

*Going Digital: **ncse***

A one day symposium to discuss **ncse** and digital scholarship.

Birkbeck College, 25 February 2006..

www.ncse.kcl.ac.uk

Registration is £8, inclusive of refreshments and lunch.

Program

10:00 Registration with light refreshments.

10:15 Introduction: Laurel Brake (ncse)

Laurel Brake is Professor of Literature and Print Culture at Birkbeck College and the Director of **ncse**.

10:30 Session One: ncse

Chair: Isobel Armstrong

Isobel Armstrong is Emeritus Professor of English at Birkbeck College, a Senior Research Fellow of the Institute of English Studies University of London, and a Fellow of the British Academy. She is a member of the **ncse** team.

Paper One: Jim Mussell and Suzanne Paylor (ncse).

Jim Mussell and Suzanne Paylor are the researchers for **ncse**.

Paper Two: Harold Short (CCH).

Professor Harold Short is Director of the Centre for Computing in the Humanities at Kings College London and Technical Director for **ncse** Project Team.

Discussion.

12:00-12:05 Break

12:05 Session Two: Going Digital.

Chair: Ed King (British Library).

Ed King is Head of the Newspaper Library and part of the **ncse** Project Team. He is also Project Director of the JISC / British Library's "British Newspapers: 1800-1900."

Paper One: Karen Spärck Jones (University of Cambridge).

Karen Spärck Jones is Professor Emeritus of Computers and Information at Cambridge and was the Chair of the British Academy's Policy Review on E-resources for Research in the Humanities and Social Sciences, published April 2005.

Paper Two: Peter Mandler (University of Cambridge).

Peter Mandler lectures in history at Cambridge University. He is part of the British Library's "British Newspapers: 1800-1900" and a member of the **ncse** International Advisory Board.

Paper Three: Andrew Cutting (London Metropolitan University).

Andrew Cutting is Senior Lecturer at London Metropolitan University and has published on Henry James and the digital age.

13:45 Discussion and working lunch.

15:00 Session Three: User Perspectives

Chair: Simon Tanner (CCH; Kings Digital Consultancy Services).

Simon Tanner is Director of Kings Digital Consultancy Services and part of the **ncse** Project Team.

Paper One: Linda Kaye (British Universities Film and Video Council).

Linda Kaye is the Senior Researcher on the BUFVC's "Cinemagazines and the Projection of Britain."

Paper Two: Beverley Kemp (Women's Library).

Beverley Kemp is Head of Library Services at the Women's Library.

Paper Three: Graeme Easdown

Graeme Easdown is a member of the Department of Education at Keele University. His recent publications report the findings of a study of History teacher attitudes to ICT and History teaching, undertaken in collaboration with Staffordshire LEA.

Discussion with tea and coffee.

17:30 Respondent: Marilyn Deegan (CCH).

Marilyn Deegan is Director of Research Development at CCH and was previously Digital Resources Director of the Refugee Studies Centre at Oxford University. Marilyn is part of the **ncse** Project Team.

18:00 Symposium ends.

There will then be a dinner at Vineopolis [www.cantinavinopolis.com] at 20:00. The price will be c. £30 per person, please notify us in advance if you would like to join us for dinner.

Directions and Travel:

A map of the symposium venue (The Clore Management Centre) and the surrounding area can be found at:

<http://www.bbk.ac.uk/bbk/findingbirkbeck/maps/immediateareawebmap.html>

Travel directions to the campus can be found here: <http://journeyplanner.tfl.gov.uk> and entering “Torrington Square” or “WC1 7HX” as the destination.

Abstracts

Session 1

Jim Mussell and Suzanne Paylor
Birkbeck College

ncse: selecting the core

The Nineteenth-Century Serials Edition (**ncse**) was initially conceived to provide digital editions of six runs of nineteenth-century periodicals. Advanced metadata and concept maps would provide interpretive links between diverse content in order to highlight the interconnected nature of nineteenth-century print culture. As the project has progressed, the team discovered that two of our six titles published more than one edition per week, substantially increasing the total pagination. Our resources, both in terms of time and money, will not allow us to undertake the advanced work as originally envisaged on the whole of the publications. This paper investigates the conceptual issues that have shaped what we have come to call ‘the core.’

