

Mid-Term Theme Report

Title of Theme:	e-Science in the Arts and Humanities
Theme Leader:	Lorna Hughes, with Sheila Anderson, Tobias Blanke and Stuart Dunn
Start Date of Theme:	1st April 2007
Report Date:	November 1st 2007
Elapsed Time:	7 month of 12 months
Target Audience:	Arts and Humanities researchers who are using e-Science tools and methods (early adopters); Arts and Humanities researchers who are interested in e-Science tools and methods, but need exemplars/case studies in order to further develop their ideas; researchers from scientific disciplines, as well as with the computational science community, including both computer and computational scientists and software engineers and technical experts. A primary goal is to get the disparate communities collaborating and discussing their differing approaches in order to learn from one another.

Objectives:

- To present a programme of activities and events at the eSI that will provide a comprehensive, case study based, overview of the ways in which arts and humanities (A+H) researchers are engaging with e-science technologies and the emerging UK-wide e-Infrastructure in order to enhance the research that they are doing.
- To use these activities and events as the basis for a series of high-profile workgroups, and an international expert seminar, that will allow invited participants to engage in considered debates around the fundamental research questions presented by the challenge of engagement in the e-Infrastructure by A+H researchers.
- To examine the ways in which this work is embedded in research practice and research needs, in order to address some fundamental questions:
 - What are the new challenges for research in the A+H, and how is this new research agenda to be defined?
 - How can it be innovative and push barriers, yet be sensitive to those less engaged, and expand beyond the "early adopter" communities?
 - What is the international context, and how does this activity have resonance beyond the UK research community?
 - What are the strategic considerations for researchers, students and funders as this agenda is taken forward, and how are the emerging communities of practice now actively engaging with the e-infrastructure to be supported in the future?
- To develop and support the ongoing debate and dialogue between A+H researchers and researchers from scientific disciplines and the computational science community (including computer and computational scientists and software engineers and technical experts).
- To identify points of commonality between the A+H and the scientific communities who are using the e-Infrastructure, and to develop a forum for the development of dialogue, partnerships and collaborations between the communities - whilst also examining the key differences and points of divergence between the communities.

- To encourage participants in this Theme to consider not just what e-science can do for the Arts and Humanities, but what the Arts and Humanities can do for e-Science, by demonstrating that that multiple areas of commonality exist. These include:
 - The use of highly dispersed content (including large scale images, moving images, sound) joined up through the grid and appropriate tools;
 - The semantic grid and ontologies;
 - Deep data and text mining (inc. image and sound) using different methods, including deep log analysis for connecting;
 - Folksonomies; and annotation;
 - Collation, visualisation, and simulation.
- To develop a training and outreach programme in several key aspects of e-Science, which is specifically aimed at A+H researchers. Materials developed for these training sessions will be based around A+H exemplar materials and case studies, and will be developed by AHESSC in close collaboration with eSI staff. We hope that these will form the basis of the development of a "core curriculum" of training materials for e-Science in the A+H which can be re-used by eSI and partner institutions in the future. This will be an important community building and outreach activity, which will bring many members of the A+H community in contact with the eSI, and be an important legacy of the Theme.

Chronology of Events:

The centrepiece of the theme so far has been the summer series of lectures, an intensive and wide-ranging examination of the relationship between e-Science and the Arts and Humanities, and the establishment of a workshop series to examine the e-Science infrastructure and assess training requirements.

Events in the Theme to date:

e-Science Institute Public Lecture - A Potential for All: e-Science for the Arts and Humanities 30 April 2007

Why is e-Science important for the humanities? It can provide funding for tools and generic resource development, an injection of new technologies, and collaborations between computer scientists and arts and humanities researchers. Primarily, it can help meet the challenge presented by the dispersed and heterogeneous nature of the typical humanities data resource, achieving integration by means of a combination of top-down and bottom-up developments.

This lecture by Sheila Anderson and David Robey presented the current state of the art in distributed computing infrastructure for artists and scholars. A scoping survey performed in 2006 has helped to define the fundamental principles of the arts and humanities e-Science agenda, and set out the challenges that will be addressed in this eSI theme.

