

Geographical Information Systems e-Science: Developing a Roadmap

Workshop Report

Introduction

Over recent years the range of electronic data sources of potential interest to arts and humanities scholars has proliferated at a dramatic pace. Both the sheer volume of cultural heritage content has increased, as has the multimedia nature of these materials. In the past e-resources were generally restricted to electronic texts whether qualitative or quantitative. Now a plethora of material in a wide variety of formats is available including, for example, high-quality photographs of paintings, 3-D views of museum objects, sound files of regional accents, and video footage of arts performances. To date there has been an emphasis on creating new e-resources with a view that existing resources were not sufficiently comprehensive and that scholarship was restricted by the paucity of key electronic sources. In fact it is now apparent that researchers in the arts and humanities are currently not able to make full use of the current range of electronic research materials. All too frequently cultural heritage resources reside in bespoke dissemination systems, use different metadata standards, are computerised using different technologies and are made available in varying and inconsistent formats. As analogue data continue to be digitised, and increasing amounts of data are born digital, these difficulties can only become more severe unless strategic actions are taken now to develop techniques to extract information and understanding, and in turn address research debates, using e-resources.

Objectives

The objectives of the workshop were simply stated but belied the potential complexity of the task ahead. The workshop aimed to establish whether 'GIS e-Science' or 'Spatio-temporal e-Science' might be of value to scholars in the arts and humanities, and more specifically whether GIS technology might facilitate the location, retrieval and interrogation of e-resources made available through the Data Grid. The workshop aimed to identify potential exemplar projects and examine the scope, utility and

complexity in terms of implementation of these projects. It was considered important not to develop spatio-temporal e-science methodologies if these were already in place elsewhere. If a consensus was found amongst attendees the workshop would attempt to develop a roadmap to allow a GIS e-Science project to move forward. It was emphasised that any roadmap should be strongly empirically based and would be composed of a series of practical work packages.

Funding

An application to AHRC under their bespoke Research Workshop e-Science Call was made in January 2006. Under the call AHRC proposed to fund up to six workshops to a maximum fEC of £15,000. The GIS e-Science workshop received £12,500 (the maximum allowable with a requirement that a stated minimum percentage of PI salary costs should be included in the project). The funding, kindly agreed by AHRC, allowed relevant e-Science and GIS researchers from the UK, Ireland, mainland Europe, Asia/Pacific and North America to attend a two day workshop with subsistence, accommodation and transport costs met by grant funds.

In addition to AHRC funding, Information Services at Queen's University kindly hosted an evening reception for workshop attendees with the Vice Chancellor, the Director of Information Services, and the Head of the School of Geography, Archaeology and Palaeoecology in attendance. The workshop was also supported by the Spatial Technologies Research Forum – a unit established by the University to act as a interdisciplinary focus for the use of GIS and GIS related research across the University. The Forum is directed by Paul S Ell, PI of the workshop grant. This further support allowed additional attendees to take part in the workshop, and for follow-up work to be carried out.

Attendees

The workshop was organised on an invitation only basis, with the proviso that experts contacted could nominate relevant scholars for attendance. In this way we were

confident that attendees would not only be drawn from the immediate, albeit extensive contacts, of the applicant and his close collaborators. It was considered essential to attract a multidisciplinary group with expertise in one or more areas relating to e-Science (in the Social Sciences, Arts or Humanities), in GIS, in e-resource development, in humanities computing, in information management, in research computing, or stakeholders responsible for the development of key e-resources or representatives of core data providers.

The core co-ordinators of the workshop were Ell who was grant PI, Professor David J Bodenhamer and Dr Ian N Gregory. Ell is director of the Centre for Data Digitisation and Analysis at Queen's Belfast. His unit has created a number of key research e-resources in collaboration with a number of partners nationally and internationally. These resources include an Act of Union Virtual Library of materials relating to the union of Britain and Ireland; the full debates of the Stormont Parliament from 1921 to 1973 (supported by an AHRC Resource Enhancement award); A Vision of Britain through Time which contains maps, census data, historical gazetteers, travellers' tales and other material; hispop – an archive of all of the British and Irish printed census reports from 1801 to 1939, and more. Ell has used some of the e-resources developed in research which has typically used GIS to integrate, analyse and map information. Bodenhamer is Executive Director of the Polis Center at Indiana University Purdue University Indianapolis. His centre has attracted more than \$40 million over the last 15 years and has extensively used GIS to provide access to multimedia sources over the Internet. Finally, Ian Gregory is senior lecturer in Digital Humanities at Lancaster University. He has used GIS extensively in research and was the architect of the Great Britain Historical GIS, the first national historical GIS to be developed.

