

2. Keynote Speakers

Professor Andrew Prescott

Head of Department, Department of Digital Humanities, King's College London

Andrew Prescott is Professor of Digital Humanities at King's College London. Andrew was from 1979-2000 a Curator in the Department of Manuscripts at the British Library, where he acted as British Library co-ordinator for a number of digital projects, including most notably *Electronic Beowulf*, edited by Kevin S Kiernan of the University of Kentucky. From 2000-2007 he was Director of the Centre for Research into Freemasonry in the Humanities Research Institute at the University of Sheffield. He has also worked at the University of Wales Lampeter and University of Glasgow. He has served on the advisory boards of many digital humanities projects in Britain and America.



Made in Sheffield: Industrial Perspectives on the Digital Humanities

Plenary Session 1: Thursday 6th September, 11.30 - 13.00

A comparison between the advent of digital technology and the arrival of print has become one of the clichés of modern life. Yet it took a long time for the technology of moveable print to influence society on a large scale. As Raymond Williams long ago pointed out, it was only the arrival of the steam press at the beginning of the nineteenth century that turned print into a mass medium. Comparisons with, and elegies on, Gutenberg are not the most helpful starting point for thinking about digital culture and society. It will be suggested that a more useful historical comparison is the industrial and economic changes that occurred in Britain at the end of the eighteenth century. The computer, controlled by systems devised to power industrial looms, is simply the most sophisticated product of the industrial revolution.

In considering the relationship between the digital and industrial revolutions, there can be no better starting point than Sheffield, one of the world's great crucibles of industrialization and modernity. By examining the conditions which made Sheffield an important centre of industrial and manufacturing innovation in the late eighteenth century, we gain insights into many issues of current concerns in the digital humanities. There are questions of scale: how huge new complexes were created from small craft traditions. Sheffield provides us with insights into the role of

2 Digital Humanities Congress 2012, The University of Sheffield

experimentation in developing new technologies. The growth of Sheffield also helps us understand the role of networks in innovation. The great manufacturing traditions of Sheffield have a lot to teach us in a new digital world. Indeed, if 3D printing and other new manufacturing technologies indeed create, as has been suggested, a new Industrial Revolution, it may be that the traditions of Sheffield prove newly and unexpectedly relevant to us in negotiating the next stages of digital change.

3 Digital Humanities Congress 2012, The University of Sheffield

Professor Lorna Hughes

University of Wales Chair in Digital Collections at the National Library of Wales

Lorna Hughes is the University of Wales Professor of Digital Collections at the National Library of Wales. She took up her appointment in January 2011, and leads a research programme based around the Library's digital collections. Prior to moving to the National Library of Wales in Aberystwyth, Lorna worked at King's College, London, most recently as Deputy Director of the Centre for e-Research. From 2005 to 2008, she was Programme Manager of the Arts and Humanities Research Council (AHRC) ICT Methods Network. She has worked in digital humanities at New York University, Arizona State University, Oxford University and Glasgow University. She is the author of *Digitizing Collections* (Hughes, 2004) and the co-editor of *Virtual Representations of the Past* (Greengrass and Hughes, 2008).



Live and Kicking: The Impact and Sustainability of Digital Collections in the Humanities

Plenary Session 2: Friday 7th September, 16.00 - 17.30

Digitisation initiatives in libraries, archives, museums and educational institutions have created a 'deluge' of data in the humanities that has transformed the information landscape and the way it is navigated for research and teaching. The use of digital collections for scholarship – using ICT based tools and methods – has been the basis of transformative and innovative research across the disciplines, allowing enhanced access to materials, and supporting new modes of collaboration and communication. Digital collections development has required a huge investment of public and private funding, and has, in turn, engendered an even greater requirement for continued investment to expand, enhance and augment the existing digital knowledge base, and to embed these collections in research and scholarship. More importantly, issues of sustainability and preservation of digital resources over the long term have an impact on the long term use and impact of digital collections.

This presentation will discuss the research programme in digital collections at the National Library of Wales. This is developing an evidence base for the use, value and impact of the digital collections of Wales, and investigating the relationship between impact and the long terms sustainability of digital collections.

4 Digital Humanities Congress 2012, The University of Sheffield

Professor Philip J. Ethington

Professor of History and Political Science, University of Southern California and
Co-Director of the USC Center for Transformative Scholarship



Philip J. Ethington is Professor of History and Political Science at the University of Southern California, North American Editor and Multimedia Editor of the journal *Urban History* (Cambridge University Press), and Co-Director (with Tara McPherson) of the USC Center for Transformative Scholarship. An interdisciplinary historian, Ethington's scholarship explores the past as a cartography of time. His recent published work include theoretical work on a spatial theory of history; sociological studies of residential segregation; large-format maps of urban historical change; online interactive Web 2.0 tools, archives, and publications for urban studies; and museum exhibit collaborations. He is co-PI of the HyperCities project. Most recently, Ethington co-wrote the award-winning film *Visual Acoustics: The Modernism of Julius Shulman* (Art House Films, 2009), which is narrated by Dustin Hoffman. His photography and cartography have been published and exhibited internationally. He is currently completing a large-format graphic book, interactive online publication, and public art exhibit *Ghost Metropolis: Los Angeles, since 13,000 BP*.

The Transcendence of Genre: Multimodal Publishing After Cervantes

Plenary Session 3: Saturday 8th September, 14.00 - 15.30

The categories of humanistic knowledge typically follow the categories of expression—written, printed, painted, cinematic, televisual, performed, built, and so on. Multimodal online publications, (mixing typographic text, archival images of archival documents, video, still images, interactive examinations of virtual 3-d objects, shared platforms, author-reader interactions, and other functionalities), intentionally bring different genres together to produce new interpretations of past and present cultural production. This presentation begins with the problem of genre as Miguel Cervantes posed it in *Don Quixote* (1605-15). By reflectively juxtaposing the dominant existing genres of his era: chivalric, pastoral, picaresque, moralist, plus poetry and drama; by the use of multiple narrators, Cervantes in effect abolished literary genres, even while parodying them, in his *invention of the novel*. Novels, henceforth, would incorporate genres within their form--but as a form, transcend genre. Such a transcendence, at a broader level among all modes of communication (multimodality), is now underway on a global scale in the networked Digital Age. Publications delivered via online multimodal “platforms” such as *Scalar*

5 Digital Humanities Congress 2012, The University of Sheffield

and *HyperCities*, reach deeply and directly into digital archives, through open channels mediated by transnational metadata. They reach widely out to multiple publics and multiple authors. They transform the economic basis of publishing, and they challenge the boundaries of disciplinary critical review. The resulting works--narrating histories, annotating film clips, interacting with maps—cannot be reduced to genre, and thereby, call on multiple orders of knowledge. Yet the “online platform,” like *Don Quixote*, struggles to find an identity on the virtual shelves of the world’s artistic and intellectual labor. There is no steady stream of such works in academia, no tradition of reviewing them, gaining access to them, grouping them in edited collections, being evaluated for promotion with respect to them, and so on. The need today is to institutionalize the circulation, evaluation, and reproduction, of this genre-transcendent entity, amid the upheaval of the revolution in publishing peer-reviewed scholarship involving university presses and open-access institutions in collaborative environments. Ethington will draw on his own “deep mapping” research and his experience as Multimedia Editor for Cambridge Journals Online’s *Urban History*, and co-author of the *American Historical Review* Digital Prize publication program.

3. Conference Programme

Please note that the individual presentations for each Session can be found in section 4. *Conference Sessions* whilst paper abstracts are available in section 5. *Abstracts*.

Thursday 6th September 2012

9.00 - 11.30: Registration

11.30 - 13.00: Introductions and Plenary

Plenary by Professor Andrew Prescott

13.00 – 14.00: Lunch

14.00 - 15.30: Parallel Sessions 1, 2, 3

Session 1: *Approaches to Digitising Literary/Historical Texts*

Session 2: *Statistics and Sentiment: Advanced Text Mining*

Session 3: *Interfacing with the User*

15.30 - 16.00: Tea Break

16.00 - 17.30: Parallel Sessions 4, 5, 6

Session 4: *New Ways of Working: Using Digital Humanities to Evolve Academic Practice*

Session 5: *Visualising Complex Pathways and Datasets*

Session 6: *Developing Research Tools*

18.30: Drinks Reception

Friday 7th September 2012

9.30 - 11.00: Parallel Sessions 7, 8, 9

Session 7: *Taking Issue: Developing the Digital Humanities*

Session 8: *Digital Representation of Historic Artefacts: in Search of Detail*

Session 9: *Text and Space: the Fundamentals of the Spatial Turn*

11.00 - 11.30: Tea Break

11.30 - 13.00: Parallel Sessions 10, 11, 12

Session 10: *Assessing Impact in the Digital Humanities*

Session 11: *Text and People: Authorship and Sentiment within Textual Analysis*

Session 12: *3D Modelling in Research*

13.00 – 14.00: Lunch

14.00 - 15.30: Parallel Sessions 13, 14, 15

Session 13: *Digital Editions: Transmuting the Original*

Session 14: *Old Paradigms and New Techniques: The Impact of Digital Data*

Session 15: *Exploring Datasets using GIS*

7 Digital Humanities Congress 2012, The University of Sheffield

15.30 - 16.00: Tea Break

16.00 - 17.30: Plenary

Plenary by Professor Lorna Hughes

19.00: Conference Dinner

Saturday 8th September 2012

9.30 - 11.00: Parallel Sessions 16, 17, 18

Session 16: *Working with Image Collections*

Session 17: *Semantics and Digital Data*

Session 18: *Developing Infrastructure and Policy for Digital Research*

11.00 - 11.30: Tea Break

11.30 - 13.00: Parallel Sessions 19, 20, 21

Session 19: *Mass Digitisation and Beyond: Making Sense of Datasets*

Session 20: *New Ways of Reading: Digital Humanities and the Reader*

Session 21: *Research on the Go*

13.00 – 14.00: Lunch

14.00 - 15.30: Plenary and Close

Plenary by Professor Philip J. Ethington

4. Conference Sessions

Please note that the overall programme for the conference can be found in the previous section 3. *Conference Programme* whilst paper abstracts are available in section 5. *Abstracts*.

Thursday 6th September, Late Morning: 11.30 - 13.00

Introductions and Plenary Session 1

Andrew Prescott
King's College London

Made in Sheffield: Industrial Perspectives
on the Digital Humanities

Thursday 6th September, Early Afternoon: 14.00 - 15.30

Session 1 Approaches to Digitising Literary/Historical Texts	
Gabriel Egan <i>De Montfort University</i>	Quick-and-Dirty XML Applications for Textual Analysis
James Cummings <i>Oxford University</i>	Compromises and Flexibility in the TEI Customization
Takako Kato <i>De Montfort University</i>	Transcribing Incipits and Explicits in TEI-XML
Session 2 Statistics and Sentiment: Advanced Text Mining	
Pim Huijnen <i>Descartes Centre for the History and Philosophy of the Sciences and the Humanities</i>	WAHSP: Developing a Web-Application for Historical Sentiment Mining in Public Media
José de Kruif <i>Utrecht University</i>	Text Mining a Nineteenth Century Media Hype
Martyn Harris <i>University of London</i>	Search and Mining Tools for Linguistic Analysis
Session 3 Interfacing with the User	
Andrea Scharnhorst <i>Royal Netherlands Academy of Arts and Sciences</i>	Visual Interfaces to Collections
Mark Hall <i>University of Sheffield</i>	Navigating Cultural Heritage Collections using Pathways
Keira Borrill <i>University of Sheffield</i>	Participating in Search Design: Understanding Scholarly Practices

Thursday 6th September, Late Afternoon: 16.00 - 17.30

Session 4 New Ways of Working: Using Digital Humanities to Evolve Academic Practice	
<p>Tim Evans <i>University of York</i></p>	<p>A History of Archaeological Investigation in Post-War England: Insights Provided by Digital Resources</p>
<p>Julianne Nyhan and Anne Welsh <i>University College London</i></p>	<p>Uncovering the “Hidden Histories” of Computing in the Humanities 1949 – 1980: An Overview of Our Key Findings</p>
<p>Chiel van den Akker <i>VU University Amsterdam</i></p>	<p>History as Dialogue</p>
Session 5 Visualising Complex Pathways and Datasets	
<p>John Haggerty <i>University of Salford</i></p>	<p>Investigating Networks over Time: Matrixify</p>
<p>William Martin , Dani Abdallah, Ahmed El-Abiary, Yosef Dalbah, Simon Julier, Melissa Terras <i>University College London</i> Rob Iliffe <i>University of Sussex</i> Michael Hawkins <i>University of Sussex and University of Cambridge</i> Tim Weyrich <i>University College London</i></p>	<p>Newton Spectrum: Nonlinear Text Browsing of a Large Corpus</p>
Session 6 Developing Research Tools	
<p>Romain Janvier <i>Université of Pau</i> and Guillaume Sarah <i>CNRS</i></p>	<p>Rich Internet Application for Collaborative Numismatics</p>
<p>Andrea Kulas and Lu Yu <i>Max Planck Digital Library</i></p>	<p>From Individual Solutions to Generic Tools: Digitization at the Max Planck Society</p>

Friday 7th September, Early Morning: 9.30 – 11.00

Session 7 Taking Issue: Developing the Digital Humanities	
<p>Smiljana Antonijević and Sally Wyatt, <i>Royal Netherlands Academy of Arts and Sciences</i> and Monica Bulger and Eric Meyer, <i>University of Oxford</i></p>	Digital Humanities in Practice
<p>Erin Snyder <i>University of Oxford</i></p>	An Institutional Framework for the Digital Humanities: An Alternative to the DH Centre
<p>Toby Burrows <i>University of Western Australia</i></p>	Designing a National “Virtual Laboratory” for the Humanities: the Australian HuNI Project
Session 8 Digital Representation of Historic Artefacts: in Search of Detail	
<p>Bill Endres <i>University of Kentucky</i></p>	More than Meets the Eye: Going 3D with an Early Medieval Manuscript
<p>Guillaume Sarah and Florence Codine, <i>CNRS</i></p>	Transcribing Early Medieval Epigraphy in the Digital Age
<p>Ségolène M. Tarte <i>University of Oxford</i></p>	Cognitive Insights in Interpretation Building: Tailoring Software to Expert Practices
Session 9 Text and Space: the Fundamentals of the Spatial Turn	
<p>Anouk Lang <i>University of Strathclyde</i></p>	Mapping Miss Mansfield: Using Digital Tools to Explore the Role of Place in the Work of Katherine Mansfield and Witi Ihimaera
<p>Patricia Murrieta-Flores, David Cooper and Ian Gregory <i>Lancaster University</i></p>	Spatial Humanities: Exploring and Analysing Texts within a GIS Environment
<p>Paul Rayson, Alistair Baron and Andrew Hardie <i>Lancaster University</i></p>	Which ‘Lancaster’ do you mean? Disambiguation Challenges in Extracting Place Names for Spatial Humanities