Methods and Technologies for Enabling Virtual Research Communities 18 June 2007

This workshop kicked off a concentrated summer lecture series for this theme. It concentrated on ways of bringing geographically scattered scholars together to form vibrant and productive groups. Tools like AccessGrid can collapse the distance between researchers, and allow them to work together as if they were all in the same building. While it can be tricky to manage an AccessGrid meeting such that all members have an equal chance to participate, careful setup, attentive management and use of the jabber facility can improve the experience. When it comes to more ambitious attempts to use AccessGrid for artistic performance, considerable further development will be needed before the technology catches up with the ambitions of the artists. The Agora multimedia collaboration tool promises to enable more fruitful scholarly collaboration in a desktop environment that combines videoconferencing, collaborative editing, and the ability to share desktops and a communal whiteboard.

Ontologies and Semantic Interoperability for Humanities Data

19 June 2007

Building ontologies in the humanities is a challenge. There are few existing examples, and they are generally quite limited, lacking the richness that full-blown ontologies promise. An ontology needs to be backed by major institutions to be widely adopted: unlike the sciences, the humanities don't have such pre-eminent institutions. Even if a system were to be given an official seal of approval, the general suspicion of top-down authority in the humanities would tend to prevent it being taken up. Accordingly, humanities data is generally gathered for a specific purpose, and not re-used.

The problems of multiple definitions of concepts and shifts in meaning over time, familiar to ontologists in the sciences, are even greater in the humanities. Our contemporary understanding of historical and cultural movements and eras can be very different from the ways they were understood by those who participated in them, and the categories we apply to order our knowledge can shift radically. There is a wealth of historical material from the Eleven Years Truce, for example – but none of the people involved knew it was going to last for eleven years. The problems of creating ontologies that are useful to and trusted by practitioners in the humanities are also substantial: with a tradition of individual scholarship, and a dearth of authoritative institutions, widespread acceptance of such grand intellectual designs can be hard to achieve.

This workshop covered some of the problems of semantic interoperability for arts and humanities data. Its central aim was to discuss existing case studies and establish a research agenda for linking arts and humanities data in a semantic metadata management system across multiple heterogeneous collections. The lecture covered an introduction to ontologies, followed by a set of examples for ontologies and annotation standards in general for the humanities. Afterwards, the scope of the investigation was expanded and other disciplines included in order to present new semantic data and information integration approaches. The last part of the workshop considered the requirements for a humanities metadata management system.

Collaborative Text Editing

20 June 2007

A key message from this summer series of lectures is that practitioners in the arts and humanities need tools that support the distinctive features of their work. Scholars and performers have priorities and concerns that are often different from those of scientists and engineers. This workshop concentrated on one of the core activities of the humanities: the production of high quality critical editions of texts. With e-Science technology, scholars can produce open source critical editions. "Open Source" is a reference to publishing and making transparent the source code as well as the source data for the decisions, with the explicit expectation that this be reused and recirculated in future studies. More than just texts, these editions can incorporate sources, supporting evidence and accounts of the methodology used in creating the edition. With collaborative text editing come questions of quality control and accountability. There is a range of possible approaches, but it is imperative that the desire to improve a work must not be allowed to supersede the need to preserve the intellectual history of a developing field.

Grid Enabling Humanities Datasets

2 July 2007

The volume of digital data generated in humanities is growing rapidly, and there is an increasing need for computation-intensive tasks, for example linguistic analysis on very large text corpora, or processing for moving image and audio data. Although data volumes are small in comparison to many scientific disciplines, the data tends to be highly complex and multi-format, context-dependent, incomplete and "fuzzy", subjective and open to interpretation, and moreover highly distributed and fragmented. In recent years, work has begun on addressing these issues, with national and international bodies explicitly recognising the special needs and importance of the humanities in regard to grids. This workshop covered recent developments in computing infrastructure and middleware specifically aimed at meeting the needs of humanities research.

e-Science and Performance

6 July 2007

The development of e-Science, and the adoption of this agenda by the arts and humanities, is helping to break down a long-standing division between science and art. What are the implications for performing and screen arts? This workshop looked back at work that took place the previous summer on developing live performance practices within AccessGrid environments, and forward to this year's eSI event on Digital Representations of Performing Arts, in order to determine the most important issues to consider.