Attendees are listed below and were selected on the basis of the stakeholder criteria outlined above

- Bob Abrahart (Geography, University of Nottingham)
- Roy Bradshore (Geography, University of Nottingham)
- Michael Buckland (Information Management, University of California Berkeley)
- Paul Cullen (Institute for Name-Studies, University of Nottingham)

- John McDonough (Digital Humanities, UCD, Republic of Ireland)
- Stuart Dunn (AHESSC, King's College London)
- Ian Johnson (Archaeology Computing Centre, University of Sydney, Australia)
- Andreas Kunz (Institute of European History, Mainz, Germany)
- Norma Menabaney (Social Sciences Librarian, QUB)
- Gavin Mitchell (Information Services, QUB)
- Ricky Rankin (Information Services, QUB)
- Andy Turner (Centre for Computational Geography, University of Leeds)
- Humphrey Southall (Geography, University of Portsmouth)
- Deirdre Wildy (Arts and Humanities Senior Subject Librarian, QUB)
- Elles Willemse (Netherlands Historical Data Archive (DANS), Netherlands)
- Matthew Woollard (AHDS History, University of Essex)

Attendance was restricted to no more than 20 to better facilitate responsive group presentations and open discussion. In total 19 attended the workshop with the Director of IS, and the Head of Geography attending some sessions.

Programme

A focussed programme, detailed below, was developed by Bodenhamer, Ell and Gregory. This report follows the format of the workshop in following sections. All PowerPoint presentations are attached as an appendix to the report.

Monday, 3 July

- | | |
|--------|--|
| 2.00pm | Welcome
Paul S Ell (CCDA, QUB), Ian N Gregory (CDDA, QUB), David J Bodenhamer (IUPUI) |
| 2.15 | Introductions |

2.30 Session I
 e-Science in the humanities and arts
 Speaker: Stuart Dunn (Arts and Humanities e-Science Support Centre,
 King's College London)

 Defining GIS e-Science
 Speaker: Paul S Ell (CDDA)

3.20 Break
 Tea and Coffee available outside room 207

3.40 Session II: Current GIS e-Science applications in the humanities and
 arts
 Chair: Ian N Gregory (CDDA)

 Short presentations will be given by Ian Johnson (University of
 Sydney) and Humphrey Southall (Geography, University of
 Portsmouth)

5.00 Comments and close

Tuesday, 4 July

9.30 Session III: Barriers to using GIS within e-Science
 Chair: Paul S Ell (CDDA)

10.45 Break
 Tea and Coffee available outside room 207

11.00 Session IV: The potential of the use of GIS tools within an e-Science
 framework
 Chair: Bob Abrahart (University of Nottingham)

Short presentations will be given by Michael Buckland (UC Berkeley) and Paul Cullen (Institute for Name-Studies, University of Nottingham)

12.30 pm Lunch
Peter Froggatt Centre

1.30 Session V: Developing a roadmap for GIS e-Science
Chair: David J Bodenhamer (IUPUI)

Having evaluated current usage of GIS e-Science, barriers to its full development and an assessment of the tangible rewards of implementing e-Science GIS, the workshop will consider whether an implantation strategy to overcome brakes on progression and encourage judicious advancement of the technology can be developed. This will include an assessment of steps necessary to improve resource collection, organisation and access for e-resources in the future and, crucially, how to make better use of existing e-resources through, for example, ontologies and gazetteers. It will assess the necessary developments within GIS to make use of the data grid and, in analytical terms, the computational grid.

4.00 Close

Session I: e-Science in the humanities and arts

E-Science remains a relatively new concept in the arts and humanities which many scholars will not be experienced with. The workshop selected attendees based on several criteria, as stated above. Not all attendees were expert in e-Science and the workshop organisers were also conscious of significant variations in terminology between, in particular, Europe and Asia and North America. Introductory

presentations were, therefore, given by Dr Stuart Dunn (Arts and Humanities e-Science Support Centre, King's College London) and Dr Paul S Ell (CDDA, Queen's Belfast) to familiarise all attendees with the concepts of e-Science, working terminologies, and the relevance of e-Science in the humanities and arts.

