Symposium

A Practicable Future for Computing in the Humanities: An International Symposium.

The University of Newcastle, July 2001

Background material

- **Abstracts of papers**
- **Biographical notes**
- **Symposium Program**

Abstracts

Keynote Humanities computing is being reborn. No longer need we be satisfied with computer-based equivalents of established scholarly procedures in our several disciplines. We are now in a position to open up new and previously unimagined paths. Our speakers will substantiate this claim.

Putting humanities electronic resources into the digital library
Susan Hockey
For much of the time since its inception in the late1940s until about 1990 humanities computing was largely the province of individual scholars or projects doing their own thing, firstly for research and then also for teaching. Typically a scholar would create an electronic text and then subject that text to different kinds of analyses, most often with a concordance program and perhaps some custom programs. The electronic texts, created as the by-products of these individual research projects, usually reflected the theoretical viewpoints of the scholars carrying out those projects. Some of these resources have found their way into archives or digital libraries, but issues of reusability really only came to the fore in the late 1980s, when it became generally accepted that a common encoding format would make it much easier for researchers to exchange electronic texts. The TEI project was formed to create this format. It did achieve a good measure of success in spite perhaps of being ahead of its time.
In the 1990s the situation changed rapidly with the widespread use of the Internet and, more specifically for this discussion, the efforts being made by many libraries to digitize their collections. Over the next years it seems likely that the world's major library collections will be converted into electronic format. In many ways it makes sense for electronic text technology to move into the library world, given the expectations of permanence of institutional structures and of information within libraries and archives, but it is perhaps disappointing to many of us in humanities computing that most of these digitization efforts largely ignore the intellectual advances that we have made. Many libraries seem to see the end-product of their efforts as collections of digital images on the Internet, accessible via the catalogue or perhaps some electronic finding aids. A few have been creating full-text transcriptions of documents, but these transcriptions tend to be treated as a kind of reference work –

the only functionality provided in the software consists only of looking up words and phrases. This leads to the major question I would like to address in this seminar: Can the tools, techniques and intellectual advances made in specific humanities computing projects be generalized to form part of a global digital library for humanities scholarship? That library clearly needs to maintain all the scholarly standards that we would expect in high-quality non-electronic scholarship, but what else should it be able to offer in the electronic world? Is it feasible, or indeed necessary, to try to bring together a variety of advanced tools and theoretical perspectives into a general purpose system? Furthermore, the Internet opens these resources to a much broader community. Can the needs of schools, lifelong-learners and 'the citizen' be met by the same digital library system?
Some specific issues:

1. HTML has introduced many more people to the concepts of structured markup, but to those of us who have worked with markup technologies for some time, it is a step backwards, certainly at an intellectual level. But new markup schemes based on XML are difficult when it comes to the meaning of tags and, after all, it is the meaning that determines how the user understands and uses encoding. Will we ever arrive at a global encoding scheme for digital humanities material that really satisfies all needs? It seems that it must then include a substantial amount of interpretive encoding, but if it does, how can we make sure that users from different communities understand what that encoding means?
2. What kind of metadata is appropriate for the humanities digital library? Traditional cataloguing methods do not really translate well into the digital environment, but they do provide an overarching format for the hybrid library. How do new initiatives such as the resource description framework fit the needs of the humanities? How can we use metadata to help make a seamless link between locating the item and analyzing it with software tools?
3. Looking in a different direction, in the 1990s research in computational linguistics moved very much from speculation about 'toy' sentences to the analysis of real-world data in the form of corpora. These analyses began with concordances (tools used for a long time in humanities computing), but statistical modelling and linguistic analysis tools are now routinely used for analysis of modern spoken and written material. In only very few cases have they been applied to humanities source texts. Would this approach work for literary material? Is it something to be built into the humanities digital library – as well as into tools for searching journals and other secondary material, as is now happening?
4. What will happen when more scholars write for the electronic rather than the print medium? Linking mechanisms in HTML are weak. If we have pieces of XML-encoded data on the network is it feasible for scholars to grab or point to the pieces that they want, and then mix and merge them in new forms of publication together with annotations and other links?

A research topography for humanities computing
Willard McCarty

Jonathan Culler, in Framing the Sign: Criticism and its Institutions, argues that contrary to the "myth of foundationalism", vigorous dispute rather than unanimity characterises the healthy beginnings of an academic discipline. So, it seems, humanities computing is off to a good start, with little ontological consensus about it.

Its vigour is attested more than negatively, by activity which has spread from beginnings in philological, literary and linguistic studies to all areas of the humanities. But does this activity have a scholarly centre?