The most important message from this workshop was that there is little point in using digital technology to do something that can be achieved by other means: artists should exploit the new and unique possibilities of the technology. Computers can be more than just tools – they can be set up to take an active part in the creative process. Finally, the importance of play was emphasised. It is vital to set aside time for free-ranging artistic experiment: a relentless focus on results can kill off creativity.

Aspects of Space and Time in Humanities e-Science

23 July 2007

This set of lectures was the prologue to the AHRC ICT Methods Network Workshop on Space and Time: Methods of Geospatial Computing for Mapping the Past, which immediately followed it. The presentations gave an overview of issues in handling spatial and temporal data, with detailed examples from archaeology and a discussion of emerging problems of trust, authority and provenance of information.

AHRC ICT Methods Network Workshop on Space and Time: Methods of Geospatial Computing for Mapping the Past

23 - 24 July 2007

As knowledge is increasingly created from the bottom up, the authority of traditional forms of geospatial understanding is eroding. For practitioners in the field, embracing this new approach may provide insights that were never before possible - but there may be dangers as well as opportunities. This workshop provided a forum for researchers from the geospatial community to examine these issues.

The geospatial mashups that Google Earth and similar web services make possible have proved to be revolutionary. Rich resources of spatially organised data can be created by the aggregated micro-efforts of thousands of strangers, and while their accuracy and reliability might be questioned, the results are manifestly useful. The consensus of the workshop was that there is no point in fighting the rise of the mashup: that battle is already lost, and researchers have to adjust to new ways of thinking about their data. This means worrying less about what people are going to do with the data, and concentrating more on getting it to them in the first place. It also means trusting that users are intelligent enough to realise there is plenty of poor quality information in the world.

Researchers are often restricted by the need to respect intellectual property rights, which can be frustrating when serious fallacies are being spread and the research community is unable to marshal the counter-arguments in public. Archaeology and history in particular are prone to being abused for political ends, or distorted in the service of fantasies. In discussions of these issues, workshop participants insisted that for all the advantages of giving up control of geospatial information, researchers still have a responsibility to assert the correctness of their own data in the face of challenge and controversy.

Agent based modelling is a significant new development that involves simulating the behaviour of people in some place and time, then integrating the results to give an overall picture of how a community might evolve. By modelling individual households, and scaling up to a community or a region, it is possible to study how a society would be affected by events such as a prolonged drought or a difficult harvest. Assumptions about the resilience of communities under these conditions can be critical in assessing possible explanations for archaeological data, and this method allows such information to arise naturally from the smallest scale. While the past cannot be re-run, it is possible to experiment on the simulation, testing which kinds of processes lead to which kinds of outcomes.

Other Activities:

We are holding a series of three e-Infrastructure meetings in collaboration with JISC funded "Enabling Uptake of the e-Infrastructure" project to scope the training requirements of the arts and humanities community:

22 August - e-Infrastructures for the Social Sciences and Arts & Humanities Meeting (Manchester)

This meeting was initiated after discussion between Lorna Hughes, Sheila Anderson, Tobias Blanke and Mike Daw in order to encourage more collaboration between e-Infrastructure work carried out for the social sciences and the A&H (including the DARIAH European project).