Dunn focussed on the work of the Arts and Humanities e-Science Support Centre (see Appendix for a copy of the PowerPoint presentation) and related AHRC and Research Council initiatives. He adopted the RCUK definition of Grid Technologies and e-Science as

'The development and deployment of a networked infrastructure and culture through which resources - be they processing power, data, expertise, or person power - can be shared in a secure environment, in which new forms of collaboration can emerge, and new and advanced methodologies explored.'

but adapted this to

'A loose federation of methods and technologies which enable advanced research via the internet.'

Dunn described the three EPSRC demonstrator projects – 'Virtual Vellum: Online Viewing Environment for the Grid and Live Audiences' (Sheffield); 'A Virtual Workspace for the Study of Ancient Documents' (Oxford), and 'Motion Capture Data Services for Multiple User Categories' (Newcastle). He also outlined the other AHRC-funded e-Science workshops: User Requirements Gathering for the Humanities (Oxford); Performativity/Place/Space: Locating Grid Technologies (Bristol); The Access Grid in Collaborative Arts and Humanities Research (Sheffield); Building the Wireframe: E-Science for the Arts Infrastructure (UCE), and ReACH: Researching e-Science Analysis of Census Holdings (UCL).

Dunn introduced attendees to the three elements of Grid Computing. He described the role of the Computational Grid, allowing access to high-end computers and computational intensive software packages, through reference to the Archaeology Data Service. The Data Grid, facilitating access to dispersed e-resources, was

illustrated using the Jean Froissart Project as an exemplar. The project brings together the key geographically dispersed Froissant manuscripts held by a number of universities and museums allowing digitized images to be compared over the Grid. Previously researchers would have had to visit each institution in turn, and even then so valuable are the manuscripts that they may not have been granted access to them. Finally Dunn used the 'Interplay: Dancing on the Banks of Packet Creek' to demonstrate the potential of the Access Grid in allowing scholars to interact through virtual meetings.

Ell built upon Stuart Dunn's introductory paper in his presentation 'Defining GIS e-Science'. Ell, for the benefit of attendees who were not familiar with the technology, reiterated definitions for the three areas of Grid Technologies. Taking AHRC's definition of e-Science as a starting point he suggested that the Data Grid was likely to have the biggest impact on Arts and humanities research methodologies. Somewhat controversially, he argued that the Access Grid was a form of distance learning 'with a better internet connection' and that scholars were unlikely to form new research relationships based solely on this means of communications. He suggested, however, that the Access Grid would strengthen research collaborations which had already been established through conventional means such as meetings at conferences or site visits. He argued that the Data Grid had the potential to have the greatest impact on arts and humanities research. However, he suggested that the Data Grid in the arts and humanities would have a different emphasis than in other disciplinary areas. The current emphasis on the Data Grid as primarily a technology and infrastructure to allow very large amounts of data to be gathered and accessed by researchers is of less relevance in the arts and humanities where e-resources tend not to occupy large amounts of storage space. The bulk of arts and humanities e-resources tend to be text based – and large quantities of statistical data, or text in ASCII format, simply does not occupy very much disk space. With a move to more complex e-resources involving, for example, high quality images, sound files, or video footage, there was likely to be a change in the physical size of arts and humanities e-resources, but in the foreseeable future humanities datasets were extremely unlikely to rival the size of those in the Sciences. Rather, the focus of the Data Grid in the arts and humanities is concerned with the inter-linkage of disparate e-resources, a focus which is far more challenging than the issues impacting on the adoption of Data Grid technologies in

other subject groupings. Already, scholars in the arts and humanities are suffering from a 'digital deluge'. AHRC data suggests that around 50 per cent of research grants awarded result in some form of electronic resource. These outputs tend not to be in a standard format, will not have consistent and, in many cases, detailed associated metadata, and will not be available in one location.

Ell proposed that some of these issues might be resolved through using the Data Grid within a Geographical Information System based data model. He suggested that all humanities and arts datasets had three essential components – location, chronology and subject. He argued that these elements could be used to retrieve, manage and integrate e-resources within a Data Grid environment. This was made possible by advances in GIS technology which allowed GIS software to be viewed as a data management tool rather than, necessarily, a tool for spatial analysis and visualisation. GIS software had developed to make it far more user friendly than in the past, and far more able to deal with multimedia materials and, significantly, increasingly able to deal with spatial uncertainty.