This session of the Symposium will begin with an argument for such a centre and will be devoted to exploring a provisional topography for it. Although the session will not deal with the details of institutionalisation, the proposed topography can be implemented. Indeed, it is derived from experience in a functioning academic department at King's College London. It is quite clear that humanities computing is not a discipline, whatever we may ordinarily mean by that. It makes no sense as yet another isolated field of its own. Rather it is best conceptualised as an interdiscipline, i.e. an academic field symbiotic with and inseparable from research and teaching activities across all the disciplines of the humanities. The unifying perspective that makes this symbiosis possible is methodology, i.e. "the science of method" (OED), in particular the study of directions and implications in an area of research. On the one hand we study praxis in the humanities as it translates into mechanical form, on the other the implemented methods of computer science as these answer to problems in the humanities.

The results speak to the humanities as a whole because, as experience has shown, the disciplines share computational techniques by virtue of shared data-types (discursive and tabular text, image, sound). In other words, the centre of activity is a methodological commons. In this commons humanities computing serves as trader in techniques, exporting them from their disciplinary origins, systematising them, seeing to their adoption elsewhere and developing a common language of interchange. Such collegial service is powered and informed by a scholarly practice constituted by enquiry into the gap between what is known in the disciplines and what may be demonstrated by mechanical methods. Thus humanities computing views applied computational methods as modelling -- i.e. the use of manipulable simulacra in order to study phenomena we cannot reach directly -- and so derives its research agenda from the failures of well-crafted attempts to model what we know.

In the session a tripartite taxonomy for its research activities will be proposed, based on the uses of automatic methods, metatextual encoding and re-presentational transformations of humanities data. In the introductory talk each of these will be defined and its specific areas of research identified and illustrated. Participants will then be invited to put this taxonomy to the test of their own experience, and particularly to reflect on the interdisciplinary perspective of humanities computing in relation to their own disciplinary orientation.

The primary objective of this session will be to establish humanities computing for the Symposium as fundamentally of as well as in the humanities -- and to provoke healthy dispute about its nature. The proposed taxonomy is intended to be a tentative instrument toward a better articulation of what makes the field scholarly the raising of better questions, not definitive solutions.

Text and Markup: More illumination in one lifetime than humanists deserve.
Allen Renear

I will begin by telling a story to illustrate my contention that the interdisciplinary and methodological nature of humanities computing has already resulted in extraordinary new intellectual achievements, and that futhermore these have been achievements in areas that seemed so highly evolved, as disciplinary fields, that qualitatively new results were thought to be unlikely or at least rare. The basic story, if not the inference, is actually already well known to computing humanists: it is the story of the

emergence of SGML markup, particularly in the context of the Text Encoding Initiative, as device for stimulating and structuring qualitatively new and distinctive investigations into fundamental features of textual communication. In retrospect these new directions have yielded more illumination than humanists are accustomed to expecting. We now know a lot more about textuality, and rather profound things too, than we did in, say, 1960, and there is more to come.

This contention is of course in opposition to the pessimistic – but well-reasoned -- appraisals of humanities computing by credible leading figures such as Tom Corns and Mark Olsen. If persuasive, my analysis should bring some relief to those of us in this community who may be been uncertain of ourselves. However it brings a challenge as well. For I am serious when I say that humanists don't deserve the illumination they have gotten. For the most part they remain quite set in their familiar ways -- deploying tired (however various) analytical strategies and rhetorical practices, that, when combined a faint-heartedness (or is it diffidence?) about engagement on unfamiliar ground ensures that promising wonderful new lines of research will be mostly fumbled rather than developed. At the heart of this are, I think, some hard problems about the nature of interdisciplinarity. I hope that the members of this seminar, if they have followed me this far, will help me say just what those problems are.

Formal Models, Digital Scholarship, and Hybrid Libraries
Harold Short

Many of the computing applications most widely used in arts and humanities research share at least three characteristics that at first sight would seem to make them inimical to humanistic research: precision, explicitness and consistency. In many cases, perhaps even for the most part, the data that form the object of study for the arts and humanities researcher has exactly opposite characteristics. How is it then that such applications may be used effectively in such research, to create new forms of scholarship and new forms of scholarly output, bringing new insights and opening new avenues of enquiry?

In this session I will argue that the use of applications such as relational databases and text mark-up and analysis imposes the necessity of more rigorous modelling of the data than has usually been required in pre-digital methodologies, and that while the rigour and narrowness of this modelling involves the loss of significant data, at its best it pushes the researcher to new and sometimes clearer or deeper understanding of the source materials. This new understanding proceeds in an iterative fashion, in each of the stages of analysis, design, implementation/data entry, and retrieval/output, with new insights or patterns leading back at any time to renewed analysis or new design.