Attendees were: Mike Daw, Manchester, Tobias Blanke (AHDS/AHeSSC), Craig Bellamy (AHDS/ICT Guides), Neil Chue Hong (Director OMII-UK), Stuart Dunne (AHeSSC), Alex Voss(Enabling Uptake), Andrew Richards (Exec Director NGS), Matthew Mascord (Project Manager Use Cases), Mercedes Argüello Casteleiro (Use Cases), Yuwei Lin (Manchester)

The first part of the meeting was a round of presentations explaining areas of work. Presentations covered the NCeSS e-infrastructure work managed by Mike Daw, the upcoming European ESFRI project DARIAH that will link up different arts and humanities data centres across Europe. The National Grid Service and OMII UK presented on their support for e-infrastructure initiatives.

The second was a very useful discussion of areas of common interest. The conclusion of the meeting agreed that more meetings were necessary and that they should be working meetings (i.e. with resulting actions) involving people 'on the ground' to ensure that all groups work more closely together.

Research Outputs:

1. Publications

i. Journal Articles

Papers from the theme lectures have been accepted for inclusion in a special edition of DHQ (Digital Humanities Quarterly), a peer-reviewed journal of digital humanities. The description of the volume appears below:

"E-Science for the arts and humanities is a new and extremely dynamic area. The AHRC-JISC-EP SRC e-Science Initiative's first phase of activities is coming to a close, and new research projects will be announced in the Spring of 2007. This special issue of DHQ therefore presents an important opportunity to reflect on developments within arts and humanities e-science, to showcase current success stories, and to look to future work in the area. In its virtual pages, we wish to consider what has been achieved so far, what impact those grid technologies inherited from the 'hard' sciences are making, and what are the critical developments of new technologies and research within the arts and humanities disciplines themselves. What are the technological and cultural challenges currently facing the field? How can these challenges be met? Only with an eclectic and inter-professional range of expertise can we hope to approach these questions, which are by no means exhaustive. We are therefore inviting selected academics, teachers, researchers, policy-makers and administrators to contribute article-length pieces under the headings of Infrastructure, technology and collaboration, and Subject and content specific studies. Under these broad banners, we wish to give those engaged with the field an opportunity to present their research and discuss their ideas. The volume will thus reflect a wide range of viewpoints, and form not only a valuable reference work for all concerned directly and indirectly with arts and humanities e-science, but also a contribution in itself to the development of this exciting and fast-growing field".

1. Stuart Dunn (KCL) and Tobias Blanke (KCL): Editorial Infrastructure, technology and collaboration
2. Sheila Anderson (KCL): Overview: e-Science in the Arts and Humanities
3. John Unsworth (Urbana Champaign): The ACLS Cyberinfrastructure report
4. Mark Hedges (KCL): Grid enabling humanities datasets

5. Nicolas Gold (KCL): Opportunities for A&H and Service Oriented Architecture and Web Services
6. Greg Crane (Tufts): Challenges for the Perseus project
7. David Shepherd (Sheffield) and Andrew Prescott (Lampeter): Virtual research communities - TBC
8. Peter Ainsworth (Sheffield): VREs – use of the Virtual Vellum VRE on site
9. Melissa Terras (UCL): The REACH project
10. Dave de Roure (Southampton): The Semantic Web

Subject/content specific examples

11. Lorna Hughes (KCL): Overview: communities of practice in A&H e-Science
12. Gregory Sporton (Birmingham) and Angela Piccini (Bristol): Performing in virtual venues
13. Hugh Denard (KCL): The London Charter Initiative
14. Juan Garces (British Library) and Gabriel Bodard (KCL): Collaborative text editing and annotating - TBC
15. Michelle Wienhold (Statistical Research Inc. / UCL) and Stuart Dunn (KCL): Accessing complex and federated 3D geospatial materials in networked environments
16. Stuart Jeffrey (York): Aspects of space and time in mapping the past - TBC
17. Tobias Blanke (KCL): The DARIAH project

ii. Book

We are developing a proposal in the Methods Network book series, published by Ashgate, entitled: "Digital Research in the Arts and Humanities", based on the events held in the Theme. We propose to hold a seminar on e-Science Methods as part of the Theme, which form the basis of this publication, including presentations and discussion which will include the following topics: visualization, data mining, data modelling, content based image retrieval, etc. This proposal will be presented to Ashgate in the new year for peer review.