The six titles were selected on the grounds of their diversity: although valid historical criteria can be established to justify their inclusion, it is the combination of different print forms that **ncse** is designed to accommodate. Just as these six titles were selected from a much larger archive, we have developed principles to further delimit page spans from the titles for further processing in the core. Derived from the titles themselves, the selection criteria identify representative samples from the runs while also emphasizing common aspects of nineteenth-century print culture.

ncse has evolved into a two-stream resource. The first phase of the project will include page images and uncorrected OCR text, and is due to be launched later this year. The ‘core’, due to be published after the end of the project in 2008, features structural and advanced metadata, as well as concept maps. We will also provide our users with the option to include or exclude the multiples in their searches of the edition. In selecting material for this advanced treatment, we have deliberately moved away from models that focus on complete runs in favour of a non-linear collection of partial runs. Although we have tried to accommodate continuous sequences where possible, we argue that this approach emphasizes generic aspects of the ‘periodical.’ This paper accounts for the choices we have made in developing selection criteria and aims to consider the wider implications of this approach for further work in the digital humanities.

Harold Short

Centre for Computing in the Humanities, King’s College London.

The **ncse** Project: technical framework and technical research

The Centre for Computing in the Humanities at King’s College London (CCH) is an international leader in the application of technology in research in the arts and humanities. It is an academic department in the School of Humanities at KCL, and operates on a collaborative basis across discipline, institutional and national

boundaries: it has collaborative relationships across the College and the University of London, and with a large number of institutions and bodies in the UK and internationally. CCH is involved in more than 30 major research projects, with funding from the Arts and Humanities Research Council (AHRC), the Leverhulme Trust and the Andrew W Mellon Foundation. CCH is responsible for MA programmes in Digital Culture and Technology, and Digital Humanities; it also runs an undergraduate Digital Humanities programme.

CCH has been involved in the digitization and delivery of complex printed content since 1997. It was a major partner in the Forced Migration Online project (with the Refugee Studies Centre at Oxford University, and other partners worldwide), and has hosted the British Library Newspaper Pilot since 2000. (See www.forcedmigration.org and www.uk.olivesoftware.com). In both these projects we have worked closely with Olive Software, using their innovative software for document processing, search and retrieval. We have helped Olive to define new products and new ways of manipulating textual data. We have also been working on a range of developments in information retrieval from large bodies of data, and worked closely with linguists from Lancaster University in the Keyword Project which researched and evaluated a number of different research methods for extracting keywords from structured and unstructured textual data.

CCH's role in the NCSE project is to lead the technical research and to develop technical frameworks for research and delivery of what has turned out to be more than 100,000 pages of data. The technical team is researching and developing tools to assist in understanding the data, using keyword extraction techniques and concept mapping, and in collaboration with Olive is building an interface that will offer easy access to the data for a wide range of users. There will effectively be two resources made available to users: the first will offer the full set of all the periodicals with searchability at page level page, but not at article level (107,000 pages). For the second resource, 30,000 selected pages will have enhanced functionality. These pages will be segmented into individual items – by automated means as far as this may be possible - and some additional metadata will be added. Linguistic techniques will be used to extract keywords and to provide input to the concept mapping process. Existing databases of 19C history and culture will be used as a basis for automated processes attempting to identify persons, institutions and events.

Session 2

Karen Spärck Jones
University of Cambridge

Surfing the Digital Wave: lessons from the IT world

IT seems to offer wonderful new opportunities for bigger and better research resources and nicer and niftier research tools to manipulate them with. Now, IT appears to be implying, you can really organise, enrich, exploit the materials that are out there but can't be brought together and developed until they are in digital form. The lessons from the IT world are yes, you can do wonderful things, but not in the ways you expected or hoped.

So what are sensible ways to invest in IT research resources? The Web is particularly instructive. No one built the set of pages on the Web like a conventional data resource. No one built the Web engines like a conventional resource description and access mechanism (indeed the reverse). The Web as a resource, and the Web engines as resource management tools, respond to what arrives on the Web and what Web users do. The Web world does not seek, pre-emptively, to impose a model of the knowledge it contains that the user is obliged to adopt.

There are certainly minimum requirements for a useful digital world like the Web. The pages have to be in viable formats (but the range of acceptable/convertible formats is increasing); and they have to have *some* access keys that users will naturally use. But beyond that, it's far better to go for redundant, simple access keys than for purist, elaborate labelling schemes, and for flexible, a posteriori tools for using these keys than for rigid, a priori ones.

Resources for digital scholarship may seem to call for, and justify, fancy modes of organisation and use. The lessons of IT so far are that investing up front in fancy modes of organisation is a mistake: when digital resources actually exist, the prior predictions about what people will do and want too often turn out to be wrong.

