ROUTLEDGE

Cultural Heritage Infrastructures in Digital Humanities

Edited by Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes

Cultural Heritage Infrastructures in Digital Humanities

What are the leading tools and archives in digital cultural heritage? How can they be integrated into research infrastructures to better serve their intended audiences? In this book, authors from a wide range of countries, representing some of the best research projects in digital humanities related to cultural heritage, discuss their latest findings, both in terms of new tools and archives, and how they are used (or not used) by both specialists and by the general public.

Agiatis Benardou is a Senior Research Associate at the Digital Curation Unit, Athena Research Centre, Greece, and a Teaching Fellow in Digital Curation at Panteion University of Social and Political Sciences.

Erik Champion is UNESCO Chair of Cultural Heritage and Visualisation at Curtin University, a steering committee member of the Curtin Institute for Computation (CIC), and a member of the Australia Asia Pacific Institute (AAPI).

Costis Dallas is Director of the Collaborative Programs and Associate Professor at the Faculty of Information at the University of Toronto, Canada, and a Research Fellow at the Digital Curation Unit, IMIS, Athena Research Centre, Greece.

Lorna M. Hughes is Professor in Digital Humanities at the University of Glasgow.

Digital Research in the Arts and Humanities
Series Editors: Marilyn Deegan, Lorna M. Hughes, Andrew Prescott and Harold Short

Digital technologies are becoming increasingly important to arts and humanities research, expanding the horizons of research methods in all aspects of data capture, investigation, analysis, modelling, presentation and dissemination. This important series will cover a wide range of disciplines with each volume focusing on a particular area, identifying the ways in which technology impacts on specific subjects. The aim is to provide an authoritative refection of the 'state of the art' in the application of computing and technology. The series will be critical reading for experts in digital humanities and technology issues, and it will also be of wide interest to all scholars working in humanities and arts research.

Also in the series:

Digital Scholarly Editing
Elena Pierazzo

Copyrighting Creativity: Creative Values, Cultural Heritage Institutions and Systems of Intellectual Property
Edited by Helle Porsdam

Critical Gaming: Interactive History and Virtual Heritage
Erik Champion

Literary Mapping in the Digital Age
Edited by David Cooper, Christopher Donaldson and Patricia Murrieta-Flores

Historic Newspapers in the Digital Age: "Search All About It!"
Paul Gooding

Cultural Heritage Communities: Technologies and Challenges
Edited by Luigina Ciolfi, Areti Damala, Eva Hornecker, Monika Lechner and Laura Maye

Generative Systems Art: The Work of Ernest Edmonds
Francesca Franco

Cultural Heritage Infrastructures in Digital Humanities
Edited by Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes

www.routledge.com/history/series/DRAH

Cultural Heritage Infrastructures in Digital Humanities

Edited by Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN: 978-1-4724-4712-8 (hbk)
ISBN: 978-1-315-57527-8 (ebk)

Typeset in Baskerville
by Swales & Willis Ltd, Exeter, Devon, UK

This book is dedicated to those tireless, enthusiastic and all too often anonymous and invisible volunteers and researchers who have spent millions of hours recording, collating, collecting and preserving aspects of cultural heritage for the benefit of others.

Contents

Figures

Contributors

Alexandra Angeletaki is a classical archaeologist (MA from the University of Oslo) and now works as a Senior Research Librarian at the Norwegian University of Science and Technology in Trondheim, Norway. Since 2008 she has worked with digital educational design and immersive technology as project manager of Mubil and ARK4 at the NTNU University library. She worked as a field archaeologist for the KA Greek Ephorate of antiquities in Greece, from 1982 until 2003, including museum registration work in Delos, Paros, Samos, Santorini. She moved to Norway in 2000 and worked until 2006 as a lecturer for archaeology students at NTNU University in Trondheim, teaching courses on Classical Archaeology, European Archaeology, Archaeological Theory and Cultural History. Her projects at NTNU promote the historical archives of the university library and its digital resources.

Agiatis Benardou holds a Ph.D. in Ancient History and Classical Archaeology and is a Senior Researcher at the Digital Curation Unit, IMIS-Athena Research Centre in Athens, Greece. Agiatis has carried out extensive research on scholarly practices and user requirements in the context of various EU initiatives, such as Preparing DARIAH (www.dariah.eu) and EHRI (www.ehri-project.eu/). She was involved in the development of the NeMO ontology, in a team led by Panos Constantopoulos at the Digital Curation Unit in Athens (http://nemo.dcu.gr). Moreover, she is co-chairing the Community Building Working Group of DARIAH-EU and leading the user requirements work for the Erasmus+ Strategic Partnership for the Digital Humanities Reference Curriculum (http://dariah.eu/teach/index.php/2015/05/21/welcome-to-dariahteach/). Agiatis is leading the team responsible for assessing researchers' needs and ensuring community engagement for Europeana Research (http://research.europeana.eu/).

Tobias Blanke is a Reader in the Department of Digital Humanities at King's College London. From 2014 to 2016 he served as a Director of DARIAH-EU. Dr Blanke's academic background is in philosophy and computer science, and before joining King's he held various positions in

the digital industries. His particular research expertise is text mining and information retrieval on heterogeneous textual collections. He works on several international projects and committees and has published over 50 papers in peer reviewed journals and conference proceedings.