2. Reports from e-Infrastructure workshops:

AHeSSC is developing detailed reports from each workshop. These will be available on the AHeSSC website. We also envision a joint future funding proposal to JISC or other funding bodies for setting up a common infrastructure for arts and humanities and social sciences. Ann Borda from JISC has been invited to the first workshop in December to widen the publicity of the joint initiative.

3. Consultation papers

We are developing a "roadmap" for e-Infrastructure in the arts and humanities, which will encourage greater involvement of the arts and humanities disciplines in the emerging e-Infrastructure, and also promote a greater understanding of the emerging interdisciplinary research opportunities. This "roadmap" will be concluded at the end of the Theme, and all Theme events will form the basis of the content of this document.

4. Training Materials

As the Training aspect of the Theme is concluded, training materials will be made available and placed on the e-SI website at the conclusion of the Theme.

5. Postgraduate Training Materials

We will develop postgraduate training materials, based on materials produced by a number of initiatives. This will include Methods Network activities; AHDS training materials; e-Science A+H theme activities; and the outputs of the "Enabling Uptake to e-Infrastructures" project. We will organize these by subject area and methodologies, and invite a subject expert to work with us on developing these for the following areas: History; Archaeology; Visual Arts; and Performance.

Progress to Date:

The events outlined to date have been successful in that we have achieved the aims and objectives associated with each individual event, and addressed the methodological concerns related to each topic. These are outlined in detail under the section devoted to each event above (as they events cover a variety of disciplines and methods, they are described fully under each section above for clarity).

What has emerged from these events is that there is a useful debate and dialog between arts and humanities researchers and their scientific/computational counterparts that needs to be nurtured and maintained. We have experimented with using the Methods Network's Virtual Community, www.arts-humanities.net, for supporting this collaboration, and we hope to devote more attention to how we might more creatively use this tool throughout the remainder of the Theme.

In the question of developing training materials for the Theme, we have come to the conclusion that there are a number of challenges that we need to address before we can develop useful training materials that will be relevant for the Arts and Humanities communities. Our conversations to date with Arts and Humanities researchers, tools developers, and those involved in developing existing training materials has made clear to us the need for a good deal more research into the question of audience and range for these training materials (a question that is complicated still further by the need to develop materials for postgraduates). Accordingly, we have developed a collaborative approach with two other JISC funded projects that have training at their core: ICT Guides, based at King's College, and the Enabling Uptake of the e-Infrastructure Project, based at King's College, London, the e-SI and NCESS.

Accordingly, the work completed to date has made clear that here is a need to scope the existing infrastructure that supports training, and that with some additional research there are new opportunities for possible enhancements to the theme by developing a collaborative approach to training with existing e-Infrastructure services. The realignment to our original goals for training that has resulted is that instead of simply developing training sessions and running them, we are going to first scope the issue by running a series of workshops: E-Infrastructures in the Arts and Humanities and Social Sciences Workshops. These are collaborations between the Adoption of e-Research Technologies Theme and the Arts and Humanities Theme:

1. Grid-enabling data sets

December 5th, 2007

The workshop follows on from some initial discussion and a meeting in August on e-infrastructure initiatives across social sciences and arts and humanities. The purpose of the workshop is to identify common interests and overlapping areas of work. The National Centre for e-Social Science is preparing a national e-infrastructure for social sciences and AHeSSC is involved in the European ESFRI project DARIAH (<http://www.dariah.eu> <<http://www.dariah.eu/>>), which will link several arts and humanities data centres across Europe.

The workshop will cover Grid-enabling datasets, Authentication/authorisation, and content modelling/metadata. It will take place in the context of the Arts and Humanities e-Science Theme in Edinburgh. Agenda items (unconfirmed) include:

- Short introductions to the problem of Arts and Humanities and Social Sciences e-Infrastructures to give people an overview of the programmes
- Grid-enabling datasets for social science - future plans and challenges
- E-Science empowered digital repositories for the Arts and Humanities and institutions
- Case Studies:
 - Shibboleth and Grid
 - OGSA-DAI
 - STFC Data Grids
 - EDINA

Confirmed speakers include:

Tobias Blanke (AHeSSC), Lorna Hughes (AHeSSC), Mike Daw (NCeSS), Mark Hedges (CerCH), Neil Chue Hong (OMII), Andrew Richards (NGS), Pascal Ekin (NCeSS), and others to be confirmed.