Peter Mandler

University of Cambridge

Digital Divides

Much has been written (though less lately) about the 'digital divide' between those with internet access and those without. As digitization of the print universe creeps forward, new 'digital divides' are opening up that pose challenges to scholarship. The mixed commercial and public or open-access provision creates a 'digital divide' between those who have institutional access to commercial and/or proprietary resources and those who don't. The partial nature of digitization programmes creates a 'digital divide' between resources that have been digitized and those that haven't (and especially those that, partly for commercial reasons, may not be for a long time). The technical limitations of digitization create a 'digital divide' between scholarly resources that can easily be accessed by this means and those that can't. This talk will (fleetingly) touch on all three of these divides.

Andrew Cutting

London Metropolitan University

More Than a Resource? Extending the **ncse** agenda

Flash animations, especially in the form of online games, have become an important growth area in educational multimedia design. They exemplify a shift in user expectations towards kinds of highly visual interactivity that can make other digital resources seem static, un-engaging and dated. The design of scholarly e-resources for the Humanities struggles to keep pace with such ongoing techno-cultural changes, even when careful attention is paid to development of sophisticated user interfaces. Partly this is a matter of limited funding and of long lead-times required to develop complex new products. And partly this lag reflects the cultural rooting of Humanities scholars in printed text – in contrast to many of today's students (tomorrow's researchers?) who are more fundamentally at home with other media. This emerging audience is also divided in its attitudes and access to technologies and education, however. In my experience at a teaching-led University, resistance to 'going digital' is common among students and academics, while students are frequently resistant towards book-centred learning, too. Designing for such a mixed audience, in such a fast-

changing environment, is obviously difficult. AHRC-funded e-projects, such as nese, primarily seek to provide a resource serving established researchers, rather than actively address pedagogical or social justice agendas. Focussing their remit in this way helps Humanities e-projects achieve their immediate aims. Arguably this is at the cost of missing other opportunities: to engage potential scholars of the future and to open up new critical understandings through attempting a more radical adaptation from printed text to new media.

Session 3

Linda Kaye

Senior Researcher, 'Cinemagazines and the Projection of Britain'

Changing the Way We Look at Words: some practical and theoretical aspects of user interaction

Drawing on the experience of developing the multimedia British Universities Newsreel Database [BUND], Linda Kaye will discuss questions raised by the practical and more theoretical aspects of user interaction. Why is web design so important and what makes it good? To what degree does this lead the user to ask certain questions of the material? What is the difference between examining a document as an object, the original or copy, and as an image, on screen. Is visual literacy a necessity for 'digital scholars'?

Beverly Kemp

The Women's Library, London.

Digital Resources for Women's History: user perspectives

Researchers of women's history require access to a range of online sources to support their work, including resources specific to women as well as more general resources to place women's history in the broader context. As far as database searching, retrieval and record display is concerned women's history researchers share many of the same user requirements as other researchers. Currently however there are a limited number of full text online sources specialising in women's history, whilst the more general resources are often not conducive to searching from a gender perspective. In this paper I will focus on databases currently available in the field of women's history, highlight what works well within those databases for the user, and consider some of the advantages the more general databases can offer to the women's historian.

Graeme Easdown

Keele University

Using Digital Resources in Educational Settings: a user's perspective

The case for using digital resources to support the development of pupils' understanding is very persuasive. These resources can provide access to a rich source of primary evidence in a form that is easily accessible and which enables interactive engagement for users. Carol McGuiness's paper From Thinking Skills to Thinking Classrooms (Department for Education and Employment, Research Briefs, Research Report No 115) makes the point that

use of multi-media resources can ‘enhance children’s understanding and powers of reasoning’ and ‘permit students to form rich images of problem situations ... and prompt alternative perspectives.’

Government policy makers have been persuaded by the case and committed increasingly large sums to projects that promote the use of multi-media technology to support teaching learning. Many schools are now technology rich environments in which pupils and teachers have access to a diverse range of resources, which include both local and networked services such as Virtual Learning Environments (VLEs), repositories and intranets, providing a range of opportunities for collaborative learning. The need for shifts in teacher development and support has also been addressed, with varying degrees of success, through government-funded initiatives such as the New Opportunities fund ICT Training initiative and the Enhancing Subject Teaching Using ICT (ESTU ICT) project.

There is a clear educational and pedagogical case for the use of digital resources, policy makers have committed large sums to the promotion of these materials yet the research and Inspection evidence consistently shows that the potential of resources such as those being developed by the NCSE is not being exploited by teachers. This paper, in addition to elaborating the case for ‘Going Digital’, will explore the reasons for this situation and offer some tentative and provisional thoughts about how the issue might be addressed.