uk/culturelab/research/amuc.htm>). In AMUC motion capture technologies and Grids are used to support performance arts
- Allison Clark (UIUC) spoke about the US HASTAC project (<http://www.hastac.org>) and the Seedbed initiative.
- Stephen Beck (LSU) described the work at the Center for Computation & Technology at LSU

Summary of Workshop Presentations

There has been little to no uptake of grid technologies in archaeology, musicology or literary studies. The workshop examined early attempts by these user communities from the arts and humanities to find their ways into the world of high-performance computing and distributed data access to data. There was a clear need for a workshop like this, as there is a growing demand in these research areas, simply because the amount of data in arts and humanities is growing rapidly. The Bush administration will have produced over 100 million emails by the end of its term. These can provide the basis of new types of historical and socio-political research that will take advantage of

computational methods to deal with digital data. But for arts and humanities, an information is not just an information. Complicated semantics underlie the new multimedia archives of human reports. As a simple example, it cannot be clear from the email alone which Bush administration or even which Iraq war is hashed out. Also, new retrieval methods for such data must be intuitive for the user and not based on complicated metadata schemes. They have to be specific in their return and deliver exactly that piece of information the researcher is interested in. Arts and humanities additionally need the means to on-demand reconfigure the retrieval process by using computational power that changes the set of information items available from texts, images, movies, etc. The presentations of the day addressed these issues by looking at concrete case studies of arts and humanities research supported by advanced computational means.

In the first presentation David de Roure from Southampton pointed to the successful collaboration between musicologists and computer scientists in Online Musical Recognition and Searching (OMRAS) and the Music Information Retrieval Evaluation Exchange (MIREX). Here, a service-oriented architecture provides a solution to complicated copyright issues. In MIREX test collections cannot be freely distributed owing to copyright issues. The community is therefore experimenting with creating a service-oriented architecture to support the labour-intensive evaluation work. The International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) at UIUC collaborated with the Automated Learning Group in the development of a Music-to-Knowledge (M2K) music mining and evaluation framework. M2K has now begun experimenting with web services, so that submitters may deal with their collections locally. David de Roure also drew on his vision of an integrated collaborative work environment to point to the idea of a closer integration of grid environments with so-called web 2.0 applications – web based applications that allow communities to share distributed information and work together on them. Web 2.0 is a fast-growing community effort compared with the more standards oriented development of grids. A lively discussion of the wide uptake of web 2.0 technologies in the Arts and Humanities community was the result. The experience here should lead to a HASS-RG white paper on web 2.0 technologies.

In a later presentation Alex Voss from the UK's National Centre for e-Social Science investigated Collaboration Support for Virtual Research Communities in the Arts and Humanities. He introduced his idea of collaboration 'upperware' to ensure seamless and instant collaboration in networks, beyond organisational boundaries. Collaboration aware tools could reduce the complexity of collaboration in dynamic work environments and enable users for creative and knowledge intensive tasks. Alex Voss also challenged the idea that web 2.0 technologies could provide a solution to many grid problems. First of all, they are not necessarily simpler to use and secondly, they do not provide a solution for secure data sharing. A mashup of third party software will most likely not reassure medical scientists that their patients' data is safe.

Mark Hedges presented the UK's Arts and Humanities Data Service approach to grid-enabling humanities data. For the hard sciences fast access to large data sets is most important, in the humanities however the datasets may not be that large, but the data is highly complex, contextual, and often fuzzy and scattered, as presented above. In order to satisfy the needs of a highly diversified community, the AHDS needs to look at virtualisation technologies that improve sharing of resources as well as accessibility. The AHDS have led such problems to the flexible preservation system Fedora that supports the representation of compound digital objects and aggregations of in principle arbitrary complexity, and that allows multiple heterogeneous metadata schemas to be associated with an object. It contains built-in support for semantically representing (as an RDF/OWL graph) the internal structure of compound digital objects and relationships between objects. The architecture of Fedora is essentially service-orientated, with all functionality being exposed as web services; in particular, all data and metadata stored within a Fedora object are made available via web services. The AHDS now looks at an iRods integration for Fedora. iRods stands for Rule Oriented Data System, a successor to SRB. This might be more promising as the data grid software will be able to make use of the complex metadata stored within Fedora. Data management and preservation processes on an iRods data grid can be coded as rules constructed from and controlling smaller atomic actions called micro-services. The sequence of actions performed when executing a rule can be changed by adding, removing or replacing individual services. iRods provides an abstraction for data management policies and processes, as SRB does for storage and data. So Fedora (and other) repositories can become structured data resources within the grid and grid technologies can provide access to the contents of distributed repositories belonging to different administrative domains.