2. Authentication/authorisation, Portals, Services (including e-Framework), Workflow Environments (to be organised by Mike Daw)

This workshop will cover the services and workflow tools that exist to work with existing arts and humanities data.

3. User requirements (definition, comparability, identifying representative users/samples, knowledge gathering), user support, authentication/authorisation, Grid-enabling datasets, Data resources/content modelling/metadata (to be organized by Tobias Blanke)

4. Sustainability and endangered digital resources: technical standards and user communities, organized by Tobias Blanke: Pascal Ekin, Pete Edwards, Keith Cole, Mustafizur Rahman (and others to be confirmed). This event will be organized by the Adoption to e-Research Theme.

We hope that, after these events and discussions have concluded, we will go ahead with the development of training materials as planned, but with a greater understanding of our audience, and of existing resources and materials. This means that we hope to be able to extend the Theme beyond our originally intended completion date of March 31st, 2008, and instead conclude our activities at the end of 2008.

Future Activity:

1. Lectures:

April 2008, Lorna Hughes, AHeSSC

" The future of e-science in the Arts and Humanities: the emerging strategic agenda".

2. Workshops:

Upcoming E-Infrastructures in the Arts and Humanities and Social Sciences Workshops (see above)

3. Training Sessions

Training sessions will begin in the Spring of 2008. The programme is being developed under the workshops described above, however, likely topics include:

Collaborative tools and technologies;
Motion capture tools and methods;
Visualization Tools and Methods
Data Mining tools and methods.

4. Postgraduate Training Workshops

2. We will hold a series of postgraduate workshops on arts and humanities e-Science to take place near the start of the 2008-9 academic year. We will invite their input into the development of postgraduate training materials developed as outlined above, and provide training in core e-Science technologies in collaboration with e-SI staff.

5. International Expert Seminar

An International Expert Seminar on e-Science Methods will take place in February 2008. The content of this event will form the basis of book we will develop on e-Science in the arts and humanities, as described above. The Seminar will include presentations and discussion on the following topics: visualization, data mining, data modelling, content-based image retrieval, etc

6. E-Learning Module

Currently, King's College London's Centre for e-Research is developing a new MA in Digital Asset management. Further extensions are planned to the existing portfolio of postgraduate courses, such

as a course in Cultural Heritage Informatics. These courses will be partly taught online as distant learning ones. In order to support this kind of teaching activities, we will to the end of the Theme evaluate the requirements for transforming parts of the collected materials into an e-learning module. This module would then be discussed in a follow-up working group and be published for further use in higher education.

Any Other Observations:

The implementation of the first phase of this Theme had coincided, unfortunately, with a period of great uncertainty for centralized support for ICT in the Arts and Humanities, with a decision taken by the AHRC in May 2007 to withdraw funding from the AHDS, the host organization of AHeSSC. This has had implications for the progress and structure of the Theme, as assumptions about the national e-Infrastructure for the Arts and Humanities that were in place at the start of this Theme have been challenged, and indeed, thrown into doubt. King's College, London has subsequently agreed funding for a Centre for e-Research, which will be the new institutional home for AHeSSC and a suite of other national support services for advanced ICT Methods in the Arts and Humanities (and which will continue to employ the Theme leaders). This has presented us with new opportunities to address in greater depth the questions related to support for e-Science in the arts and humanities, and to raise several important questions about the strategic direction for the e-Infrastructure in the arts and humanities. We hope at the end of the theme to have developed a "roadmap" for e-science support in the arts and humanities, and to address how an agenda for innovative research and collaboration can be taken forward as we face a severe lack of centralized support.