Ell argued that while subject-related organisation of research materials was well developed in the arts and humanities, building on decades of work in Libraries and in Information Management, and to an extent chronological organisation of material had also developed, organising information by location had largely been overlooked. This was perplexing as all arts and humanities e-resources are referenced in space. These references may be vague or quite precise. They might reference datasets at a national level – information for the UK for example. At the other extreme they might contain a precise Ordnance Survey six-figure grid reference, or a collection of grid references to identify an administrative unit. At its most precise, these grid references may be 'date stamped' to take account of changes in administrative boundaries over time. Most commonly, however, arts and humanities e-resources are referenced by a place-name. Place-names thus form a common language in which to interrelate and associate e-resources. In fact, even when the most precise references to a location, as either a point or polygon are given, these data are also accompanied by a place-name. It might appear straightforward, therefore, to use the data grid, and place-names, to retrieve relevant e-resources across the arts and humanities. Regrettably this was not

currently the case due to the lack of comprehensive place-name gazetteers in an electronic format.

Ell identified exemplar projects which showed the potential of organising, managing and studying e-resources within a spatio-temporal GIS framework. A number of these projects had sent representatives to the workshop. The Electronic Cultural Atlas Initiative (ecai.org) based at UC Berkeley had around 1,000 scholar affiliates in the arts and humanities with interests in examining spatially referenced data they held in context with other e-resources. ECAI had developed metadata that allowed registered datasets to be retrieved by users in real time within a distributed environment. The group had also developed software – TimeMap – which allowed retrieved data sources to be visualised. The Vision of Britain Through Time project demonstrated how it was possible to transform a ‘traditional’ historical GIS project composed of time-variant administrative boundaries and census data into a multimedia resource containing historical maps, historical texts and historical gazetteers. The system had the potential to incorporate virtually any spatially referenced data and allowed these data to be interrogated by location, chronology or subject.

Ell stressed that to develop spatio-temporal GIS e-Science basic e-infrastructure would be required and in particular place-name gazetteers, chronological look-up tables and subject indexes. He stressed that enhanced metadata for source material would be required or ‘intelligent’ context-sensitive searching. He also asked the workshop to consider whether a geo-temporal data browser might be required.

Session II: Current GIS e-Science applications in the humanities and arts

In this session there was discussion of early developments of distributed multimedia spatio-temporal e-resources, their software environment including middleware to assist with access and integration via the grid.

Representatives were present from the Electronic Cultural Atlas Initiative (ECAI) and the developer of the associated TimeMap geo-data browser and the Vision of Britain Through Time project.

Dr Ian Johnson (Archaeology Computing Laboratory, University of Sydney) presented a paper on TimeMap and mapping and GIS (see appendix for PowerPoint slides). He described a number of projects which made maps available online interlinked with other materials. He also outlined the development of TimeMap and summarised its aims and functionality as:

- A means of displaying the results of scholarly research in space and time
- A means of making text, images, hyperlinks, tabular data, and multimedia applications from anywhere on the Internet available in one interface
- A tool to query the data catalogued in the ECAI clearinghouse and search the Internet to find and access ECAI projects
- A way to superimpose diverse ECAI projects on one another in a single time-aware map
- A means of downloading selected data from Internet datasets to a consistent GIS format on a local drive for further analysis with other software packages
- Control of access to datasets through password and encryption
- Methods of display and animation of maps through time

Dr Humphrey Southall (Geography, Portsmouth) described the development of the Vision of Britain Through Time project from the first comprehensive national Historical GIS to a user-friendly multimedia resource. The system contains more than 11.5 million statistical data values drawn from the HGIS. This is supplemented by descriptive gazetteers, images of historical maps including Ordnance Survey series, the General Reports to every census and 13 Travellers' Tales. He made clear that constraints were placed on the project by the funders – the NOF Digitisation funding stream. As a result a widely accessible web based resource had to be created that did not include plug-ins, javascript, or ActiveX components and the site had to work with a low resolution monitor and a non-broadband internet connection. Southall described the data model developed for the Vision of Britain site. He focused on When: Date Objects, Where: Gazetteer and What: Data Documentation System.

In discussion there was broad agreement that the projects highlighted, and the work carried out by the Electronic Cultural Atlas Initiative, did have considerable potential.