Friday 7th September, Late Morning: 11.30 – 13.00

Session 10 Assessing Impact in the Digital Humanities	
Simon Tanner <i>King's College London</i>	New Approaches to Measuring the Impact of the Digital Humanities
Jonathan Blaney <i>University of London</i>	The Citation Problem in the Digital Humanities
Max Kemman, Martijn Kleppe, Stef Scagliola and Renske Jongbloed <i>Erasmus University Rotterdam</i>	Mapping the Use of Digital Sources Amongst Humanities Scholars in the Netherlands
Session 11 Text and People: Authorship and Sentiment within Textual Analysis	
Francesca Benatti , <i>Open University</i> and Justin Tonra , <i>University of Virginia & NUI Galway</i>	Who Killed Christabel? Can Authorship Attribution Solve the Case?
Mitsunori Ogihara <i>University of Miami</i>	Analyzing the Carlyle Letter Collection
Hinke Piersma <i>NIOD Institute for War, Holocaust and Genocide Studies</i>	The Second World War as a Benchmark of Political Morality in Post-War Political Discourse in the Netherlands
Session 12 3D Modelling in Research	
Sushardjanti Felasari and Chengzhi Peng <i>University of Sheffield</i>	Connecting Digital Representations: A City's Urban Spaces and its Collective Memory
Yang Yu and Chengzhi Peng <i>University of Sheffield</i>	Exploring the Boundary of Architectural Enquiry through Mixed Reality Modelling

Friday 7th September, Early Afternoon: 14.00 – 15.30

Session 13 Digital Editions: Transmuting the Original	
Michael Meredith University of Sheffield	Representing Norwegian Emigration in the Virtual World Through the Eyes of Ola Nordmann
Tobias Schweizer, Ivan Subotic and Lukas Rosenthaler <i>University of Basel</i>	Building Digital Editions on the Basis of a Virtual Research Environment
Jennifer Stertzer <i>University of Virginia</i>	Working with the Financial Records of George Washington: Data and Database
Session 14 Old Paradigms and New Techniques: The Impact of Digital Data	
Ester Appelgren, Helge Hüttenrauch and Gunnar Nygren <i>Södertörn University</i>	Data Journalism – Implications and Opportunities
Christopher Dingle , <i>Birmingham Conservatoire</i> and Laura Hamer , <i>Liverpool Hope University</i>	False Memories and Dissonant Truths: Digital Newspaper Archives as Catalyst for a New Approach to Music Reception Studies
Claudia Favero <i>Open University</i>	Digital Historians: Perspectives and Approaches in Research and Teaching
Session 15 Exploring Datasets using GIS	
Jennifer Bullock , <i>Adam Matthew Digital</i> and David Heyman , <i>Axis Maps</i>	Space and Place in Victorian London: Interactive Mapping for Humanities Research
Hilde De Weerd <i>King's College London</i>	Citation Networks and Topic Maps: Digital Readings of Imperial Chinese Notebooks
Karina Rodriguez-Echavarria, Leah Armstrong, Catherine Moriarty and David Arnold <i>University of Brighton</i>	Using GIS Technologies to Explore the Disciplinary Reach and Geographic Spread of British Designers

14 Digital Humanities Congress 2012, The University of Sheffield

Friday 7th September, Late Afternoon: 16.00 – 17.30

Plenary Session 2

Lorna Hughes
National Library of Wales

Live and kicking: the impact and sustainability of digital collections in the humanities

Saturday 8th September, Early Morning: 9.30 – 11.00

Session 16 Working with Image Collections	
<p>Marcio Emilio dos Santos and Cicero Inacio da Silva <i>Federal University of Juiz de Fora</i></p>	Analyzing Big Cultural Data Patterns in 4,000 Covers of Veja Magazine
<p>Leigh Garrett and Marie-Therese Gramstadt <i>University for the Creative Arts</i></p>	KAPTUR: Examining the Importance and Effective Management of Research Data in the Visual Arts
<p>Michael John Goodman <i>Cardiff University</i></p>	'Art to Enchant': The Creation of a Digital Archive
Session 17 Semantics and Digital Data	
<p>Marc Alexander and Ellen Bramwell <i>University of Glasgow</i></p>	Mapping Metaphors of Wealth and Want: A Digital Approach
<p>Jeremy Huggett <i>University of Glasgow</i></p>	Promise and Paradox: Accessing Open Data in Archaeology
<p>Damiana Luzzi <i>Fondazione Rinascimento Digitale</i></p>	Reperio: a Collaborative Knowledge Environment for Digital Humanities
Session 18 Developing Infrastructure and Policy for Digital Research	
<p>Erik Malcolm Champion <i>Aarhus University</i></p>	Research As Infrastructure
<p>Ann Gow and Laura Molloy <i>University of Glasgow</i></p>	Ahead of the Curve: Digital Curator Vocational Education
<p>Clare Scott <i>University of Sheffield</i></p>	Getting Rights Right! - The University of Sheffield Library Experience

Saturday 8th September, Late Morning: 11.30 – 13.00

Session 19 Mass Digitisation and Beyond: Making Sense of Datasets	
Giles Bergel <i>University of Oxford</i>	A Sense of Tradition in the Digital Archive: The Example of Broadside Ballads
Stephen Brown, Simon Coupland and David Croft <i>De Montfort University</i>	Improving Record Matching Across Disparate Historical Resources
Tom Storrar <i>The National Archives</i>	A Semantic Knowledge Base for the UK Government Web Archive: Opportunities for Researchers
Session 20 New Ways of Reading: Digital Humanities and the Reader	
Bronwen Thomas and Julia Round <i>Bournemouth University</i>	Researching Readers Online
Simon Rowberry <i>University of Winchester</i>	Vladimir Nabokov's <i>Pale Fire</i> and the Problem of Interface
Elisabeth Salter <i>Aberystwyth University</i>	In the Mind's Eye: Reflections on Generating Reader Experience c 1350 – 1600
Session 21 Research on the Go	
Genovefa Kefalidou, Mercourios Georgiadis and Suchith Anand, <i>University of Nottingham</i> and Bryn Alexander Coles, <i>Open University</i>	Crowd-Sourcing our Cultural Heritage
Adam Park <i>University of Sheffield</i>	The Production of a 'Locative Digital Trail' as a Creative, Collaborative Methodology to Investigate or 'Map' Place-Identity
Melissa Terras and Steven Gray <i>University College London</i>	Building Textal: What can Apps do for Digital Humanities?

Saturday 8th September, Early Afternoon: 14.00 - 15.30

Plenary Session 3 and Close

Philip J. Ethington
University of Southern California

The Transcendence of Genre: Multimodal
Publishing After Cervantes

5. Abstracts

Please note that abstracts for the three plenary sessions are available in section 2. *Keynote Speakers*

The Scholarly Edition in the Context of Social Knowledge Production: A 'Social' Edition of the Devonshire MS (BL Add Ms 17492)

Ray Siemens, *University of Victoria*

We need to extend our understanding of the scholarly edition in light of new models of edition production that embrace social networking and its commensurate tools. This paper offers an understanding the scholarly edition in the context of new and emerging social media from two pertinent perspectives: the first, from the perspective of its theoretical context and particularly as that context intersects with a utility-based consideration of the toolkit that allows us to consider the social edition as an extension of the traditions in which it is situated and which it has the potential to inform productively; the second, from the perspective of an iterative implementation of one such edition, carried out via a research team operating in conjunction with an advisory group representing key expertise in the methods and content-area embraced by the edition. The implementation that will form the basis of the example is an edition of the Devonshire MS (BL Add Ms 17492), the first sustained example of men and women writing together in the English literary tradition.

Thursday 6th September, 14.00 – 15.30: Session 1

Quick-and-Dirty XML Applications for Textual Analysis

Gabriel Egan, *De Montfort University*

Not every set of questions we want to answer about literary/historical texts is best addressed by comprehensive transcription of all the features of the documentary witnesses. There is much that can be done in XML without undertaking a comprehensive encoding, and this paper will illustrate one such roll-your-own XML application for answering questions about the compositors who typeset the first editions of Shakespeare. Forty-six years ago Fredson Bowers foresaw computers transforming this field: "It will be a blessed day in the future when one can press a button and give such a lordly command as 'List for me every time compositor B follows his copy in spelling win as win or winne, every time he changes a copy spelling win to winne, or winne to win, and distinguish in each case what he does in setting prose and setting verse". We've been rather slow at addressing this problem, and existing transcriptions of the early editions are distinctly unhelpful here. The paper will show how far a lone scholar can get towards a tool that will answer Bowers's questions using freely available materials and just a little knowledge of XML.

Thursday 6th September, 14.00 – 15.30: Session 1

Compromises and Flexibility in the TEI Customization

James Cummings, *Oxford University*

The ongoing efforts of the Text Encoding Initiative (TEI) have produced a wide-ranging set of Guidelines for encoding textual phenomena. However, standardisation attempts such as this always make compromises in order to achieve their goals. The recommendations of the TEI are no exception to this and this paper examines at some of these compromises along with the benefits and drawbacks they bring. The TEI's extremely flexible method for documenting customisations and extensions will be also considered and how it might benefit interoperability amongst sub-communities of users.

Thursday 6th September, 14.00 – 15.00: Session 1

Transcribing Incipits and Explicits in TEI-XML

Takato Kato, *De Montfort University*

The AHRC-funded Project "The Production and Use of English Manuscripts 1060 to 1220" aims to identify, analyse and evaluate all manuscripts containing English written in England between 1060 and 1220. We are currently transcribing the incipits and explicits--the opening and closing words of manuscripts--in TEI-compliant XML. What is the real value of transcribing them in TEI-XML? How can the user optimise our detailed descriptions of 205 manuscripts? Theoretically our transcriptions and descriptions will allow the user to conduct searches such as "where, by whom and for whom were the manuscripts that contain 'First Sunday after Easter' produced, what are the typical sizes and the layouts of those manuscripts, and what other texts are included in these manuscripts?" In reality the user will be able to find the answers to such complex questions, but it will take more than the clicking of a few buttons. Theoretically, our transcriptions being in TEI will allow integration with other available resources. In reality, it will take a while before the user can benefit from the integration. This paper explores research potentials and limits of such a resource by contextualizing the English Manuscripts Project in the digital electronic research culture of medieval manuscripts.

Thursday 6th September, 14.00 – 15.30: Session 1

WAHSP: Developing a Web-Application for Historical Sentiment Mining in Public Media

Toine Peters, *Descartes Centre for the History and Philosophy of the Sciences and the Humanities*

This case study concerns a user oriented model of collaboration between humanities scholars and ICT developers. Experiences in this trans-disciplinary project aiming at advanced forms of text mining in a large historical dataset of Dutch newspapers provide lessons for a smooth integration of perspectives. The technical challenge has been to convert a specific text mining technology, so-called 'sentiment mining', into a user-friendly web-application addressing research questions of the intended user group of historians. The interdisciplinary project-team (historians, linguists, computer scientists) has tailored existing tools to the specific needs of digital humanities research, with a special focus on opinions/perceptions regarding the use and abuse of drugs between 1900 and 1945.

The starting point for this project has been the open-source software infrastructure xTAS (<http://xtas.net>) developed by the Intelligent Systems Lab (Maarten de Rijke) at the University of Amsterdam. Building on this infrastructure, and in collaboration with historians at Utrecht University, a CLARIN-supported web-application for historical sentiment mining in public media has been built (<http://washp.nl>). The development of an open-source mining technology that matters to domain users requires continuous negotiation of methodologies, practices and comprehension between historians and computer scientists. A pre-requisite is in-process collection and analysis of hands-on experiences with the mining tool under construction. Therefore, every developmental task and research activity envisaged within the project has been a trans-disciplinary co-production: from selecting and filtering out meaningful lexical items, carrying out text-mining tasks, training the algorithms to meet the needs of the domain users up to in-process feed-back loops. I will show that the process of articulating the needs and demands of users in relation to available technical options crucially depends on including program mediators who bring a strong background in the humanities as well as state-of-the-art text-mining expertise in the research team. The mediators will focus on aligning specific technical choices and innovations of computer scientists with the specific research needs of the participating historians.

Text Mining a Nineteenth Century Media Hype

José de Kruif , *Utrecht University*

This paper will record the results of research into publications (pamphlets and newspaper articles) on a nineteenth century political conflict: the 'aprilbeweging' of 1853, a passionate dispute between the government and protestant and Roman Catholic citizens. Digitized documents containing texts on this issue were researched with the help of state of the art text- and data mining tools. Thus, this research is also testing the suitability of modern text mining tools for research into historical documents. This software normally functions within commercial companies for analysis of for instance call centre databases. What do the customers think of our products? And within governmental institutions such as security, customs and police, for analyses of e-mails and phone taps.

But in this case, the instrument was employed to systematically take stock of the publications in this nineteenth-century media hype. It was used to answer such questions as: What contribution did the specific form of the nineteenth-century media landscape make to the development of the hype? What format of publication seems to fit the contents best? And which genres can be distinguished and what kinds of argument do people use in these publications? Sentiment analysis was used to list opinions on crucial actors and issues. Finally, the results of the text mining of media content, could systematically be compared to metadata like authorship, publisher, genre etc. by using data mining tools like cluster analysis.

Texts were thus surveyed on content as well as on other characteristics. Not only will the knowledge we thus acquire on the way conflicts were fought in the media be discussed, but the issue which text mining components are a suitable method for this kind of research, will be addressed as well.

Search and Mining Tools for Linguistic Analysis

Martyn Harris, *University of London*

Samtla (Search And Mining Tools for Linguistic Analysis) is a research tool providing domain-specific search using statistical language modelling, and is designed to assist researchers in the Humanities with the task of quantifying historic corpora through phrase searches and comparative analysis. Although there are many examples of search tools applied to language corpora, for example, the works of Shakespeare, many of these are specific to small groups of researchers and to the specific corpus. In addition, the user interface of these systems can be complex and parameter-laden, and they only provide basic retrieval capabilities.

Samtla adopts methods developed in Information Retrieval and Data Mining, including character-based suffix trees, probabilistic language models, and statistical analysis; presenting search results according to the underlying principles and structure of the language present in domain specific corpora. *Samtla* is designed to be language agnostic allowing it to be generalisable to any body of texts in any language, due to the probabilistic nature of the approach adopted. The current version of *Samtla* features exact, partial, and approximate query matching. Each text is also supplemented by relevant metadata. *Samtla* is currently being used for the study of Aramaic Magic Incantation texts from late antiquity (6th to 8th CE) to aid an international group of researchers with the analyses of the diverse subject matter and literary formulae that occur within these texts that are written in a number of related dialects; including Jewish Aramaic, Mandaic, and Syriac. Their work involves understanding the dialects in terms of their wealth of lexical and grammatical forms, and in ascertaining correct readings in order to produce reliable transcriptions, translation and literary analysis. *Samtla*, will enable these researchers to expand on their current approaches and methodologies through the use of digital technologies.