In pursuing the argument, I will use as specific examples some major projects at King's College London, in fields such as prosopographical studies, art history, and cultural history, and will refer to others, e.g. the work produced over a number of years by Fellows at the University of Virginia's Institute for Advanced Technology in the Humanities. The projects involve different types of source data, have differing research objectives, will be aimed at a variety of audiences, and employ a range of technologies. For the purposes of the seminar, they provide a framework in which to ask questions about how data, purpose and technology inter-relate and how these shape and are shaped by formal modelling methods.

The 'digital resources' that are the end-products of projects such as those illustrated are often large-scale, multi-source, multi-discipline and multi-technology - all characteristics of what John Unsworth has called 'second-generation digital resources'.

Their creation typically requires new modes of collaboration, between scholars, and between scholars and technology specialists. The integration of their content and their technologies is often a major challenge, with implications for the initial data modelling processes, as well as for the design of the end product. The presentation will conclude with consideration of the 'publication' of these digital resources. Without pursuing too far the question of what 'publication' now means, there are important issues related to how the resources may be located and used, if they are to take their place as new tools of research and teaching in the burgeoning 'hybrid library' – hybrid in terms of media and of location. These issues include resource identification and description ('metadata') and uniform access ('interoperability'). They too have implications for data modelling and resource creation which should not be overlooked.
The session will have a very practical focus, and those attending will be invited to share their own experiences as well as to engage with the views and materials presented.

Knowledge Representation as a Core Activity of Humanities Computing
John Unsworth
"Knowledge representation is the application of logic and ontology to the task of constructing computable models for some domain. . . .

1. Logic provides the formal structure and rules of inference.
2. Ontology defines the kinds of things that exist in the application domain.
3. Computation supports the applications that distinguish knowledge representation from pure philosophy."

[from the preface to Knowledge Representation by John Sowa]
Humanists have always engaged in knowledge representation--by constructing taxonomies and other systems of classification, by indexing and cataloging, by constructing encyclopedias and other reference works. With the application of computer technology to the subject matter of the humanities, though, knowledge representation moves to center stage in a broad range of humanities disciplines, because of the need to provide formal statements of the knowledge of these disciplines in order to make that knowledge tractable for the purpose of computer processing and analysis. Knowledge representation in the humanities involves logic, philosophy of language, visual representation, bibliographic methods, information design, visual and textual models of epistemology, aesthetics and metaphysics of form. This presentation will examine, in detail, one concrete example of knowledge representation in humanities computing from a project at the University of Virginia's Institute for Advanced Technology in the Humanities, and will discuss the role of knowledge representation in the new MA in digital humanities at the University of Virginia.

Language Resources and Humanities Computing:scientific, technical, strategic relationships
Antonio Zampolli
After a period of about thirty years, in which the role of lexica, corpora, and large grammars was practically ignored in the field of Natural Language processing, the paradigm of LRs (Language Resources) has progressively gained ground in the context of language industry to the point where some surveys identify NLP with the statistics-corpus based approach.
We will analyse five major axes of development:

- The trend towards the specification of de facto standards for LRs
- The creation of large LRs coordinated among different countries/languages
- The identification and setting up of mechanism for the distribution of LRs
- Research and experimentation of methods for automatically acquiring knowledge from text
- The specialisation of general www meta-data for the classification of LRs and coordination among national projects.

The scientific and strategic goals of projects like EAGLES, ISLE, PAROLE, SIMPLE, ELSNET, SPEECHDAT, SALA, ENABLER will be discussed. Methods, standards, tools, techniques, data developed in this paradigm have immediate applications in various humanistic disciplines, offering not only the possibilities of new scientific insight, but also the opportunity to link the humanity resources and disciplines to the cultural, economic, social, strategic activities characterising the contemporary ICT based society.

Biographical sketches of session-leaders

Susan Hockey, Professor and Director, School of Library and Information Studies, University College, London.
Previously Professor and Director of the Canadian Institute for Research Computing in Arts at the University of Alberta (1997-9), and Director of the Center for Electronic Texts in the Humanities (CETH), Rutgers and Princeton Universities (1991-7), where she founded and co-directed (with Willard McCarty) the CETH International Summer Seminar on Methods and Tools for Electronic Texts in the Humanities.
Chair of the Association for Literary and Linguistic Computing from 1984-97; member of the Steering Committee for the Text Encoding Initiative, twice serving as Chair of that Committee. Author of four books, the most recent being *Electronic Texts in the Humanities: Principles and Practice* published by Oxford University Press in December 2000, and numerous articles on text analysis, humanities computing, text encoding and digital library technologies for the humanities.