The 'what, when, where' structure of organizing information were regarded as key and it was felt that 'what' and 'when' elements could be developed for humanities and arts e-resources fairly easily. The 'where' element would remain challenging however. There was discussion on whether place-names were key to spatially referencing resources or whether some other system might be used such as a grid reference. It was agreed that the vast majority of arts and humanities research material was generally reference by a place-name. It was accepted that place names changed over time both in terms of the names themselves and in regard to the spatial footprint they occupied. Birmingham in 1801 would have, for some data sources, a radically different and far smaller spatial footprint than Birmingham in 1971. It was also considered that only on relatively rare occasions did humanities sources refer to a very specific location. Census data was a clear example of where information did pertain to a clear spatial unit/area. Generally though place-names were used to refer to a location in more colloquial terms. It was suggested that in many ways the colloquial expression of location common in the humanities might allow space to be dealt with more easily. Traditional HGIS projects were extremely challenging to develop because there was a clear need to develop a precise spatial footprint for each census tract and these tracts changed over time. A colloquial approach did not require this level of accuracy.

Session III: Barriers to using GIS within e-Science

This session examined the barriers to using a GIS approach to manage electronic materials over the Data Grid.

It was agreed that a system for resource discovery and data integration would be exceptionally useful to researchers in the arts and humanities and that the projects described in Session II illustrated the potential of such a system. However, they also illustrated the challenges in that significant work needed to be carried out with e-resources to ensure that they could be used within a geo-temporal framework. TimeMap required datasets to have specific metadata applied at object level. This metadata was non-standard and was likely to take considerable time to implement for large and complex sources. To a degree this complexity was marked by the relatively

small number of datasets that had had suitable metadata applied and were, therefore, available to TimeMap. In the case of the Vision of Britain project much manual work had been done to ensure that materials were linked correctly by location.

It was agreed that the data model – that at the most abstract level all things can be considered as objects located at geo-referenced points with spatial and temporal markers– was applicable but that spatial and temporal expressions were non-standard often relying on colloquialisms. To deal with these colloquialisms a series of, in effect, comprehensive look-up tables were required. Some place-name gazetteers already exist such as the Getty Thesaurus of Geographic Names (TGN) and most AHDS data holdings are referenced by location. However, existing digital gazetteers were felt to be inadequate for arts and humanities scholars, particularly those with interests in historical data as few gazetteers incorporated changes in place-name over time. However, the volumes of the English Place-Names Society, which contain detailed time-variant place-names down to field level might prove to be an invaluable resource. These are currently not comprehensively available on line.

Session IV: The potential of the use of GIS tools within an e-Science framework

In this session discussions were focussed on the potential ability to interrelate distributed information within a single software environment, the availability of powerful retrieval and analytical tools, and the capacity to cope with almost any conceivable form of data. Attendees felt that GIS software was suitable as an environment in which humanities researchers can conduct the bulk of their e-research. It was felt that this might not involved the use of off the shelf GIS software but the use of software, possibly specially developed as with TimeMap, to provide GIS-like functionality in terms of data management, integration and visualisation. In parallel with theoretical discussions of the potential of a GIS e-Science agenda the application of the technology to specific research debates and agendas in arts and humanities disciplines was considered with particular attention given by workshop participants to their own areas of research.

Short presentations were given by Professor Michael Buckland (UC Berkeley) and Dr Paul Cullen (Institute for Name-Studies, University of Nottingham)

Professor Buckland began his presentation with the following comment: “This proposal is intended to result in significant methodological advances in arts and humanities scholarship through the integration of two key developing technologies e-Science and Geographical Information Systems.”

In terms of focus we are dealing with an intersection between two fields with very different characteristics:

- The humanities are concerned with PLACES and GIS with SPACES.
- Spaces are physical constructs; Places are cultural constructs. It is culturally easier to deal with imaginary places than imaginary spaces.
- But, actually, we are not directly concerned with either arts and humanities or GIS, but with some relationship between them.
- Place name gazetteers are pivotal both for illustrating the difference and for linking them, like a bilingual dictionary.

There is a large agenda for tool development. It is remarkable how undeveloped (and how unready for e-Science) relevant tools are. Based on experiences at UC Berkeley:

- Still (in 2006) no national or international standards for place name gazetteers. (OGC is said to be discussing this.)
- Time period directories, a tool for mapping between eras and calendars (another dual-naming system analogous to place and space). One was constructed at Berkeley, but this may be unique.
- Biographical dictionaries and “Who’s who”s summarize individuals’ activities through places and time. They are a long established genre but there appears to be no documented “best practices”, let alone standards.