Visual Interfaces to Collections

Andrea Scharnhorst, Royal Netherlands Academy of Arts and Sciences

Humanities research often relies on rich collections of artifacts as found in musea, special collections of research institutions, archives, and libraries. Most of these 'knowledge spaces' can be accessed on-line meanwhile. However, when sending off text-based queries the usual response takes the form of lists of hits, albeit based on sophisticated information retrieval methods. In some cases returned information is enhanced with images of the objects, now and then 'woordles' of related search terms are offered as visual feedback for searching and browsing. But, in most cases, the size and composition of the interrogated collection remain hidden to a virtual visitor.

In this paper, we propose a different visual interface to collections. We present two recent experiments: interactive visual interfaces to EASY, and visualization of NARCIS content (both services of Data Archiving and Networked Services - DANS). EASY is a large digital archive mainly containing datasets from the humanities and social sciences, among them a rich collection of archeological material. Some of their depositors also feature in NARCIS, which is the Dutch national research information system.

We discuss possible benefits of visual enhanced browsing for different user communities, including humanities scholars. We also discuss the co-construction of new interfaces in a collaboration between humanities fields (as producers and users of content) and information specialists. Eventually, we discuss how visual interfaces to different collections can be used to create one common knowledge space to resources, objects, projects, publications and researchers in the field of digital humanities.

Thursday 6th September, 14.00 – 15.30: Session 3

Navigating Cultural Heritage Collections using Pathways

Mark Hall, *University of Sheffield*

Significant amounts of cultural heritage material are now available through online digital library portals. However, the size and lack of organisation of these collections can be overwhelming for many users who are provided with little or no guidance about how to access the information they are interested in or even what is contained in the collection. Potentially useful and relevant content is hidden from the users who are typically offered simple keyword-based searching functionality as the entry point. The situation is very different within traditional mechanisms for viewing cultural heritage content (e.g. museums) where items are organised thematically and users guided through curated collections.

The PATHS project (www.paths-project.eu) is developing a system that acts as an interactive personalised collections guide, supporting exploratory search and discovery through existing digital library collections. Navigation in this system is based around the metaphor of paths (or trails) through the collection. These paths are more than simple lists of items; they also contain links to similar items and connections to information both within and outside the collection that provide context and encourage exploration of the collection. Both experts and general users can create such guided paths and they can be based around any theme, for example artist and media ('paintings by Picasso'), historic periods ('the Cold War') or places ('Sights not to miss when visiting Venice'). The paths system currently provides access to two large collections of cultural heritage items (Europeana and Alinari 24 Ore). The talk will describe the overall project objectives, including demonstration of a prototype system, with emphasis on its functionality, development and evaluation. This is joint work with Nikos Aletra, Paul Cough, Samuel Fernando, Nigel Ford, Paula Goodale, Mark Hall and members of the Paths consortium.

Thursday 6th September, 14.00 – 15.30: Session 3

Participating in Search Design: Understanding Scholarly Practices

Keira Borrill, *University of Sheffield*

The AHRC-funded project 'Participating in Search Design: a study of George Thomason's newsbooks 1649-1653' is a collaboration between the Humanities Research Institute and the departments of History, English and Sociological Studies at the University of Sheffield. The principal aim of the project is to investigate and understand how search is used within humanities research methodology and to apply this knowledge in the design of better search interfaces for online resources which genuinely meet the needs of the research community. We are exploring the use of a Participatory Design approach in developing search tools through focus groups and design groups, which will ensure that potential end users are actively engaged at every stage of the technical design process as creative agents rather than reactive consultants. Our test dataset is a sample of the Thomason tracts, currently only available as facsimiles through Early English Books Online.

This paper will report on the first phase of this 2-year project which began in February 2012. It will explain the aims, objectives and methodology of the project in detail and identify where the research fits within current knowledge and issues in the digital environment. It will then present a summary of the data gathered from a UK-based landscape survey of research practices of around 500 academics and postgraduates in the fields of History, English Literature, English Language and Linguistics, Politics and Journalism and from a series of focus groups held around the country. An initial analysis of key themes arising from the survey and focus groups will be discussed, followed by an insight into what those findings might suggest for the next steps of the project, including how best to approach the design of the test resource through a series of Design Groups.

Thursday 6th September, 14.00 – 15.30: Session 3

A History of Archaeological Investigation in Post-War England: Insights Provided by Digital Resources

Tim Evans, *University of York*

In recent years, the amount of archaeological digital data available online has increased significantly, particularly as resources that have been deposited with the Archaeology Data Service (ADS). Amongst these resources are several databases relating to archaeological investigations: the NMR Excavation Index, The English Heritage Geophysical Database, The OASIS project, Grey Literature Library and the CBA Index to Radiocarbon Dates from Great Britain and Ireland. These digital resources can, and have been, used to aid conventional research on period or site-type specific research questions, such as 'Bronze Age sites in Southern Britain' and the results presented through familiar distribution maps and academic syntheses (cf, Bradley 2007; Fulford and Holbrook 2011). However, the same resources also represent a substantial corpus of data relating to the practice of archaeology itself, for example where investigations have taken place, in what numbers and by whom. This presents us with a fascinating insight into the development of the discipline and a comparative analysis of the datasets highlighting reveals distinct socio-political and geographical biases manifest in the archaeological record.

Such reflective meta-analyses are increasingly important for a knowledge-based discipline such as archaeology. In much the same way that recent work in metric analysis of academic journals has been used as a tool for highlighting cultural trends within archaeology (Bateman and Jeffrey 2011); this paper suggests that a similar process can be applied to the broader record of archaeological work itself. This is of importance, given the post-war diversification and fragmentation of practice, recording and publication (Jones *et al* 2001), and that previous overviews of 'archaeological work' have tended to rely on piecemeal literary analysis, anecdotal evidence or a single dataset to present the 'history of archaeology' as a process of logical linear development (Gardiner and Rippon 2009). It is arguable that archaeological practice is not always amenable to such simple analysis, but it is a partial and subjective discipline with a complex relationship with society and priorities and biases reflecting its antiquarian predecessors.

Uncovering the “Hidden Histories” of Computing in the Humanities 1949 – 1980: An Overview of Our Key Findings

Julianne Nyhan and Anne Welsh, *University College London*

The origin of the field now known as Digital Humanities (DH) is usually traced back to 1949, when Fr Roberto Busa, an Italian, Jesuit priest, began work on an index variorum of some 11 million words of medieval Latin in the works of St Thomas Aquinas and related authors (Hockey 2004). McCarty has argued “For computing to be of the humanities as well as in them, we must get beyond catalogues, chronologies, and heroic firsts to a genuine history. There are none yet.” (McCarty 2008: 255). Our Hidden Histories project has aimed to make a contribution to the writing of such a history.

The project was a pilot that used a mixed bundle of methodologies, including bibliographic analysis and oral history to identify sources and information that are not mentioned in the published, scholarly literature of the field. In this paper we will present the key findings of this work.

We have identified a list of scholars active in the field from 1949 to 1980 by compiling a bibliographical database of early publications including *Computers and the Humanities* (1996-2004); *Literary Data Processing Conference Proceedings* (1965) and *The Computer in Literary and Linguistic Research* (1971). This list will be discussed with reference to names from the earlier period that are often forgotten or overlooked. The paper also sets DH scholars' publishing patterns in such outlets in a comparative context with the Humanities as a whole. Our key findings include

- (1) the rates of single and joint publication in DH compared to available data on the traditional Humanities;
- (2) the trajectory of publication patterns in *Computers and the Humanities* from a comparative perspective in terms of the Humanities as a whole along with *Literary and Linguistic computing* and *DHQ*
- (3) information and analysis of the levels of formal training in both computing and humanities among what we call first and second generation practitioners (i.e. up to 1980);
- (4) the myth and reality of the commonly expressed notion that computing was perceived as being dangerous to one's academic career;

30 Digital Humanities Congress 2012, The University of Sheffield

(5) discussion of the administrative and policy work that has been so central to the field but tends to be of lower profile than the formal scholarly work.

Thursday 6th September, 16.00 – 17.30: Session 4

History as Dialogue

Chiel van den Akker, *VU University Amsterdam*

“The goal [of digital history] is not to displace argument, synthesis, interpretation, and understanding in favor of a celebration of infinite possibility, but to broaden the participation in a dialogic process of engagement, questioning, and reflection on answers.” – Michael Frisch¹

It is indisputable that the virtual reconstruction of Rome, the real time simulation of the Apollo 11 flight, and the searchable archive of social life in early modern London, are important and exciting means of presenting history with the use of new media. However, these genres are not digital alternatives to *academic history writing*.

For a long time, the book has been history’s medium and the monograph its genre. The monographic narrative, however, no longer seems appropriate in a digital environment. I will argue that the dialogue as a genre is the most likely candidate to achieve online what the monograph aimed at before: providing a comprehensive synthesis of the past. This online dialogue might develop out of the already existing discussion networks.

Two arguments will be put forward. 1) What is regarded as typical of digital media – it is variable, interactive, dynamic, collaborative, and consisting of hypertext – is in agreement with the dialogue. 2) Historical monographs provide panoramic interpretations of part of the past. Such historical *understanding* can also be achieved by means of an online dialogue.

This second argument allows me to partly criticize the supposed rapprochement between the sciences and the humanities in digital humanities. Admittedly, there might be such rapprochement on the level of *methods of analysis*, however, on the level of *understanding and writing*, the sciences and the humanities will remain distinctive realms of thought.

1. Cohen, Frisch, Gallagher, Mintz, Sword, Taylor, Thomas III, and Turkel, “Interchange: The Promise of Digital History”, *Journal of American History* Vol. 95 no. 2 (2008) 452-491.

Investigating Networks over Time: *Matrixify*

John Haggerty, *University of Salford*

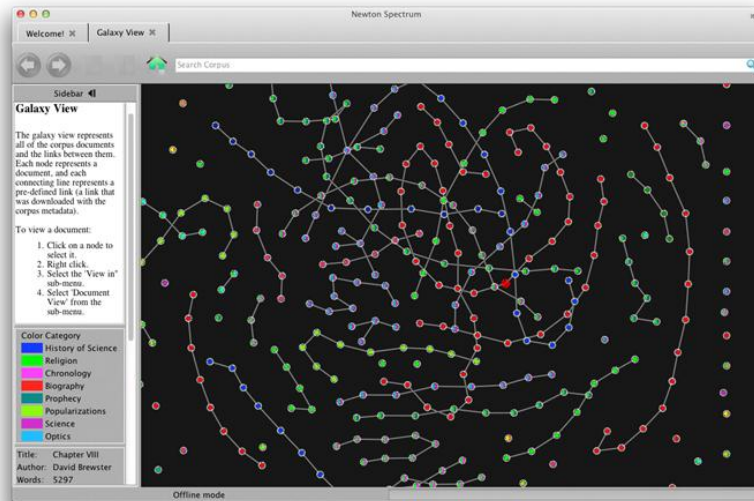
Scholars in the arts and social sciences have become increasingly interested in networks, and more recently, to problematise their understanding of them. This has led to questions concerning the use and abuse of networks, power play, and network development, all of which require relational *and* temporal analysis. At present however, network visual analytic tools are limited to static representations; they show a snapshot of a moment in time in a network.

This paper presents *Matrixify*, a new tool developed through collaboration between an historian and a computer scientist, to meet the requirements of those interested in networks. Primarily aimed at users from the arts and social sciences, this tool has been designed to allow users with limited knowledge of visual analytic applications to view and assess how their network changes over time. The tool provides an interactive visual interface with which a researcher may explore and view their data, usually on one page. This is important because it facilitates the asking of new and important questions of the data. For example, are some actors active for only a short time, if so why? Conversely, why are some actors inactive in the network at certain points? Why is a network denser during certain periods? Who represents the network and why (are there cliques?, interlinked subnets?, etc.)? What endogenous and exogenous events shape their involvement? How does shifting actor engagement impact on access to information and financial and social capital? This paper will use a historical data set, membership of the African Company of Merchants trading from Liverpool over the period 1750-1810, to demonstrate *Matrixify*. In this way, it will show how these questions might be asked via the tool and what answers may be suggested by the analysis. More importantly, it will share the experience of conducting a digital humanities project from two very different disciplines' viewpoints.

Thursday 6th September, 16.00 – 17.30: Session 5

Newton Spectrum: Nonlinear Text Browsing of a Large Corpus

William Martin, Dani Abdallah, Ahmed El-Abiary, Yosef Dalbah, Simon Julier, Melissa Terras
 University College London, Rob Iliffe, University of Sussex. Michael Hawkins, University of Sussex
 and University of Cambridge and Tim Weyrich University College London



When he died in 1727, Isaac Newton left over five million unpublished words on theology, alchemy, physics, optics and mathematics. Composed over a lifetime, Newton's archive contains hundreds of thousands of novel or re-used blocks of texts that make up various 'projects', often without a strong topical focus or a defined sequential order. Studying such an unprecedentedly complex and high-granularity dataset provides unique insights into Newton's findings, opinions, and intellectual behaviour. The *Newton Project* at the University of Sussex (<http://www.newtonproject.sussex.ac.uk>) is the largest effort to diplomatically transcribed and to make available many of Newton's writings online.

However, the sheer size and complexity of the corpus renders traditional manual approaches to studying it impractical and potentially problematic. Furthermore, modern digital scholarship tools often present text fragments in a sequential order, be it topically sorted, by date or by record number. We believe that any such ordering introduces an implicit bias that obfuscates the essence of Newton's nonconventional creative process. To overcome these issues, UCL together with the University of Sussex has developed a non-linear text browser, called *Newton Spectrum*. Using a set of interlinked graphical and textual views, *Newton Spectrum* provides researchers with a means to explore the different blocks of text and edit the relationships and arrangements between. To aid in this process, unsupervised machine learning algorithms have been used to identify topics and propose relationships between the blocks.

Newton Spectrum illustrates the kind of issues that arise when undertaking visualisation of large-scale textual corpora. In this paper we will demonstrate both the system, the thought process which led to its development, and suggest further work that will be applicable to all scholars wishing to use computational methods for researching and visualising complex corpora.

Thursday 6th September, 16.00 – 17.30: Session 5

Rich Internet Application for Collaborative Numismatics

Romain Janvier, *Université of Pau* and **Guillaume Sarah**, *CNRS*

In the last few years several institutions have undertaken the digitization of their numismatic collections, and their diffusion on the Internet. In hindsight, this has developed in an anarchic way. Each institution focused first and foremost on acquiring and modelling data according to its own needs, giving only little thought to standardization and the developing of more global solutions. This is hopefully going to change in the near future, as members of the numismatic world have started to take matters into their own hands. Most notably, a domain ontology has been created through the writing of the NUDS specification (Numismatic Database Standard). However, little attention seems to have been given to renewing the access paradigms to this data. The existing web databases offer only conventional research and filtering tools, which are often unsuited to the users' requirements.