Alexander Aschenbrenner presented the DARIAH initiative and its vision to facilitate long-term access to, and use of, all European humanities and cultural heritage digital information. DARIAH is an ESFRI project for European Union Framework 7. Its aim is to bring together arts and humanities data stakeholders across Europe like AHDS in the UK, DANS in the Netherlands, Max Planck in Germany and CNRS in France. On top of the cultural heritage collections at these centers and at other similar ones in Europe, the DARIAH collective wants to build a virtual workbench for arts and humanities research that will integrate the data with services to make use of it.

Peter Gietz portrayed the modular platform for collaborative textual editing called TextGrid. TextGrid is part of the German D-Grid initiative, and is funded by the German Federal Ministry of Education and Research. TextGrid's aim is to become a community grid for the collaborative editing, annotation, analysis and publication of specialist texts. It thus forms a cornerstone in the emerging e-Humanities. Building on existing expertise in the field of e-Science and advancing towards the Semantic Grid, TextGrid partners are developing a comprehensive toolset for researchers in philology, linguistics, and related fields using grid technologies like Globus Toolkit 4 and GAT and providing grid related services via simple standard Web Services based interfaces, by

which other projects can plug into TextGrid. Reaching out to the academic community, the project establishes an interdisciplinary platform for research.

Sally Jane Norman described the challenges for arts and humanities researchers to work in an interdisciplinary project together with computing experts and biomechanics scientists. Her Culture Lab together with the North East Regional e-Science Centre and the Centre for Rehabilitation and Engineering Studies experimented with 'Data Services for Associated Motion Capture User Categories (AMUC)'. AMUC targets the tracking and capturing of motions that go beyond the everyday use of human bodies. Grid technologies provide the infrastructure to adjust motion capture data to specific user needs and to distribute it across multiple research sites. Complex, coordinated movements produced by performing artists should theoretically benefit areas that require exact measurements of human body movement like medical engineering. Here, the complex and fuzzy data typical to humanities can assist the wider research. The AMUC collaboration first worked on the exact definition of user requirements regarding motion capture data in order to develop data retrieval methods for motion capture resources via grid technologies. This work included capture and storage with a Vicon advanced motion capture system, and advanced computational methods for analysis and visualisation of such data, which comprised methods for retrieval of the data. The newly developed retrieval method does not rely on prior knowledge of metadata annotations and is not textual. Different input devices can be used from touch screens and haptic devices towards traditional computing elements. Queries can be sent to the network of databases using a web service. Furthermore, customisable indexers can help new user groups to define their specific interests and needs. One indexer can access one or more data channels and return user-identified indexable features. Future work could include on-the-fly indexing for interactive installations.

Allison Clark spoke about the Humanities Arts Sciences Advanced Collaboratory (HASTAC) in the US committed to new forms of collaboration across communities and disciplines fostered by creative uses of technology. She also introduced 'The Seedbed Initiative for Transdomain Creativity: Exploring Human Experience Through Art and Technology', which is designed to cultivate collaboration between the arts and sciences at the University of Illinois, Urbana-Champaign. Seedbed's main aims are to investigate scholarship in the digital age, creation and performance as well as changes to perception due to digital technologies. New ways of experiencing art shall be explored as well as new ways of creating it. Seedbed can help here as it offers an open forum for collaborations of philosopher, historians, or legal scholars with artists and scientists.

The last presentation came from Stephen Beck who reported on the situation at Louisiana State University. They experimented with High Definition Video Conferencing for arts support. For distributed performances, they used a high-speed network. The advantage of HD and especially prepared high-speed networks is here that the visual input can be much more satisfying. Participants can actually decipher each other's facial expressions. Latency is

reduced, but remains an issue. Even 100 milliseconds of latency can be very disruptive for a performance.

Summary of Workshop Conclusions

The activities within the arts and humanities grid initiatives around the world show that there are specific needs that need to be addressed to make e-Science tools and methodologies work within these disciplines. The early experimentation phase delivered projects that were mostly trying out existing approaches in e-Science. They demonstrated the need for a new methodology to meet the requirements of humanities data that is particularly fussy and inconsistent, as it is not automatically produced, but is the result of human effort. It is fragile and its presentation often difficult, as e.g. data in performing arts that only exists as an event. The presentations and discussions showed that grids and related technologies might provide solutions, but we definitely need to go on to build more applications like TextGrid and AMUC to support research in specific domains of arts and humanities. OGF might provide a forum for us to coordinate the work and activities and ensure that we build the necessary synergies and avoid duplication of work.

The final discussions of the two sessions were focussed on how to achieve future collaborations among the member of HASS-RG and with the wider community. OGF will hopefully provide a forum for loosely coupling different approaches in Arts and Humanities e-Science around the world. Two concrete actions were the suggestions to prepare HASS-RG white papers in two groups on the idea of a social grid where grid technologies are married with web 2.0 applications and a second white paper on emerging standards for arts and humanities grid computing.