In terms of content there are many “low fruit” opportunities for drastically improved e-Science content but caution is advisable.

- The NGiA gazetteer indicates no golf-links in Scotland. Remedy, share existing digital data.

- There were no “Agricultural labourers” in Great Britain before 1841 according to Matthew Woollard, though clearly there were. This is a terminological issue.

- Language maps (both language area maps and isogloss maps) are just one example (like place name gazetteers) of rich opportunities for digitization in the humanities that not only move material into a digital (e-Science) environment and allows originally intended uses to be performed better (comparing different isogloss maps) but opens up new opportunities (relating language map layers to potentially related non-linguistic map layers. (Note that language maps are a surrogate for cultural maps.)

With regard to terminology the arts and humanities are massively and centrally concerned with interpretation and representation. With digital resources, natural language processing can aid the finding of connections and mitigate language problems. The humanities is rich in monumental resources, especially in the U.K.: Victoria County Histories; English Place Name Society publications; Migne’s Patrologia (Writings of the early Church fathers); Dictionary of National Biography; and on and on. Each of these has its own particular categorization and indexing practices. To make these resources effectively accessible requires (1) Serious provision and explanation of metadata and terminology; (2) Scope for computer-aided mapping of equivalences between different vocabularies.

Paul Cullan’s in his presentation described a source that would, once digitized and adapted, form an essential element of e-Science infrastructure – the work and publications of the English Place-Names Survey (EPNS). Since 1924 EPNS has been engaged in the painstaking collection and analysis of all the country’s place-names, including the names of administrative units, settlement sites, topographical features, field-names and street-names. These names are extraordinarily revealing about the cultural and social patterns of English history: the suffix ‘by’, for instance, can be seen to chart the Scandinavian settlements of the ninth and tenth centuries, while Celtic survival in Anglo-Saxon England is marked by recurrent instances of Walton ‘the settlement of the Welsh’. Topographical names and field-names record changes in landscape (e.g. extent of former woodland) and land-use. And so on. Fortunately for modern scholarship the EPNS has systematically collected place-name spellings from a wide array of textual sources: these spellings are arranged chronologically and

related to their modern forms (where these survive). Shifting spellings of the same name are dated; wholesale replacements of the name are recorded. The history and derivation of the name is discussed and where possible explained. This material is printed in, to date, over 70 volumes of the Survey of English Place-Names, which are highly regarded as standard reference works. They are of immense value to the study of history with either a national or local focus, language, geography, archaeology, family history and genealogy, literary studies, cultural studies and many other disciplines that need to understand the geographical and environmental contexts of the subjects they study or the places they live and visit. Few reference works can claim such a wide audience. The critical importance and indispensable nature of these volumes can scarcely be overstated. Nor should their ability to interlink disparate sources by location-name.

Session V: Developing the RoadMap

The final session of the workshop was devoted to development of an implementation plan to realise the potential of spatio-temporal e-Science to link research materials. The use of space and time to link disparate e-resources within a distributed infrastructure offers great potential of success, in large measure because of an increasingly large set of digital objects and the recent development of key framework elements such as spatio-temporal gazetteers and crosswalks. These elements, discussed in earlier sections, are prerequisites for effective data integration. Their emergence makes it possible to consider an agenda or roadmap for future work.

To ensure that the agenda fits a research need, it is necessary first to state the problem correctly. It may be summarized as follows:

1. Research in the arts and humanities is moving inexorably toward digital investigation and analysis, using data, text, and images accessible via the Web and other e-repositories.
2. In anticipation of this shift, funding bodies in the UK and elsewhere have underwritten the conversion of a large and rapidly growing number of strategic materials from traditional forms into e-resources.

3. These e-resources currently exist in different formats and are difficult to integrate. Current search tools are inadequate to support investigation across differently formatted data without extensive metadata development, an unlikely outcome.
4. The absence of intuitive, effective cross-searching tools is a major contributor to the under-utilization of e-resources and is a barrier to the interdisciplinary work that is at the heart of future scholarship in the arts and humanities.
5. It is not enough simply to identify e-resources, scholars require an easy-to-use means of drawing discovered data together and managing it for purposes of analysis and the creation of knowledge.
6. To date, the potential for using location as the common link among disparate resources remains untapped, as does the use of spatial technologies within a distributed environment to integrate, manage, and disseminate the products of cross-searching.