Willard McCarty, Senior Lecturer, Centre for Computing in the Humanities, King's College London.
Vice-President of the Association for Computers and the Humanities. Codirected the CETH Summer Seminar, "Electronic Texts in the Humanities: Methods and Tools", at Princeton University (1991-96), with Susan Hockey. Designed and taught a certificate programme in humanities computing for graduate students at the University of Toronto, Canada (1990-96). Since 1987, with a five-year hiatus, has served as founding editor of the online seminar *Humanist*, a forum for discussion of all matters related to humanities computing.
Co-editor of CH Working Papers, a refereed online publication series for studies in humanities computing, at Toronto and King's London. Has published widely on humanities computing in general, hypertext research, the Analytical Onomasticon Project and textual encoding, and electronic publication and communications.

Allen Renear, Associate Professor of Library and Information Science, University of Illinois, Urbana-Champaign.

President of the Association for Computing in the Humanities. He is also the Chair of the Open eBook Forum's Publication Structure Working Group. Prior to his current position he was for nine years the Director of the Brown University Scholarly Technology Group. He has also served as Director of the Brown Women Writers Project, on various working groups of the Text Encoding Initiative, and in 1998 was a Distinguished Visiting Fellow at the Oxford University Computing Unit. He has published widely in his area of speciality: how documents function as knowledge representation systems.

Harold Short, Director, Centre for Computing in the Humanities and Assistant Director (Focused Services), Information Services & Systems, King's College London.
Chair of the Association for Literary and Linguistic Computing. Organiser and editor of proceedings for the annual UK Digital Resources in the Humanities conferences.

John Unsworth, Professor, Foundation Director, Institute for Advanced Technology in the Humanities, University of Virginia.
Co-founder of *Postmodern Culture*. Commissioning editor for *Computers and the Humanities*. Involved in many national US projects in networked information and electronic scholarship.

Antonio Zampolli, founder and Director, Istituto di Linguistica Computazionale, CNR (Italian National Research Council), and Professor of Computational Linguistics at the University of Pisa.
Leader of a number of large projects in computational linguistics funded by the Italian state and by the EU, including PAROLE, SIMPLE, NERC, EAGLE, ISLE, and ENABLER. Special interest in the presentation of cultural resources on the world-wide web and in re-usable large-scale digital resources. President of ACCL, President of ELRA, member of the ICCL, President of LREC.

Symposium Program

Monday, 2 July
10.00-10.30 Welcome and coffee
10.30-12.30 John Unsworth, University of Virginia
Knowledge Representation as a Core Activity of Humanities Computing
12.30-13.30 Lunch with the Deputy Vice-Chancellor
13.30-15.30 Willard McCarty, King's College, London
A Research Topography for Humanities Computing
15.30-16.00 Afternoon tea
17.00-19.00 Drinks Party at The Customs House Hotel, Cnr of Watt and Bond Sts., Newcastle

Tuesday, 3 July
10.00-10.30 Coffee
10.30-12.30 Antonio Zampolli, Istituto di Linguistica Computazionale, CNR, Pisa
Language Resources and Humanities Computing: scientific, technical, strategic relationships

12.30-13.30 Lunch
13.30-15.30 Susan Hockey, University College, London
Putting Humanities Electronic Resources into the Digital Library
15.30-16.00 Afternoon tea
19.00-22.00 Symposium Dinner . The Brewery Restaurant, 150 Wharf Road, Newcastle Foreshore

Wednesday, 4 July
10.00-10.30 Coffee
10.30-12.30 Harold Short, King's College, London
Formal Models, Digital Scholarship, and Hybrid Libraries
12.30-13.30 Lunch
13.30-15.30 Allen Renear, University of Illinois, Urbana-Champaign
Text and Markup: More illumination in one lifetime than humanists deserve
15.30-16.00 Afternoon tea

Thursday, 5 July
09.15-11.00 Review session
11.00-11.30 Coffee
11.30-18.00 Outing to Hunter Valley vineyards and lunch
16:00-16:30 Sinéad O.Sullivan, University College Cork
Glosses, Texts and the Canons of Knowledge: Humanities computing and the early medieval glosses to Prudentius' Psychomachia with the Dean, Faculty of Arts and Social Science
2-5 July 2001 The University of Newcastle