The Université de Pau's research project on the ancient coins of Bearn¹ involved the creation of a corpus of about 3000 coins, to be studied by no fewer than twenty researchers from various fields (archaeologists, archaeometrists, historians, numismatists) and with heterogeneous skills and needs. The data had to be made accessible to all of them through a computer interface versatile enough that it would satisfy the expectations of each of them, and be an asset to their research. The existing data-accessing softwares, with their simple user interfaces, seemed suboptimal. The requirements of this research project, combined with reflexions and software developed during prior coin hoard studies, led to the creation of a few innovative and more efficient technical solutions for sharing, accessing and handling numismatic data.

In this paper, we will be presenting our first attempts at developing such tools, both from a technical and conceptual point of view, and through a discussion of the practical results of their implementation during the Coins of Bearn project. We started out by choosing to use robust open source software components commonly used on server side (Python, Django web framework, relational DBMS). As a traditional dynamic website architecture was unsuited to our goals, we created a Rich Internet Application (RIA), tailored to our field of research, by making use of the commoditization of asynchronous and real time capabilities server side, and new client sides features (SVG, HTML5, CSS3). These very recent web technologies, which have been little used in cultural heritage information systems so far, made it possible to structure the workflow around both coins and data mining, thereby creating new opportunities for collaboration between researchers.

1."Les Monnaies de Béarn (XIe-XVIIIe siècles)". EA 3002 ITEM, Université de Pau et des Pays de l'Adour – France. For further information on Numismatics content: <http://item.univ-pau.fr/live/Les+monnaies+de+B%C3%A9arn>

Thursday 6th September, 16.00 – 17.30: Session 6

From Individual Solutions to Generic Tools: Digitization at the Max Planck Society

Andrea Kulas and Lu Yu, *Max Planck Digital Library*

Max Planck Institutes are increasingly digitizing their library holdings for improving accessibility and supporting research. Although most institutes have similar needs in those respects, they often realize their own individual projects to digitize their specific objects, thereby dispensing with efficiency and re-usability. The project Digitization Lifecycle (DLC) was initiated by the Max Planck Digital Library and four Max Planck Institutes from the Humanities and Social Sciences. DLC has started in February 2011 and will last two years. The aim is to develop a generic application providing functionalities for indexing, editing and publishing digitized works as well as browsing, searching and semantic content enrichment. At the same time general guidelines are produced for digitization projects at the Max-Planck-Society.

The major challenge in DLC lies in creating a generic service to support digitization endeavours. As such DLC replaces isolated solutions at the institutes based on their individual requirements. In individual projects bibliographic metadata is stored in many different formats and degrees of detail. Full texts available in the TEI format are often tailored to very specific needs such as different genres and digitization workflows depend on the objects to be digitized. For a generic solution the partners need to agree on common formats for bibliographic data and full text, still allowing different genres to be uploaded. Functionalities such as upload and update procedures for self-ingest are implemented and an online editor for structural data allows producing a rudimentary TEI Format, which can be exported and reimported. The enrichment of the digital material in an online environment is thereby an important focus of DLC. The presentation will demonstrate the balance achieved between individual and community needs and the limits as well as benefits which can be offered by generic tools at the service of Digitization Lifecycle.

Thursday 6th September, 16.00 – 17.30: Session 6

Digital Humanities in Practice

Smiljana Antonijević and **Sally Wyatt**, *Royal Netherlands Academy of Arts and Sciences* and **Monica Bulger** and **Eric Meyer**, *University of Oxford*

This paper presents findings of a fieldwork study that explored research practices, challenges, and directions in contemporary digital humanities scholarship. The study was conducted in the period April-October, 2010, as part of two research projects of the Royal Netherlands Academy of Arts and Sciences and the Oxford Internet Institute. The studies included observations, focus groups, and in-depth interviews with digital humanities scholars, policymakers, and funders, with a focus on developers and users of digital resources for humanities research. The study involved 92 participants from over 25 institutions in 5 countries

This paper focuses on the following themes, which emerged from the empirical study as focal issues in contemporary digital humanities scholarship:

- *epistemological and methodological issues* brought by the use of digital resources in humanities research, such as the ways in which use of digital resources influences collaborative work, methods of inquiry, types of research questions asked and answered, and dissemination of scholarly results.
- *challenges in developing tools, data and infrastructure* for humanities research, including the dilemma of developing generic vs. field-specific tools; difficulties of data preservation; and the challenges of developing multidisciplinary and multinational infrastructures.
- *educational issues*, such as the design of university programs and degrees in the area of digital humanities; teaching digital humanities methods; educating senior researchers.
- *the future of digital humanities scholarship*, including infrastructural and financial sustainability, changing methods and questions, and the general profile of the field.

The informants consulted in the study are engaged in the exploration and development of various digital humanities projects and initiatives, which grants this paper a unique opportunity to comparatively explore a range of strategies, challenges and directions in contemporary digital humanities scholarship. In addition, the paper will address actions that might be taken to improve the use of digital resources in humanities research.

An Institutional Framework for the Digital Humanities: An Alternative to the DH Centre

Erin Snyder, *University of Oxford*

Digital humanities is growing in the UK. In some cases, this is evident in the formation of new centres and departments dedicated to the study of DH. However, many institutions are struggling to develop support and expertise in the digital humanities and do not have the resources to create a dedicated centre, while others are doing very fine DH work throughout many departments of the institution.

Using the Digital.Humanities@Oxford project as a case study, I will present the process of developing a decentralized institutional infrastructure for digital humanities. Oxford has considerable expertise and activity in digital humanities; this activity and expertise is also very widely dispersed across the university, presenting a heightened challenge in locating and networking it. I will discuss identification of resources (people, existing projects, support provisions and collaborative possibilities), requirements for reference materials and pathways to engagement, and creating a “banner” or umbrella for broader activity at the university. I will also discuss the problem of sustainability of digital resources in the context of a fragmented collection of projects; while this is and will remain an ongoing problem with no easy solutions, collaborative work can improve the case for institutional support of resources, and for creating increasingly sustainable resources. I will also examine the limits of the virtual network; institutional support of some sort is vital, and while there are possibilities to expand this model beyond the institution, such expansion carries further difficulties.

Decentralized DH provides an opportunity to engage a wide community, particularly as this form of DH is often, by necessity, broadly multi-disciplinary. In this way, a digital discipline can make use of digital resources to compensate for physical dispersal, and to promote new work that might otherwise have been impossible.

Friday 7th September, 9.30 – 11.00: Session 7

Designing a National “Virtual Laboratory” for the Humanities: the Australian HuNI Project

Toby Burrows, *University of Western Australia*

This paper looks at the design and architecture of the Humanities Networked Infrastructure (HuNI), a national “Virtual Laboratory” which is being developed as part of the Australian government’s NeCTAR (National e-Research Collaboration Tools and Resources) programme. NeCTAR is funding discipline-based “Virtual Laboratories” to integrate existing capabilities (tools, data and resources), support data-centred research workflows, and build virtual research communities to address existing well-defined research problems.

HuNI has been funded from June 2012 until the end of 2013. It is being developed by a consortium of thirteen institutions, led by Deakin University in Melbourne. It aims to cover the whole of the humanities (defined as the disciplines covered by the Australian Academy of the Humanities) and is designed to bring together the content from numerous digital services which have already been developed by collecting institutions (libraries, archives, museums and galleries) and academic research groups. The Linked Data framework will be used as the basis for combining this content.

A range of different Open Source tools have already been developed for working with Australian digital humanities collections. HuNI will adapt a selection of these tools to work within the Linked Data environment and will supplement them with tools developed internationally. HuNI will enable researchers to find and analyse data across a range of humanities disciplines, and to save the outputs of their analysis in a variety of forms, including compound digital objects, annotations, maps, timelines, and graphs. They will be able to share their results and outputs with other researchers.

Friday 7th September, 9.30 – 11.00: Session 7

More than Meets the Eye: Going 3D with an Early Medieval Manuscript

Bill Endres, *University of Kentucky*

Presenting 3D renderings on the Web provides scholars with intriguing possibilities, particularly with 3D models offering additional information for scholarship and efforts of preservation, like the ability to take accurate measurements of any feature of a rendered 3D image. Current solutions are not perfect, but stable and Open Source possibilities are emerging. While HTML5 and WebGL (Web Graphic Library, a JavaScript API for rendering interactive 3D graphics) appear the long-term solutions, different browsers can show varying issues when portraying 3D content, and Microsoft has yet to include support in Internet Explorer for WebGL. Solutions like JSC3D, an Open Source 3D object viewer, offer reliability but less robust solutions (and, again, lack Microsoft support). Concurrently, Adobe offers 3D capabilities for pdf's and Flash. Adobe's pdf reader includes dynamic tools for tasks like measurements. Yet, pdf's offer a cumbersome and less elegant solution for Web viewing.

In my presentation, I will examine and assess approaches for rendering 3D images on the Web as part of my online version of the St Chad Gospels, an 8th century illuminated manuscripts at Lichfield Cathedral. In my assessment, I will be sensitive to paleographic and conservatory issues like assessing cockling, measuring holes and features of interest, and magnification. While 3D images offer a sense of depth perception, native to examining a manuscript, I want to focus on 3D renderings that offer tools and magnified views that reveal significant information. I will discuss my short and long-term solutions for the St Chad Gospels and their trade-offs. Long-term solutions will prove particularly beneficial as a manuscript ages; they will offer benchmarks for identifying folios that might require special care.

Friday 7th September, 9.30 – 11.00: Session 8

Transcribing Early Medieval Epigraphy in the Digital Age

Guillaume Sarah and Florence Codine, *CNRS*

The digitization of patrimonial collections and their availability online is a main concern for curators and institutions, especially in the case of ancient artefacts which are often fragile and precious. It raises many questions regarding cataloguing standards and the aims of these new tools, both in the institutions in which the objects are preserved and among researchers. In the case of early medieval epigraphy, the graphic complexity of the inscriptions as well as their abundance has made the endeavour especially arduous. Researchers and curators have always tended to study the inscriptions of this period in a very fragmented manner, especially regarding media: numismatists study coinage, epigraphists stone as well as a few other media such as metallic objects or ivory, sigillographers seals or ivory. Museums, libraries and other institutions also traditionally consider these objects as part of separate, specific collections. Yet in all French early medieval inscriptions, the letter forms and writing practices seem to be remarkably consistent. It has therefore seemed desirable to consult with specialists of various domains – mainly numismatics, epigraphy and sigillography, but also linguistics – in order to determine the most precise and relevant letter forms and graphic peculiarities at the time and update and unify the existing standards of transcription and description.

The object of this paper is to present the current state of these discussions and the proposed perspectives of the font we hope to ultimately derive from them. We will discuss how it would solve the current difficulties encountered in paper and digital publications, databases and catalogues, as well as enable better communication between researchers and institutions and offer the opportunity to widen the scope of studies in this domain, as data could be collected and studied on a far greater scale. More generally, the preliminary work for the creation of this font, and the information which could be derived from its use seems a promising lead for a better understanding of graphic practices and structures in the early middle ages, of the ontology of letters and the semantic content of glyphs.

Cognitive Insights in Interpretation Building: Tailoring Software to Expert Practices

Ségolène M. Tarte, *University of Oxford*

Ancient textual artefacts, as individual objects and as corpora, are the substrate of our scholarly knowledge of ancient civilisations. This knowledge is discovered, extracted, created through the daily practice of interpretation of these ancient documents. The cognitive processes involved in this act of knowledge creation, and the search for strategies to support them digitally are at the core of the research presented in this paper.

Building on previous research that established the importance of encapsulating interpretation strategies in the digitisation of textual artefacts, we seek to expand the functionalities of the piece of software that enables the visualization of the digitised artefacts. Its aim is to further enable the users to record their key choices in their building of an interpretation. At the digitisation stage, it was essential to capture the aspects of the materiality of the textual artefact that enabled interpretation, namely that the script is inscribed in the artefact, creating critical zones of highlights and shadows on the object.¹ This was achieved through Reflectance Transformation Imaging,² a methodology, which, with its associated software, forms the starting point of the current research. In order to identify the various cognitive styles mobilised to extract meaning from ancient and difficult to read documents, after having observed how papyrologists approached the task, we have turned our attention to assyriologists attempting to read an as-yet-undeciphered script, Proto-Elamite.

I will present an ethnographic study of assyriologists grappling with Proto-Elamite clay tablets, and the nature of some of the cognitive processes involved in their sense-making activity. Independently of the scholars' respective cognitive styles, their interpretation building pursuit invariably involves image-based semantic perception of the script. Combined with a tailored set of schemes and critical questions (as developed in argumentation theory, a field at the crossroads of artificial intelligence and philosophy) and mind-mapping software principles (where rationale building is typically non-linear), the findings of this ethnographic study inform the design of our piece of interpretation support software.

References:

¹ S. M. Tarte, "Digitizing the act of papyrological interpretation: negotiating spurious exactitude and genuine uncertainty," *Lit Linguist Computing*, vol. 26, no. 3, pp. 349–58, 2011.

² G. Earl, P. J. Basford, A. S. Bischoff, A. Bowman, C. Crowther, M. Hodgson, K. Martinez, L. Isaksen, H. Pagi, K. E. Piquette, and E. Kotoula, "Reflectance transformation imaging systems for ancient documentary artefacts," in *Electronic Visualisation and the Arts*, (London, UK), BCS, The Chartered Institute for IT, July 2011.

Friday 7th September, 9.30 – 11.00: Session 8

Mapping Miss Mansfield: Using Digital Tools to Explore the Role of Place in the Work of Katherine Mansfield and Witi Ihimaera

Anouk Lang, *University of Strathclyde*

A New Zealander by birth who emigrated to Britain and became associated with some of the central figures of the Bloomsbury Group, Katherine Mansfield occupies a unique position within literary modernism, as her texts engage not only with the towns and villages of pre- and postwar Europe, but also the world of colonial New Zealand. As a tribute to Mansfield on the one hundredth anniversary of her birth, the Maori author Witi Ihimaera wrote the volume *Dear Miss Mansfield*, whose characters also move between New Zealand and Europe and, in addition, navigate the cultures of Maori and Pakeha (Europeans). The novella and the short stories which make up Ihimaera's volume simultaneously pay their respects to Mansfield's own stories and reposition them, introducing Maori perspectives and characters in ways that disrupt the eurocentric nature of Mansfield's texts. In different ways, the two authors are both concerned with the dialectic between modernity and tradition, a distinction that maps – both geographically and conceptually – onto different kinds of space, and that sets the urban against the rural, and the colonial against the metropolitan. Taking advantage of the fact that much of Mansfield's work has already been digitised, this paper uses a corpus of her short stories, along with a digitised version of Ihimaera's text, to produce a digital map of the real and imagined geographies underlying this literary conversation across a century. Using geo-spatial technologies to map the locations in the work of these two authors and identify patterns that may not be visible using the tools of conventional literary analysis offers scholars a much more precise indication of how each deploys geography as a symbolic resource, which in turn generates insights into where, and how, non-European cultures and perspectives can be excavated from the canonical texts of high modernism.