With the problem in mind, it is possible to develop a vision to guide the development of a proposed e-Science initiative to unlock the potential of e-resources for research. This vision has four main parts. We will tap the power of existing and emerging spatial technologies within an open standards and open source e-Science grid to enable:

1. Effective e-resource discovery
2. Integration of e-resources
3. Management of e-resources
4. Dissemination of knowledge developed through the enhanced use of e-resources.

The most pressing problem is to develop a means of effective e-resource discovery, and it should be the first step toward implementing the vision outlined above. The first project proposed under the RoadMap is to use existing and emergent tools to create an expert means of cultural resource discovery as defined by space and time for UK e-content. The audience for this project will be arts and humanities researchers. Its initial target for cross-searching will be e-content developed by UK-funded projects, but it also will aim to facilitate discovery among all e-resources relating to the UK, regardless of its provenance. Its focus, more precisely, will be to allow

expert linking of e-resources by the attributes of what, when, and where. E-tools to identify the many variants of these attributes already exist or are in advanced stages of conception. These tools include gazetteers, place-name indexes, temporal translators, and subject indexes. Geographic information systems (GIS) offers a suitable platform for integrating these e-tools, but it may not be the only platform useful for this purpose. Regardless of the framework chosen, it must be operate on open standards and it must be open source.

Implementation will require a full scoping exercise, using Volare methods or another reliable specifications development process. It will be critically important to identify user requirements and to refine these requirements through iteration during the development phase. The project should begin with the JISC-funded work currently underway at Oxford University to understand user requirements in the humanities (Building a Research Environment for the Humanities, <http://bvreh.humanities.ox.ac.uk>). Among the requirements suggested by the GIS workshop participants were:

- Intuitive, common-language interface for all e-resources
- Return relevant, reliable information in easily understood form
- Must link directly to resource

Although user requirements may influence some system requirements, a discovery tool should be built on the following functional specifications:

- Uses distributed Web services
- Uses existing resource discovery tools
- Searches existing e-resources
- Works with existing metadata
- Modular and extensible architecture
- Stable and sustainable
- Accepts expert-user comment
- Adaptive

The key to effective cross-searching among disparate resources is to generate linking identifiers dynamically. Extensive metadata is the standard way to locate information currently, but most datasets have sufficient metadata only to allow searching at the

collection level, whereas the ideal system would allow it at the object level. The work of ECAI provides an exemplar of the metadata required to facilitate object-level documentation and object harvesting using TimeMap. This level of metadata is unlikely, however, to be applied to most datasets. It is far too costly to create the metadata required for such deep linking, so the proposed e-discovery system uses location or location + time or location+time+subject as linking elements, which means it must use various discovery tools to provide the spatial, temporal, and subject tags to satisfy these criteria. These tools will operate as Web services. The system will allow users to access these services to search any digital resource, although in its first iteration the resources must be Web-accessible and, for the beta version, they will be e-resources made available through AHDS. We note, that subsequent to the workshop, continued support for AHDS by AHRC may not be in place beyond May 2008. As a result the test-bed data may need to be found elsewhere. If AHDS does not remain as a key archiver of arts and humanities e-resources the role of the resource discovery tools proposed here will be even more significant.

The proposed system will be built on a modular and extensible architecture to accommodate growth and new user requirements, such as the integration and management of discovered data. The system also must be sustainable, which will require the development of a business case. An ideal system will be adaptive, or capable of learning from the uses made of it, much as Amazon.co.uk returns recommendations based on the activity patterns of other users. Finally, the system should be capable of accepting expert commentary on the sources, the methods of locating information, or other information that can be compiled automatically into a knowledge base, perhaps in the form of a wiki, as well as serve to guide further development of the system.

Conclusions

The workshop proceeded as envisaged attracting an international group of experts to review the potential of e-Science to revolutionise the use made by scholars in the arts and humanities of electronic research resources. Of all of the Grid related technologies, it is likely that the Data Grid will have the longest long term impact.

Our proposals are not intended to require a fundamental change in the way scholars conduct their research. Rather we foresee scholars continuing to work as they have in the past. The workshop findings would suggest that spatio-temporal e-Science has the potential to deliver the basic source materials upon which research is based to the scholar more effectively, more reliably and in greater volume. The process of using e-research resources will be eased and the research of scholars in terms of its breadth and quality will improve.

Paul S Ell

Director

Centre for Data Digitisation and Analysis

Queen's Belfast