Friday 7th September, 9.30 – 11.00: Session 9

Spatial Humanities: Exploring and Analysing Texts within a GIS Environment

Patricia Murrieta-Flores, David Cooper and Ian Gregory, *Lancaster University*

In recent years a number of researchers have been working on developing automated techniques to extract place-names from large corpora so that, by linking them to a gazetteer, the places named within the texts can be mapped either within a GIS or using internet mapping technologies such as Google Earth (Grover et al 2010; Yuan 2010; Gregory & Hardie 2011). While this is a significant development it only represents the first stage of the process. Developing it further we need to be able to demonstrate that texts that have been geo-referenced in this way can be used to make a contribution to knowledge across a range of disciplines within the humanities. To do this we have to be able to identify both the geo-specificity of texts and the qualitative nature of these geographical references. Alongside this, we need to illustrate how textual accounts of place change both over time and across a generically heterogeneous corpus. This paper draws on the work of the European Research Council-funded Spatial Humanities: Texts, GIS, Places project to explore approaches to answering these questions. It explores two contrasting corpora: the first concerned with Lake District landscape writing up until 1900; the second based on the Histpop collection of material from the census and Registrar General's reports from 1801-1931. The paper will demonstrate how GIS-based spatial analysis techniques – more traditionally associated with quantitative sources – can be used to explore the way places are depicted and perceived over time, and can help us to analyse texts beyond the identification of trends and patterns. It will also show how GIS can be used to analyse texts in ways that are more traditionally associated with the humanities by assisting scholars in structuring the texts and in highlighting which parts may demand further close reading. It will argue that the combination of these two techniques will facilitate new geographical understandings both within and beyond the Digital Humanities.

Gregory I.N. and Hardie A. (2011) "Visual GISTing: Bringing together corpus linguistics and Geographical Information Systems" *Literary and Linguistic Computing*, 26, pp. 297-314.

Grover C., Tobin R., Byrne K., Woollard M., Reid J., Dunn S. and Ball J. (2010) "Use of the Edinburgh geoparser for georeferencing digitized historical collections" *Philosophical Transactions of the Royal Society A*, 368, pp. 3875-3889

Yuan M. (2010) "Mapping text" in Bodenhamer D., Corrigan J. and Harris T. (eds.) *The Spatial Humanities: GIS and the future of humanities scholarship*. Indiana University Press: Bloomington, pp. 109-123

Which 'Lancaster' do you mean? Disambiguation Challenges in Extracting Place Names for Spatial Humanities

Paul Rayson, Alistair Baron and Andrew Hardie, *Lancaster University*

It is already possible to apply simple gazetteer-based named-entity recognition (NER) techniques to textual sources in order to extract candidate place names. Combining NER techniques with Geographical Information Systems (GIS) allows the digital humanities researcher to overlay the information on a map and visualise the result. This allows us to ask questions such as “what place is this corpus talking about?”, “what is being said about different places?” and “how has the way that places are represented in the corpus changed over time?”. Together, these techniques have spawned the development of Spatial Humanities (Bodenhamer et al. 2010). However, as the application of these techniques widens and the scale of full-text datasets grows (e.g. EEBO-TCP), we need to refine the techniques to improve their accuracy with both identification of candidate place names and linking to the correct location on a map. For example, *Lancaster* (a city), needs to be extracted whereas *Lancaster bomber* (a plane or type of beer), *Stuart Lancaster* (the England rugby coach) and *Duke of Lancaster* (a nobleman or pub) do not. In addition, linking occurrences of *Lancaster* from a textual source to the correct location in the north-west of England rather than other locations in Australia, Canada and the United States would be vital. In this paper, we argue that techniques from corpus linguistics and natural language processing are essential when tackling both these issues, e.g. concordancing to examine the co-text, part-of-speech tagging to identify proper nouns, frequency analysis and keyness statistics to highlight significant differences in place names between texts. Techniques such as edit distance, letter-replacement rules and phonetic matching also permit spelling variation issues in place names to be overcome when extracting mentions (Baron and Rayson, 2008). Collocation between place names and semantic tags on the surrounding context has already proven useful in identifying topics associated with particular places (Gregory and Hardie, 2011). In addition, we argue that combining collocations between place names and distance information from GIS will provide strong evidence for the disambiguation of location.

References:

Baron, A. and Rayson, P. (2008). VARD2: A tool for dealing with spelling variation in historical corpora. In *proceedings of the Postgraduate Conference in Corpus Linguistics*, Aston University, Birmingham, 22nd May 2008.

Bodenhamer D.J., Corrigan J. and Harris T.M. (2010, eds.) *The Spatial Humanities: GIS and the future of humanities scholarship*. Indiana University Press: Bloomington.

Gregory, I. and Hardie, A. (2011). Visual GISTing: bringing together corpus linguistics and Geographical Information Systems. *Literary and linguistic computing*, 26 (3), 297-314.

Friday 7th September, 9.30 – 11.00: Session 9

New Approaches to Measuring the Impact of the Digital Humanities

Simon Tanner, *King's College London*

This paper will consider issues of impact and cultural value and new modes of measurement for the Digital Humanities. In the last twenty years, academics and cultural organisations have created or obtained a mass of digital resources and delivered these online to a wide range of users worldwide. We know that uptake is considerable, but crucially, what impact have they had on people's lives?

Recent research by Simon Tanner and Marilyn Deegan discovered a huge range of benefits and value in digital resources and collections (www.kdcs.kcl.ac.uk/innovation/inspiring.html). However, to step up to the challenges now facing the digital humanities community we have to move beyond traditional measures that have focussed on web statistics, anecdotal information or evaluations of outputs rather than outcomes. It is clear that we need to gain better evidence of impact. How have digital resources delivered positive change in a defined group and how can we demonstrate this?

This paper will seek to answer many of these questions based upon his research funded by the Arcadia Fund (www.kdcs.kcl.ac.uk/innovation/impact.html). This research seeks to construct a synthesis of existing methodologies and techniques and resolve these into a cohesive and achievable methodology for Impact Assessment of digital resources in the creative, cultural and academic sectors. In this context, our definition of impact is: the measurable outcomes arising from the existence of a digital resource that demonstrate a change in the life or life opportunities of the community for which the resource is intended. The paper will consider what constitute key indicators of value and measures of changes, will present some worked exemplars to demonstrate the cohesive methodology and reflect upon how these relate to the digital humanities.

Friday 7th September, 11.30 – 13.00: Session 10

The Citation Problem in the Digital Humanities

Jonathan Blaney, *University of London*

The Institute of Historical Research (IHR) has been collecting longitudinal data on historians' attitudes to and use of digital resources since 2003. Although the data shows gradual changes in attitude towards the use of digital resources, with increasingly positive attitudes, citation practice has not kept pace even with this slow change in research culture. This paper will use citation practice as a focus for discussion about a cultural problem that still remains in the digital humanities: lack of esteem for digital resources.

Some of the arguments for citing print over digital sources in journals and monographs will be reviewed and it will be suggested that they aim at the wrong targets and involve unclear appeals to authenticity, dominated by privileged notions of the printed text. Comparison with citations for other objects, such as manuscripts, paintings and film can point up these assumptions, as can observing current practice in the sciences.

Within the humanities citation practice varies wildly. Historical Research, the IHR's flagship journal, changes print citations to digital citations wherever possible: other, comparable history journals have a diametrically opposite editor

Friday 7th September, 11.30 – 13.00: Session 10

Mapping the Use of Digital Sources Amongst Humanities Scholars in the Netherlands

Max Kemman, Martijn Kleppe, Stef Scagliola and Renske Jongbloed, *Erasmus University Rotterdam*

The last decade we have witnessed not only an explosion of available digital databases and archives but also the creation of all sorts of tools to open up these archives in new ways, allowing researchers to study more and new sources as well as probing new research questions.

But to what extent do humanities scholars actually use these new sources? Several studies have been undertaken in gaining first insights in the use of digital databases and tools. Nevertheless, these studies generally focus on the 'success stories' and aim to formulate advice for future projects. However, this paper provides the results of an exploratory study on the actual use of digital sources by humanities scholars in the Netherlands, focusing on three research questions: 1) To what extent are digital databases and archives used? 2) Which sub disciplines use digital sources more and which less? 3) Which search techniques are applied? We will present the results of a large survey amongst Dutch humanities scholars and will illustrate the main results with video clips we made while interviewing several researchers.

This survey is part of the European FP7 research project *AXES – Access to Audiovisual Archives* (see www.axes-project.eu) which aims to open up audiovisual archives in new and innovative ways by combining techniques such as automatic speech recognition and visual image recognition. One of the targeted user groups are humanities researchers & educators. In order to gain a better understanding in their current use of digital databases and search techniques, a survey has been set up. Its results will be used to provide the system designers with more background on the users, enabling them to better design the envisioned systems.

Friday 7th September, 11.30 – 13.00: Session 10

Who Killed Christabel? Can Authorship Attribution Solve the Case?

Francesca Benatti, *Open University* and Justin Tonra, *University of Virginia & NUI Galway*

This paper examines the experience of two literary scholars, the editors of the *Thomas Moore Archive*, who seek to apply digital humanities methods in order to solve a long-standing authorship question. Moore was an irregular contributor to the *Edinburgh Review*, but like most other nineteenth-century reviewers, he often exploited the anonymity of his articles to pronounce severe or mocking judgements on new works of literature. The virulent review of Coleridge's *Christabel* that appeared in September 1816 ranks among the most controversial examples of this kind, but its uniqueness lies in its persistent anonymity. Since 1930, a number of scholarly articles, relying on both external contextual clues and internal textual evidence, have indicated either Thomas Moore or William Hazlitt as the most likely authors, but a firm consensus has yet to be reached.

Can the adoption of a digital humanities perspective bring conclusive evidence to this debate? This paper describes the process and the results of a new examination of the *Christabel* review based on digital textual analysis and author attribution technology. Central to the authors' focus is the encoding, collation, and electronic textual analysis of Moore and Hazlitt's identifiable contributions to the *Edinburgh Review* for the period 1814-1816. They will outline their methodology, discoveries, and obstacles in their search for a viable hypothesis for attribution. Particular attention will be paid to the complexities of examining for identifiable authorship patterns texts that survive only through one witness and which themselves contain and mediate other texts. In the light of the new knowledge and perspectives generated by this investigation (both digital and paper-based), should the "Christabel" review be listed among Moore's writings in the *Thomas Moore Archive*?

Friday 7th September, 11.30 – 13.00: Session 11

Analyzing the Carlyle Letter Collection

Mitsunori Ogihara, *University of Miami*

The Duke-Edinburgh Edition of *The Collected Letters of Thomas and Jane Welsh Carlyle Carlyle Letters* (Duke University Press) is a 39-volume collection of 10,000+ letters written by the Victorian historian and essayist Thomas Carlyle (1795-1881) and his wife Jane Baillie Welsh Carlyle (1801-1866). *The Carlyle Letters Online* (<http://carlyleletters.dukejournals.org/>) is its text-encoded, online version.

We (Brent Kinser, Western Carolina University; Neil Johnson, Mitsunori Ogihara, and Dingding Wang, University of Miami) have formed with our students (Yajie Hu and Guannan Zhao, University of Miami) a team to study this large volume of letters using computational analysis. Our study will be conducted from two angles. One is the timing of communication, a topic pioneered in the papers by Barabási (*Nature*, 435, 2005) and by Vázquez et al. (*Physical Reviews*, E73, 2006). In these previous studies written communications are modeled in terms of distribution of time distance between two consecutive events. We are employing the methods presented in these papers as well as developing our own approach. The other angle is the relationship between word frequencies and authorship. In particular, we are interested in finding ways to distinguish among three groups of letters: those written between Thomas and Jane Welsh, those written by Thomas to other people, and those written by Jane Welsh to other people via distant reading.

We have conducted some preliminary analysis using the data from volumes 2 through 5 and are in the process of extending the scope to include the entire corpus of encoded data (35 volumes of the print edition, with 3 more to follow later this summer). Our preliminary timing of communication analysis reveals that the mathematical communication models of Thomas and of Jane Welsh are at two opposite extremes of human behaviors, while our preliminary word-frequency analysis indicates the existence of distinctive sets of words to represent each group.

In this paper, I will present the methods we employed and some of the results we obtained, both of which will lead to a greater understanding of approaches to computational analysis and ways of using that methodology to enhance our understanding of these two prolific, and important Victorian letter-writers.

The Second World War as a Benchmark of Political Morality in Post-War Political Discourse in the Netherlands

Hinke Piersma, *NIOD Institute for War, Holocaust and Genocide Studies*

The Second World War (WW II) has been the most powerful benchmark of political morality in post-war Western societies. Political elites were in permanent need to relate to the defining experiences of warfare, political oppression and genocide. *War in Parliament* entails a study of the impact of WW II in post-war political debates and decision-making in the Netherlands. Our assumption is that an analysis of parliamentary debates will reveal how the political elites constructed narratives evoking the wartime past of their nation to make sense of the present. In this paper we will look at why, how often, when, by whom and in what context parliamentarians referred to WW II. What meanings did politicians attribute to the war-years, what lessons was the war supposed to teach and did these lessons change over time? The answers to these questions will shed a new light on the complex legacies of WW II. In collaboration with computer scientists of the University of Amsterdam, we develop tools that enable us to research large corpora and language resources (in this case *de Handelingen der Staten-Generaal* (Dutch Hansard)) by creating an advanced search engine for this dataset with an intuitive and powerful query language.

The quantitative approach underlying this project contributes to qualitative historical research on different levels. First, it establishes and analyzes patterns and trends, in order to answer questions on how WW II discourses have evolved over a longer period of time. Second, Dutch historiography points at the dominance of resistance in the Dutch political discourse in the early post-war period, while collaboration and accommodation were downplayed. Only from the 1960s onwards we see some changes. A systematic analysis of the sources may answer the question where dissenting voices can actually be found in order to confront the dominant and accepted narrative. Third, this research invites scholars to think about the differences between traditional historical research and research done in the new digital era. To what extent will this new methodology change current practices? And in addition, will a digital approach contribute to improvements in the valorisation and reproducibility of historical scholarship?

Connecting Digital Representations: A City's Urban Spaces and its Collective Memory

Sushardjanti Felasari and Chengzhi Peng, *University of Sheffield*

The concept of collective memory was first introduced in the 1920's by the French philosopher and sociologist Maurice Halbwachs (1877-1945) who focused on collective memory in relation to time, space and history. Halbwachs defined collective memory not as a socially constructed idea about the past, but rather as a socially shared notion, a way that a group conceptualised the past while in the present. He believed that all memory is socially constructed through spatial imagery. More recently, the roles of collective memory in architecture and cities were explained by the Italian architect and designers Aldo Rossi (1931-1997). According to Rossi, the city is the locus of the collective memory of its people. In her book *The City of Collective Memory* (1994), M. Christine Boyer, Professor of Urbanism at the School of Architecture Princeton University, described collective memory as the way the urban public compose their images of the city, i.e. the city as a work of art, the city as panorama and the city of spectacle.

This paper continues to explore the relational study of cities and collective memory with a focus on digital representations of urban spaces and instances of collective memory, and how the two representations may be connected to form a virtual social world in which e-learning about urban history and design may take hold.

Using Google Earth and Google Site as 3D modelling and digital asset organisational tools, we constructed a virtual *Jalan Malioboro*, Malioboro Street in the City of Yogyakarta, Indonesia, as a research site. Malioboro Street is commonly regarded as one of the most memorable places in the City of Yogyakarta. In our initial pilot study, a 300m long section of the Malioboro Street was modelled using Google SketchUp. Pieces of collective memory of Jalan Malioboro, such as records of the history of buildings and places collected from various sources in texts, images, videos, are embedded into the 3D virtual street model. In so doing, we produce an initial example of what we call Collective Memory Enhanced Virtual City (CREATI). The paper concludes with a critique of CREATI with reference to The People's Archive – a photo-centric collective memory gathering project developed by Birmingham City Council, Service Birmingham, In Cahoots and Clusta.

Exploring the Boundary of Architectural Enquiry through Mixed Reality Modelling

Yang Yu and Chengzhi Peng, *University of Sheffield*

The recent popularisation of 'Tangible Interfaces' and 'Tangible Interaction' driven by breakthroughs in the fields of human computer interaction, interaction design, and internet of things has opened up exciting opportunities for architectural researchers and designers to explore possible fusion between physical model making and 3D virtual modelling. This paper examines the affordance of such fusion of physical and virtual media in architectural enquiry as a design thinking and making process.

There are several reasons why such explorations can be worthwhile from a design pedagogy viewpoint: (a) architectural physical model making remains widely practiced in design thinking and communication even in the digital age; (b) physical architectural models retains superior usability to 3D virtual models in terms of enabling agile holistic spatial understanding of how complex environments such as hospitals, university campus, user-configurable office blocks perform; and (c) increasingly, physical architectural models can be augmented with novel tangible interfaces to provide multi-modal and multi-dimensional interactions with a virtual world representing a dynamic context of architectural enquiry.

The paper presents the results from our current working with a group of postgraduate students at the School of Architecture in three small-scale site-specific Mixed Reality Modelling (MRM) projects: (1) Houghton Le Spring Health Centre, (2) School of Architecture, Floor 16, Arts Tower, and (3) The Western Bank surrounding areas of the Sheffield University campus. The three groups share a similar MRM platform but they need to address different issues arising from the thematic and locative content specific to their projects.

As a prototype design, our current MRM platform is built on a combination of architectural physical model-making, 3D virtual modelling, and elements of interactive physical computing that drives real-time tracking and multi-modal projection systems. This basic MRM configuration can be adapted and fine-tuned to an architectural project's specific information content and modes of interaction. The paper concludes with the lesson learned from the current experiment and how the present pedagogical objectives should be refined for future projects which could be of an interest to the wider digital humanities research community.

Representing Norwegian Emigration in the Virtual World through the Eyes of Ola Nordmann

Michael Meredith, *University of Sheffield*

Ola Nordmann Goes West charts the journey of a fictitious but historically informed Norwegian emigrant leaving the area of Voss in Western Norway in the 1880s and travelling via Bergen, Hull and Liverpool to a new life in the USA. As the project recreates his experience within a virtual world, the use of historical documents and accounts are used to construct 3D models to populate it, representing key points along this journey. The virtual world also includes a social networking facility, enabling the project to engage with the descendants of those who followed a similar migration route to enhance and develop the received historical picture.

A speculative hope of the project is that the virtual representation will invoke memories in its visitors who will pass along knowledge that contradicts or confirms the historical accounts that the project is built, and indeed building upon. We also hope that such a virtual world will allow the project to reach a wide and dispersed population of descendants and interested parties to better our knowledge and understanding of the human impact that existed during the numerous emigration journeys.

This paper will present the project and the making of the world of Ola Nordmann and discuss the use of virtual world technology in historical research and representation.

Friday 7th September, 14.00 – 15.30: Session 13

Building Digital Editions on the Basis of a Virtual Research Environment

Tobias Schweizer, Ivan Subotic and Lukas Rosenthaler, *University of Basel*

Since many years the importance of facsimile reproductions has been growing in textual scholarship. The systematic integration of facsimiles in print editions has made them transparent and verifiable for the reader. Furthermore, the conceptual relation between the document and its textual content has been changing: the emergence of the diplomatic transcription emphasized the possibly complex spatial relations of manuscripts. Thus, the constitution of texts cannot be regarded as being unambiguous.

In digital editions, the systematic integration of facsimiles makes even more sense, since digitization projects usually create digital representations of analogue documents by reproducing their visual appearance. Therefore, the digital facsimile has become the primary reference in the world of digital editing.

Within SALSAH (System for Annotation and Linkage of Sources in Arts and Humanities), a web-based Virtual Research Environment (VRE) for the humanities, we are developing a topographical transcription method for digital facsimiles addressing their spatial relations. The method is based on a tool originally developed to define "regions of interest" (ROI) within digital images. Multiple regions can be defined on a digital facsimile (regarding the visual coherence of textual information) and transcribed individually. Each region's transcription consists of a character stream and markup information representing the text's visual attributes (e.g., underline, bold, color etc.). Since XML-based markup such as TEI does not allow overlapping attributes, we have chosen an index-based approach where the markup data is stored separately from the text. Additionally, we are now developing methods to relate parts of the transcription like single or multiple words to digital objects stored in SALSAH. Digital objects may represent concrete objects like books, pages etc. but also abstract concepts without a representation such as a digital facsimiles (e.g., a person, location or event). In this way, a transcription is not only a text but a more complex, networked construct expressing knowledge about interrelations between objects of different kind. Consequently, the digital edition becomes an integral part of SALSAH.

Working with the Financial Records of George Washington: Data and Database

Jennifer Stertzer, *University of Virginia*

The Financial Papers of George Washington Project (FPGWP) is a born-digital project that will publish all of Washington's financial records – a total of several thousand distinct documents. These are not only numerous, but very diverse and complicated documents, including receipts, invoices, day books, and double entry ledgers that each comprise thousands of discrete transactions. These documents record in depth and in detail every aspect of Washington's business enterprises, from the account books detailing activities at Mount Vernon and his farms to the books he kept while away at war and during his presidency.

In order to realize their vision, FPGWP is working with the developers of DocTracker. This paper will discuss the following topics: (1) different types of financial documents are formatted in distinct, though standardized, ways, and the formatting of financial documents carries implied meanings; (2) transactions are full of dittos, abbreviations, and short hand, that raise a question of what kind of fields should be created to capture the transcription and clear text; (3) the account books present issues of currency, valuation, and barter; and (4) a hierarchy of documents exist, and therefore the same transaction may be recorded in a day book, account, and ledger, etc., generating multiple instances of the same transaction. How these challenges should be explained and visualized for the user and accounted for in the development of the database and final display of information will also be examined.

This presentation will explore these issues and the technological solutions that will make George Washington's financial documents not only available (as a free online resource) but also intellectually accessible so that users will be able to interact with, search for, and restructure the data to suit their needs.

Data Journalism: Implications and Opportunities

Ester Appelgren, Helge Hüttenrauch and Gunnar Nygren, *Södertörn University*

In 2003 the European Commission adopted the Public Sector Information (PSI) directive. It encourages re-use of authorities' data to create new services. Sweden has a long tradition of free access to public records, but with the PSI directive the formerly paper-based access rights are now expanding to digital records. This provides new opportunities to journalists and changes the role in engaging the general public. Data driven journalism is a form of storytelling where elements of text, interactive graphics and datasets often are combined into one interactive journalistic project. Furthermore, the general public is often encouraged to submit their own data for publication within the project. Data driven journalism is not a new area, but it is due to the expensive production process and the advanced skills needed by the journalists, still fairly uncommon in the media landscape.

This study is based on a joint-research project together with SVT Pejl, the data journalism department at the Swedish public service television company SVT. The main aim of the paper is to explore this emerging form of interactive journalism based on public sector data, and how the general public is engaging with it. How can data journalism be perceived by the general public? What are the implications of the current development for their interpretation of the data selected, cleaned and presented by journalists?

The paper is based on a four month transdisciplinary case study of SVT Pejl and its' data journalism services, taking place in 2011-2012. The methods used are deep interviews, observational studies and a web-based survey. Furthermore, interviews with experts at NY Times, ProPublica and Investigative Reporters and Editors (IRE) have been carried out.

False Memories and Dissonant Truths: Digital Newspaper Archives as Catalyst for a New Approach to Music Reception Studies

Christopher Dingle, *Birmingham Conservatoire* and Laura Hamer, *Liverpool Hope University*

Musicologists primarily use music criticism as the raw material for understanding the reception of works and musicians, or to gain a greater understanding of the cultural contexts of music, its institutions, and audiences. However, the British Music Criticism since 1945 (BritCrit) Project, has thrown up a number of findings which seriously undermine some of the accepted 'facts' of post-war music history, and open the door to a more nuanced perspective. The BritCrit Project initially aimed to critique the widely-held assumption – often centred upon anecdotal claims – that music criticism has declined in both quantity and quality over recent decades by examining the evolving trends of music criticism in British newspapers since the Second World War. The project had no *a priori* revisionist agenda; the expectation was that the findings would confirm the prevailing view of music criticism being in decline. That the trends which it has already uncovered appear to challenge many of the accepted understandings of music since 1945 is surprising and suggests that new insights into reception might be gleaned through interrogating digital archives.

A blended methodology was developed which enabled qualitative observations to be underpinned by quantitative, statistical results. In the first phase, music criticism from *The Times* was sampled across the past 67 years and subjected to qualitative analysis. Qualitative observations were then placed within a framework of quantitative evidence through a database created using metadata from the reviews. Contrary to common perceptions, British music criticism has seen expansion in recent decades rather than decline. This paper briefly discusses the methodological challenges faced by the project, along with the advantages and pitfalls encountered in working with digital newspaper archives. It also considers how digital archives have acted as a catalyst for an unexpected new approach to reception studies in music, and the resultant opportunities and challenges.

Digital Historians: Perspectives and Approaches in Research and Teaching

Claudia Favero, *Open University*

The purpose of this research is to investigate the emergence of “digital history” as “the use of digital media and tools for historical practice, presentation, analysis, and research” (Wikipedia, 2011). Research suggests that most historians have been reluctant to explore the opportunities provided by Internet technologies (Cohen, 2004a, b): the “digital” historians, who use and promote digital history, still constitute a small minority. At the same time, methodologies and tools of digital history have also not yet been studied in a theoretical perspective in general terms (Boonstra et al., 2004).

This study focuses on two contexts known to the researcher compares and contrasts the situation in the UK and Italy and investigates in particular the emergence of other actors, such as non-academic and “amateur” historians. Its main aim is to consider how tradition and innovation compete to determine the present and future of the epistemology and *modus operandi* of the discipline.

In this paper I will discuss data gathered in the first phase of the investigation, through in- depth interviews with digital historians and humanists, both inside and outside academia, as well as a documentary analysis of online historical projects, websites, podcasts, blogs and repositories. Several concepts have started to emerge: for instance, the relative isolation of these individuals as far as their initiatives in this field are concerned, but also a potential for innovation in history teaching and research, through, possibly, new relationships between academia and civil society, based on a common interest in historical themes and the new communication and cooperation tools provided by technology.

The research methodology is inspired by Classical Grounded Theory (Glaser and Strauss, 1967; Glaser, 1998, 2002) and aims at the creation of a substantive theory of digital history as practice and in relation to the changes and developments in the profession of historian.

References

Boonstra, O., Breure, L., Doorn, P. (2004), 'Past, present and future of historical information science', *Historical Social Research / Historische Sozialforschung*, vol. 29, No. 2; available on http://www.dans.knaw.nl/nl/dans_publicaties/ppf/ (accessed 24 June 2008)

Cohen, D. J. (2004a), 'History and the second decade of the web', *Rethinking History*, vol. 8, no. 2, pp. 293-301

Cohen, D. J. (2004b) 'Digital history: The Raw and the Cooked', *Rethinking History*, Vol. 8, no. 2, pp. 337-340.

Glaser, B.G., & Strauss, A. L. (1967). *The discovery of grounded theory; strategies for qualitative research*. Chicago, Aldine Pub. Co.

Glaser, B.G. (1998), *Doing grounded Theory: issues and discussion*, Mill Valley, CA : Sociology Press

Glaser, B.G. (2002), *Conceptualization: On Theory and Theorizing Using Grounded Theory*, *International Journal of Qualitative Methods*, 1 (2) 23-38

Wikipedia (2011), *Digital history*, http://en.wikipedia.org/wiki/Digital_history [accessed April 10, 2011].

Friday 7th September, 14.00 – 15.30: Session 14

Space and Place in Victorian London: Interactive Mapping for Humanities Research

Jennifer Bullock, *Adam Matthew Digital* and David Heyman, *Axis Maps*

This paper is a collaboration between Adam Matthew Digital and Axis Maps, both award-winning publishers of interactive digital resources in the humanities. It will demonstrate the added value technology can bring to humanities research by discussing Adam Matthew Digital's online resource 'London Lowlife' and specifically the ground breaking interactive maps created for the resource by Axis Maps. It will explore the way in which interactive maps are used in 'London Lowlife' to facilitate the interpretation of primary sources in new and exciting ways that are only possible within the sphere of digital humanities. Within the London Lowlife map, there are three conceptual parts that work seamlessly together through a single interactive map interface. Map users can browse primary source documents that have been 'spatialized' to show their position within London. Allowing users to browse the entire collection by location reveals relevant documents related by their proximity to one another; an important relationship that a text-based search might have missed. Pseudo-3D streetviews allow a user to place themselves inside a map of Victorian London to literally look around, giving a 'sense-of-place.' This immerses the user into the map experience, revealing detail otherwise missed in a static environment. Finally, in addition to providing insight into the qualitative aspects of Victorian London geography, this map also explores the changing character of the population through quantitative thematic mapping. Simple demographic measures, viewed over time, can provide a major insight into the changing nature of a city. These three aspects of the map show the range of possibilities that exist to engage those interested in the humanities through interactive mapping.

Friday 7th September, 14.00 – 15.30: Session 15

Citation Networks and Topic Maps: Digital Readings of Imperial Chinese Notebooks

Hilde De Weerd, *King's College London*

Notebooks (biji) became in the twelfth and thirteenth century an established medium in which individuals engaged with published texts and with information shared orally. Their very constitution was based on intertextuality and conversation. The genre therefore lends itself particularly well to an analysis of the networks through which information was exchanged and serves as a major source for a project on the longue-duree history of relationships and structures of information exchange in imperial China. Based on an encoding of a sample of five notebooks and the linking of notebook data with large prosopographical and historical GIS datasets, this paper first surveys prior work on the social makeup of informants, their geographic distribution, as well as the extent, frequency, and temporal distribution of sources cited in each notebook. The presentation focuses on recent work on the design of topical taxonomies, the application of topical markup in TEI, and the visualization and analytical potential of topical frequencies.

Friday 7th September, 14.00 – 15.30: Session 15

Using GIS Technologies to Explore the Disciplinary Reach and Geographic Spread of British Designers

Karina Rodriguez-Echavarria, Leah Armstrong, Catherine Moriarty and David Arnold, *University of Brighton*

The use of Geographical Information System (GIS) technologies to map locations of interest has been widely adopted. However, its use to support humanities research has not been fully explored. This paper will present the application of these technologies in order to analyse the contents of yearbooks for the Chartered Society of Designers. The main objective is to support the exploration of the disciplinary reach and geographic spread of British designers over an eighty year period. For this, a combination of technologies has been used, mainly to digitise the books and create a database with geographic information of members of the Chartered Society of Designers. This information is then displayed in a web based map using the Google Maps API. The interface provided allows the user to browse the data using a variety of criteria, such as classifying the data by year, membership type, discipline, category or gender. In so doing, the project represents the first structured attempt to analyse the image, identity and composition of the design profession. Although the project is still on-going, the first results have shown that presenting the information in a map allows for a better analysis and understating of the data. Furthermore, the project has also identified some challenges with regards to the technology, such as dealing with legacy data and privacy issues.

Friday 7th September, 14.00 – 15.30: Session 15

Analyzing Big Cultural Data Patterns in 4,000 Covers of Veja Magazine

Marcio Emilio dos Santos and **Cicero Inacio da Silva**, *Federal University of Juiz de Fora*

The increasing capacity of computational data analysis is driving computer scientists and designers into the development of new features to visualize and understand cultural artifacts in a different manner. Social scientists, digital humanities researchers are investigating how to create what we can call "cultural algorithms" to discover or reveal new trends about a field of investigation that can be related to film studies, literature, communication and so on and so forth. Following the theoretical approach created by Lev Manovich about his studies related to "cultural analytics", this paper will present a one year research on the visualization of the entire collection of covers of the *Veja* magazine, considered the most important weekly magazine in Brazil. The visualization that we created in our Lab at the Federal University of Juiz de Fora (www.ufjf.br/sws) analyses and demonstrates practical uses for cultural visualization, since we can have critical analytical details about all the covers such as the gender that is more frequent in the covers (masculine), the colors that the magazine uses regularly etc. We can also use image recognition algorithms so we can cross data with Wikipedia, for example, and discover who was more frequently featured in the covers: e.g. politicians or media stars.

Saturday 8th September, 9.30 – 11.00: Session 16

KAPTUR: Examining the Importance and Effective Management of Research Data in the Visual Arts

Leigh Garrett and Marie-Therese Gramstadt, *University for the Creative Arts*

Research data is seen as a valuable resource and, with appropriate curation and management, it has much to offer learning, teaching, research, knowledge transfer and consultancy activities in the visual arts. However, very little is known about the curation and management of this data: none of the specialist arts institutions have research data management policies or infrastructure in place and anecdotal evidence suggests that practice is ad hoc, left to individual researchers and teams without support or guidance. In addition, the curation and management of such diverse and complex digital resources is, in itself, challenging.

Led by the Visual Arts Data Service, a Research Centre of the University for the Creative Arts, in collaboration with the Glasgow School of Art; Goldsmiths College; and University of the Arts London, and supported by JISC, the KAPTUR project (2011 - 2013) seeks to address the lack of awareness and explore the potential of research data management systems in the arts, by: uncovering the nature of research data in the visual arts; investigating the current state of the management of research data; developing a model of best practice applicable to both specialist arts institutions and arts departments in multidisciplinary institutions; and to applying, testing and piloting the model with the four institutional partners.

Utilising the findings of the KAPTUR project to date, this paper will consider the nature and importance of research data in the visual arts; explore the potential benefits its effective management may offer researchers, academic colleagues and students; and provide the audience with the opportunity to consider how developing practice in the visual arts could inform and enhance similar endeavours more widely across the humanities.

Saturday 8th September, 9.30 – 11.00: Session 16

'Art to Enchant': The Creation of a Digital Archive

Michael John Goodman, *Cardiff University*

The aim of this paper is to use my doctoral research as a case study to explore the implications of creating online digital archives that enable students, academics and the wider public to engage dynamically with historical texts and images. I am currently mounting and creating an open access database of the illustrations that appeared in Victorian editions of the works of William Shakespeare. This paper will explain and analyse the issues surrounding such a project, for example, the complex relationship between text and image in illustrated editions, the construction of Victorian notions of Shakespeare, and the ways that the digital archive allows these issues to be explored and enables new and challenging research questions to be asked of the material. The paper will explain how this project provides the opportunity for an interrogation and critical insight into Shakespeare illustration that would be impossible without the tools this archive and Digital Humanities can offer and I will propose that the process of creating a digital archive is a valid means of scholarly enquiry in itself. Ultimately, I will argue, that digital archives are only as good as their underlying structures and interfaces, that their structure should derive from the individual project, and that both the project's content and form should exist symbiotically to form part of an aesthetic whole. Additionally, I will suggest, that when projects are thought of in these terms they are not only more user friendly and attractive to a general audience but, when used by researchers, an intelligently designed system can also generate new knowledge.

Saturday 8th September, 9.30 – 11.00: Session 16

Mapping Metaphors of Wealth and Want: A Digital Approach

Marc Alexander and Ellen Bramwell, *University of Glasgow*

The AHRC-funded *Mapping Metaphor with the Historical Thesaurus* project aims to computationally provide data on the extent and development of metaphor across the history of English. It uses the full database of the *Historical Thesaurus of English*, which extensively categorizes and classifies the recorded vocabulary of the English language from Old English to the present day. By using this database to map semantic categories onto one another, and thus showing lexical overlap in different conceptual fields, we aim to provide results which will demonstrate the widespread, systematic and far-reaching impact of metaphor on English.

Mapping Metaphor is one of the first projects which has originated following the 2009 completion of the *Historical Thesaurus* database, which contains almost 800,000 meanings arranged into 230,000 semantic categories. As such, this paper will outline the statistical and computational methodologies used by the project, and will present a case study of the semantic categories of wealth and poverty, demonstrating the metaphorical links between these categories and the rest of the language.

Website: www.glasgow.ac.uk/metaphor

Print version of the Historical Thesaurus: Kay, Christian, Jane Roberts, Michael Samuels and Irené Wotherspoon. 2009. *Historical Thesaurus of the Oxford English Dictionary*. Oxford: Oxford University Press.

Saturday 8th September, 9.30 – 11.00: Session 17

Promise and Paradox: Accessing Open Data in Archaeology

Jeremy Huggett, *University of Glasgow*

The transformative effect of open access to archaeological data online is in many ways incontestable: the availability of national monuments records has changed the way in which regional and local studies are undertaken, although the availability of site-based data (excavation records, small finds data etc.) has yet to have an equivalent effect. However, as more and varied data are incorporated into online archaeological data archives, the information is increasingly deconstructed into semantic structures which, while they can seem logical, even common-sense, seek to capture the consequences of interpretation, classification, and identification which nevertheless remain largely tacit. Infrastructure solutions for the management and presentation of information can be characterised in terms of two extremes: the integration of information through core ontologies and a range of associated artificial intelligence techniques, or the looser coupling of data which seeks to support diversity and avoid domain-level standards. Each has its own advantages and disadvantages, and each carries implications for the end user in terms of their ability to access and use the data as well as their perception of the level of control and authority implicit in each approach. Indeed, the status of the end user as no more than a consumer becomes increasingly problematic as the tools formalising the information for delivery are increasingly automated. These issues risk overlooking the three key characteristics of knowledge-based work: that it is situated, distributed, and social (N. Van House 2003) and these frequently gives rise to competing, contradictory demands which an information infrastructure needs to incorporate rather than over-ride. Like standards, information infrastructures are necessary, but once they become integrated within practice the forms, decisions and assumptions which underpin them tend to become invisible; consequently there needs to be proper consideration of these factors given the potential for such infrastructures to shape future practice.

Saturday 8th September, 9.30 – 11.00: Session 17

Reperio: a Collaborative Knowledge Environment for Digital Humanities

Damiana Luzzi, *Fondazione Rinascimento Digitale*

In our speech we present Reperio¹, an online collaborative work environment that provides several tools to support research in several field of Digital Humanities and Cultural Heritage sector.

Reperio has a modular and plug-in-based architecture. The main tools are:

1. Ontology Editor
2. Ontology population or data input
3. Text and image editing and management. It makes possible to:
 - a. manage text transcription;
 - b. work, sharing the ontological model, with and on texts.
 - c. insert in both the text and image: semantic, morphological and syntactical annotation, as well as links to related resources, notes, comments, variants and, comparison of several witnesses, etc.

The characteristics of Reperio will be presented on the results of the activities carried on for the management of the complex and articulated documental material owned by Historical Archives of the Pontifical Gregorian University²: printed texts glossed by the author and with notes of the censors; manuscripts (e.g. the seven manuscripts of Christophorus Clavius) with glosses, erasures, insertion of papers and fragments.

This is an exciting challenge which involved Reperio on two levels:

1. To represent the complex physical, structural and textual description of the documents, often printed-manuscript "hybrids", whose levels of description are at least four: published text, glosses, later interventions, correspondence; and to collect at the same time data related to the binding, the state of conservation and restoration of the documents;
2. to reconstruct, starting from the use of manuscript and later elaborations, the history of education and its evolution within the Roman College (the school of Gesuiti).

To manage these levels we developed an ontology, based on international standards³, in order to make available an open and flexible system in which information, with a different granularity, can be integrated and continuously updated by the users themselves.

¹ For information on Reperio you see the website: <<http://www.reperio.it>>.

² Historical Archives of the Pontifical Gregorian University, for the history and the description of the documental material owned in it you see the website: <<http://www.archiviopug.org>>.

³ ISBD <<http://www.ifla.org/publications/international-standard-bibliographic-description>>, REICAT <http://www.iccu.sbn.it/opencms/opencms/it/main/pubbl/pagina_57.html>, MANUS handbook <<http://manus.iccu.sbn.it/>> , TEI <<http://www.tei-c.org/>>, AACR2 <<http://www.aacr2.org/>>, FRBR <<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>>, FOAF <<http://www.foaf-project.org/>>, CIDOC CRM <<http://www.cidoc-crm.org/>>, EDM <<http://pro.europeana.eu/edm-documentation>>.

Saturday 8th September, 9.30 – 11.00: Session 17

Research As Infrastructure

Erik Malcolm Champion, *Aarhus University*

In the edited book *Debates in the Digital Humanities*, edited by Matthew Gold, the chapter "The Digital Humanities or a Digital Humanism" by Dave Parry raised the controversial question as to whether Digital Humanities should be the application of computing, or an inquiry as to how digital media has irrevocably changed the Humanities. While this may appear to be a very theoretical issue, the debate has major practical consequences. For example, I have been entrusted with managing the development of a national research infrastructure for the Digital Humanities. This task may seem to involve logistics, technical details, and general funding issues. However, before we even get to that stage we have major fundamental, political and theoretical challenges.

We currently have four universities as partners, the national library (or libraries) should be joining soon, and hopefully the major museums will follow. Our government has asked that we include as many as possible, a noble goal, but in practice we have hit a major roadblock. How does one create a national focus while allowing academics and other researchers to pursue their own specific goals? This also raises a deeper question, what are the boundaries of the Digital Humanities pertinent to our researchers, beyond which we should not tread? Having discovered our niche, or niches, how can we focus on key research areas important to our country in particular, without becoming cut off from international networks?

Of course there are perennial questions such as how can one develop an infrastructure five years ahead, based on catering for technology that we are not yet using? How can a distributed network allow for unified identity and individual planning? This leads us to a more pragmatic issue of which resources are best managed centrally, and which are best distributed. These more technical issues do however return us to a central problem: how one create a centre for something that has no physical centre, unifying traditionally disparate and sceptical disciplines, without restricting them or discriminating between them?

Ahead of the Curve: Digital Curator Vocational Education

Ann Gow and Laura Molloy, *University of Glasgow*

A critical and often sidelined issue within digital humanities, and the cultural heritage sector more widely, is the ability of those undertaking research in the arts and humanities to care for their data and other digital material over time. Digital humanities research creates rich digital resources but also the challenges of sustaining and managing these objects. Other professionals in the cultural heritage sector also have the responsibility of stewardship of digital material over time. But are those now professionally obliged to perform digital curation receiving the training they need? And what exactly constitutes those training needs? Another pedagogical dilemma in digital curation is whether all staff in the cultural heritage sector should become more proficient in the curation of digital assets, or whether intensely specific training should be developed to enable a distinct strain of specialists to emerge. As digital humanities scholars should we be skilled to care for as well as to create? It is in the context of these debates that the EC-funded Digital Curator Vocational Education Europe project ('DigCurV') is working.

DigCurV brings together a network of partners from across Europe and north America to address the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections. DigCurV is identifying, analysing and profiling existing training opportunities and methodologies, surveying training needs in the sector and identifying the key skills and competences required of digital curators. The project then aims to establish a curriculum framework from which training programmes can be developed.

Our paper will describe the work of the project to date, report our main challenges and the findings from our multinational research, and indicate the future directions of the project.

Saturday 8th September, 9.30 – 11.00: Session 18

Getting Rights Right! - The University of Sheffield Library Experience

Clare Scott, *University of Sheffield*

The University of Sheffield Library is implementing a new Digital Asset Management System (DAMS). This paper draws on our experience in tackling rights either side of the point of digitisation. The paper will be an example of developing best practice for navigating rights issues in the dissemination and management of archival materials as digital humanities resources.

Archive material often embodies different types of ephemera and multi-modal media alongside composite items (for instance annotated maps and plans). It is not easily predictable in terms of the complexities of its format and content. At the same time legislation concerning intellectual property (IP) can also be complex. We have also had to think about contracts and terms of donation, as this has implications for making online versions available where we own the object but not the intellectual property. One approach to getting rights right is to develop a structured policy using alert stages and risk assessment in the lifecycle of the asset: Is the physical item legal? Can we digitise it for preservation? Can we host it on the DAMS? Can a reader make a copy? Can staff make a copy on behalf of a reader? How do we proceed with commercial re-use of material? After outlining our approach to IP policy I will then touch briefly on a parallel procedure to navigate other areas of legal concern such as privacy and data protection. Finally, I will look at issues on the 'other side' of digitisation. Making digital assets accessible must be balanced against protecting our own rights, and also allowing their exploitation both commercially and non-commercially. Digital assets can quickly multiply unattributed through social networks, and we have considered the pros and cons of watermarking technologies and other preventive measures. I ask whether misuse can be prevented whilst engaging with, and benefiting from, contemporary evolving social media platforms.

Saturday 8th September, 9.30 – 11.00: Session 18

A Sense of Tradition in the Digital Archive: The Example of Broadside Ballads

Giles Bergel, *University of Oxford*

Broadside ballads - illustrated printed song-sheets - were once sold in large numbers on street-corners, in town-squares and at fairs by itinerant ballad-singers. Pinned on the walls of alehouses and other public places, they were sung, read and viewed by a wide audience, but have since survived in only small numbers in scattered collections, imperfectly-catalogued (if at all) by standard bibliographic metadata schemes. This paper will outline two, complementary approaches to digitisation: one based on the mass-digitization of historical collections and the other on the description of traditions embodied by discrete items within collections. Examples will include the JISC-funded Integrating Broadside Ballads Archives and a personal research project funded by the British Academy to digitise a ballad tradition made up of exemplars printed in England between 1634 and 1830.

The techniques involved include Linked Data; TEI markup; high-resolution imaging and automated image-recognition; analytical bibliography; computer-aided transcription; iconography; textual criticism ('stemmatology'); and book history. The paper will argue for a holistic approach to digitisation that integrates a humanistic understanding of the source materials at all stages of the digitisation process.

Saturday 8th September, 11.30 – 13.00: Session 19

Improving Record Matching Across Disparate Historical Resources

Stephen Brown, Simon Coupland and David Croft, *De Montfort University*

For many years the Galleries, Libraries, Archives and Museums (GLAM) community has put significant resources into the digitisation of their collections. These digitisation projects have opened up the collections to a much wider audience in terms of the general public and have made them far more accessible to researchers. However, the sheer quantities of information now available, combined with the difficult nature of said information is increasingly highlighting the flaws in standard search techniques. This is to say that ever expanding databases, combined with the imprecise and uncertain nature of the information stored in collection records, is rendering traditional methods such as keyword searching inadequate.

This paper details an investigation into the use of approximate reasoning techniques to successfully search through digitised collections despite the difficult nature of the data. Specifically, it describes an attempt to use approximate reasoning techniques to automatically identify co-reference between records in collections held across different institutions.

This investigation resulted in the development of a novel approach to searching and co-reference identification combining features from Probabilistic Record Linkage, Document Classification and Fuzzy Clustering. The new approach was designed to produce more relevant results than traditional search/identification systems despite uncertain, missing and imprecise record information. Application of this approach and initial results will be discussed with reference to the challenge of linking 19th Century photographic exhibition catalogue records to online museum image collections.

A Semantic Knowledge Base for the UK Government Web Archive: Opportunities for Researchers

Tom Storrar, *The National Archives*

The UK Government Web Archive (www.nationalarchives.gov.uk/webarchive) is a free-of-charge, open web archive of central government websites dating back to 1997. Its purpose is to ensure that UK central government information released on the web is preserved and perpetually accessible. For the public and government alike, it is a rich and valuable resource documenting rapid changes in politics, society, culture, and the environment.

As it contains over 1 billion documents, ensuring content is easy to find and presenting it in a useful way is a challenge. Conventional search has significant limitations when applied to web archives: it does not deal well with the “noise” created by duplicate content. It also relies on keyword-matching: a user may have to perform many highly specific searches before they gather the information they require.

Specific to our domain, a user may need to have specialist knowledge of the government web estate to successfully locate a resource. Additionally, browsing for specific subjects requires the user to know which department was responsible for which functions, at a particular point in time.

The National Archives sought an innovative alternative to conventional search. Thus, the Semantic Knowledge Base (SKB) project was conceived. Some key features include: natural language processing identifies the semantic meaning of generic entities (such as dates, people and places) as well as some government-specific concepts (for example, departments and roles within government) this data, or metadata, is linked together through ontologies, curated by those with specialist knowledge at The National Archives, which describe the relationships between entities. This means that the SKB can return accurate results to the user, even if the user is unaware of the underlying relationships between concepts and data the system uses, and contributes to, the linked open data web - duplicates are recognised, preventing them from being returned in search results, thus producing a “cleaner” user experience

We believe the system will be of great value to researchers across many disciplines, allowing a level of insight not possible through conventional search. The National Archives would like to demonstrate the concept and application of the system to the conference.

Researching Readers Online

Bronwen Thomas and **Julia Round**, *Bournemouth University*

Changes like digitisation are shaping our whole cultural heritage including the responses of future scholars to literature so we need to understand their impact. This paper will present the results of an AHRC funded pilot study that explored the activities of readers online and which considered how far understanding these new modes of engagement might transform our whole approach to researching and teaching literature.

Digital technologies have made it much easier for readers to participate in a wide range of activities around their reading, including connecting with other readers, commenting on a range of texts, and even writing stories themselves. Online communities and fan forums present researchers with a wealth of data not only on what readers are reading, but how they are reading and engaging with texts. This data is potentially invaluable, especially in a context where reading is supposedly in decline, and where new devices such as ereaders are threatening to supplant the printed book. Our project used surveys, focus groups and workshops with established researchers to investigate the implications of these changes, and how they may potentially benefit a wide range of stakeholders, including those working in education and in the creative industries, as well as academics interested in readers and audiences.

This paper will present the results of the above research with the aim of:

- Presenting possible new approaches and insights into the practices of digital readers
- Reflecting on the ethical responsibilities that come with this type of research
- Building on existing scholarship to critically assess appropriate methodologies for future research into digital reader activity.

Saturday 8th September, 11.30 – 13.00: Session 20

Vladimir Nabokov's *Pale Fire* and the Problem of Interface

Simon Rowberry, *University of Winchester*

Vladimir Nabokov's *Pale Fire* (1962) has a long and complex history of remediation dating back to Ted Nelson's apocryphal early hypertext demonstration using *Pale Fire* in 1969, as well as being the subject of two artist's books and a radio play. In the last 20 years, several unauthorized versions of Nabokov's text have been produced for the Web in both English and Russian, as well as a couple of more recent authorized e-book editions and several experiments that were only publically acknowledged, but not shared. The one constant amongst these digitization projects was the feeling that here was a novel that cried out to become an electronic hypertext. This paper will trace the novel's history of digital remediation with a particular focus on the use of interface by the amateurs, and a handful of professionals, who are trying to reproduce and enhance the network present in Nabokov's novel. Through doing so, they have not used scholarly mark-up such as the TEI standards, but rather present the text through a variety of linking mechanisms approximating the possibilities of a digital edition of the text. Since Nabokov's works are still protected by copyright, these editions represent the best current chance to understand how one of the most frequently cited print-based hypertexts can be translated into the digital medium. Much of the paratextual complexity of *Pale Fire* has been undermined through these remediations and this paper will question to what extent it is possible to represent this complexity on the screen.

Saturday 8th September, 11.30 – 13.00: Session 20

In the Mind's Eye: Reflections on Generating Reader Experience c 1350 – 1600

Elisabeth Salter, *Aberystwyth University*

I'd like, tentatively, to propose a paper for this conference as a relative newcomer to Digital Humanities currently working on two projects, one of which is concerned with the enhanced use of the digitised text of Chaucer's *Canterbury Tales* and the other is concerned with enhancing the value and searching possibilities for digitised probate records.

As a specialist in the study of reader experience and textual culture in medieval and early modern society, I have in recent years been working in the field known broadly as The History of Reading. Particularly, for me, on the ways that medieval and early modern individuals used and responded to texts and what their imaginative processes were in their encounters with texts. (see eg Salter and Wicker eds *Vernacularity*, 2011; Salter, *Popular Reading*, 2012, earlier work on probate records is in Salter, *Creativity*, 2006) This has involved detailed case study work examining groups of manuscripts, documents and books to identify perceptions of thematically connected texts, of misconstrued meanings and creative reinterpretations of particular items, of the construction of identity through documentary texts such as the will and testament, of connections between the imaginary of a text and the imagination of the reader/user. One of my favourite images which reflects the fifteenth century notion of the mind's eye is this by Flemish artist Petrus Christus:



As part of working with Keri Thomas (PhD student on the KESS funded Enhanced Digital Chaucer Project) and with Prof Lorna Hughes (project partner in National Library of Wales) I have been engaging with current debates in Digital Humanities concerned with the afterlife and continued usability of digital resources, the issue of data silos, and the need to enhance the use of historic records which have been digitised rather than seeing digitisation itself as an endpoint. [encapsulating and reviewing many of these debates is Warwick, C; Terras, M; Huntington, P and Pappa, N [2008] “*If You Build It Will They Come? The LAIRAH Study*”; as well as recent lectures and papers by Ell, Hockey, Burgess, Hughes, Prescott]

With these issues in mind, I would venture to conjecture that there is ample scope for the enhancement both of Medieval/Early Modern Studies in the History of Reading and textual culture as well as the enhancement of the use of already digitised textual resources through the creation of digital media for the visualisation of the “mind’s eye” of the Medieval & Early modern reader. One of my questions is, in how far is it possible to simulate the mental world of the fifteenth century reader of, for example, Chaucer’s *Canterbury Tales*? This might involve, for example, linking of the digitised text to a range of images such that as words or phrases are passed over, resonances from medieval social, culture or religious life including objects of material consumption, clothing, &c in the form of images and snippets of other texts might resound through this virtual “mind’s eye”. I would like the opportunity to talk about this possibility at the forthcoming Digital Humanities Congress in Sheffield.

Saturday 8th September, 11.30 – 13.00: Session 20

Crowd-Sourcing our Cultural Heritage

Genevfa Kefalidou, Mercourios Georgiadis and Suchith Anand, *University of Nottingham* and **Bryn Alexander Coles**, *Open University*

Humanities disciplines, such as Archaeology, involve the discovery and examination of ancient artefacts and cultural sites. Such activities may be enhanced by embracing technological developments which introduce rich methodologies incorporating user experiences and attitudes when interacting with their environment. Crowd-sourcing is an example of a methodology that helps to create novel and varied data through the acquisition of user data in a digitally-driven environment. Cultural Heritage is a resource for creativity and lifelong learning and has effects on the economic impact and innovation on historical research. The need for sustainable heritage and innovative learning experience seems to become necessary, even more so for less-technologically driven disciplines such as Archaeology.

As such we propose a new twofold direction in Archaeology research, involving the gathering of visitors' personal experiences of the sites they are visiting. The aim of this new approach is to promote a more interactive experience and active learning of historical events and artefacts and to provide a tool for storing and maintaining research activities. Such activities can include personal note-taking on cultural sites, photographs, videos and audio acquisition of experiences of historical artefacts and sites that will be digitised and maintained on web-based media such as dedicated research portals open to the public.

Visitors' personal experiences can contribute to the understanding of a cultural site not only from a historical and knowledge perspective but also from the location/time-awareness and experience perspective. The use of mobile technology and web development will provide a research tool framework for collecting and maintaining data. We argue that the impact of visiting a culturally-vibrant location with artefacts (whether these are physical, digital or location artefacts) will be enriched and experienced on a deeper level rather than on a dry knowledge-based level, if an interactive infrastructure is built for gathering and sharing visitors' experiences.

The Production of a 'Locative Digital Trail' as a Creative, Collaborative Methodology to Investigate or 'Map' Place-Identity

Adam Park, *University of Sheffield*

Born out the digital arts culture, locative media is the linking of digital content to a real-world location using GPS or other real-time locating system (RTLS). Media artists, performance-makers and game designers have appropriated this technology to choreograph relations between user and environment in a playful way. A 'locative digital trail' typically comprises of a series of 'events'; segments of audio-visual content relevant to the context of the participant. These may take the form of stories, traces or clues.

While the use of the computing term 'context-aware' is relevant to the technical aspects of this project, it also refers to the use of 'context' within architectural discourse. To add to these technical and architectural definitions, this paper reaches to performance studies to propose a third meaning of (context) awareness; how, in framing the everyday actions such as walking as 'performative' these emerging models of site-specific practices might "show us, as if for the first time, the world we have always had about us but ignored"¹.

The central question of this study asks whether site-specific, participatory locative practices might be co-opted into the study of the contemporary urban condition. This research aims to contribute to a timely debate on whether these emerging technological practices might (or might not) be able to be adopted as tools to create and gather knowledge about the city, what forms that knowledge might take, and how it might be interpreted.

A collaborative, creative methodology is proposed, with researcher working alongside arts practitioners in the creation of the trail. Data is gathered by asking participants to reflect on the experience of their performative 'mapping', and in doing so perhaps uncovering aspects of the sites they have negotiated.

¹ Allan Kaprow, 'The Legacy of Jackson Pollock', in *Essays on the Blurring of Art and Life*, ed. by Jeff Kelley (University of California Press, 2003), pp. 1–9 (p. 9).

Building Textal: What can Apps do for Digital Humanities?

Melissa Terras and Steven Gray, *University College London*

This paper reports on Textal, a text analysis application for iOS which is currently in development at UCLDH and UCLCASA. The app will be freely available from June 2012, and this paper will present findings evaluating the development, launch, and reception of the app, indicating how smartphone technology can increase the potential for public engagement within the Digital Humanities.



Textal (soon to be launched at www.textal.org, currently on twitter at @textal) will be a freely available smartphone application which allows users to create, share, and explore word clouds of a document, website, or tweet stream. Those in Digital Humanities have tended to sneer at word clouds, given we are used to applying robust text analysis tools, but Textal makes them interactive: once a word cloud of a chosen text is generated, users can click on words to access underlying statistics, such as frequency and collocates.

We envision Textal as a fun text-analysis-in-your-pocket app. We have built this with the general audience in mind, to bring Digital Humanities techniques to as wide an international audience as possible (we will be translating the interface into many languages). Textals can be shared via social media and there is a corresponding website where they are publicly available.

Although the app will not be launched until June this is not a promissory abstract: most of the development, save user testing, is now complete. When we launch we will be tracking use and users of the app, including geo-locating text analysis, to ascertain the potential audience for this type of app. This paper will be an up-to-the-minute account of a very recent development in Digital Humanities: we believe we are the first people to build a stand-alone app that brings DH

techniques to a wider audience. What ramifications do apps hold for DH as a discipline or a field of practice? We will report using up-to-date statistics generated from Textal as a case study.

Saturday 8th September, 11.30 – 13.00: Session 21