Marcello Carrozzino is Assistant Professor of Computer Graphics and Virtual Reality. His activities deal with the integration of Virtual Reality systems and Cultural Heritage applications. Since 2005, he has been teaching the course of Virtual Environments held at the Computer Science department of the University of Pisa. In 2006, he attained a Ph.D. at Scuola Superiore Sant'Anna in Pisa with a thesis on Efficient Management of Complex Virtual Environments. From 2006 to 2009 he has been Assistant Professor at IMT Lucca in the area of Technology and Management of Cultural Heritage. Since 2000, he has been collaborating with Scuola Sant'Anna, where he is currently heading the Art, Cultural Heritage and Education Group (see http://www.percro.org/people?field_tags_tid[]=54). He is a founder of VRMedia s.r.l, a spin-off company of Scuola Sant'Anna, and of Mnemosyne Digital Culture association. He has authored and co-authored over 70 publications in scientific journals, conference proceedings and book chapters.

Erik Champion (Ph.D. Melbourne) is UNESCO Chair of Cultural Heritage and Visualisation at Curtin University. A past ARC SPIRT Ph.D. scholarship holder, he has received Fulbright and Greece–NZ scholarships, a Distinguished Lecturer Invitation from UC Berkeley, funding from the Danish government and Apple, and facilitated or advised on major grants and awards in Europe and America from organizations such as the EU and COST, ERASMUS, European Digital Humanities, Mellon Foundation, Digital Heritage Centre funding (York, Leiden, Uppsala, Aarhus), and the National Endowment for the Humanities (NEH). He was a team co-leader of Research and Engagement for DARIAH (www.dariah.eu) and project leader of DIGHUMLAB Denmark, a five-year DKK 30 million national infrastructure project. He writes in the area of game design and virtual heritage; his books include *Playing with the Past* and *Critical Gaming: Interactive History and Virtual Heritage.*

Costis Dallas is Associate Professor and Director of the Collaborative Programs at the Faculty of Information, University of Toronto. He is also a founding Research Fellow of the Digital Curation Unit (DCU), IMIS-Athena Research Centre in Athens. His research as co-principal investigator in the CARARE, LoCloud, Europeana Cloud and ARIADNE projects has been on understanding knowledge practices and methods in the field of cultural heritage and humanities scholarship, on knowledge representation of material culture, and on the specification and use of curation-enabled digital heritage infrastructures. He leads the DARIAH-EU European scholarly practices survey project, and is a

member of the NeDiMAH Methods Ontology (NeMO) project team. His recent publications include 'Digital Curation beyond the "Wild Frontier": A Pragmatic Approach', and 'Curating Archaeological Knowledge in the Digital Continuum: From Practice to Infrastructure', and is currently working towards a monograph on the digital curation of thing cultures, shaped by increasingly distributed, pervasive and participatory information practices 'in the wild'. Dr Dallas is Chair of the DARIAH-EU Digital Methods and Practices Observatory Working Group (DiMPO), and Vice-Chair of the Archaeological Research Practices and Knowledge Work in the Digital Environment (ARKWORK) COST Action. He holds M.Phil. and D.Phil. degrees from the University of Oxford.

Alastair Dunning was coordinator of the Europeana Cloud project from 2013 to 2015, before taking up a role at the Technical University of Delft in the Research Data team. He has previously managed or overseen over 70 digital projects as part of the UK Digitisation Programme funded by the educational charity, Jisc. Prior to that, he worked at the Arts and Humanities Data Service.

Lorna M. Hughes is Professor of Digital Humanities at the University of Glasgow. Her research addresses the creation and use of digital cultural heritage for research, with a focus on collaborations between the humanities and scientific disciplines. A specialist in digital humanities methods, Hughes is the author of *Digitizing Collections: Strategic Issues for the Information Manager*, the editor of *Evaluating & Measuring the Value, Use and Impact of Digital Collections,* and the co-editor of *The Virtual Representation of the Past*. She was the Chair of the European Science Foundation (ESF) *Network for Digital Methods in the Arts and Humanities* (www.nedimah.eu) from 2011 to 2015, which developed the NeDiMAH Methods Ontology for the Digital Humanities (NeMO: nemo.dcu.gr/). Other notable digital projects include the AHRC-funded *The Snows of Yesteryear: Narrating Extreme Weather* (eira.llgc.org.uk) and the Jisc-funded digital archive, *The Welsh Experience of the First World War* (cymruww1.llgc.org.uk).

Christina Kamposiori has a Ph.D. in Digital Humanities; her thesis focused on the information practices of art historians in the digital age and was conducted at the UCL Centre for Digital Humanities. Currently, she is the Programme Officer at Research Libraries UK (RLUK). She holds a BA in Archaeology/Art History from the University of Ioannina in Greece and an MA in Cultural Heritage Management from Panteion University in Athens, Greece. She has previously worked as a Teaching Assistant at the UCL Faculty of Arts & Humanities and been a programme committee member of the AHRC funded project 'New Media, Audiences and Affective Experiences'. Before that, she worked as a Junior Researcher at the Digital Curation Unit-IMIS, Athena Research Centre in Athens, Greece, in the context of the project 'Preparing

DARIAH'. Her research interests include, but are not limited to, digital scholarship and infrastructures in the Arts and Humanities, information management, and the application of digital technologies in libraries, museums and the cultural heritage sector.

Gertraud Koch is Professor of Cultural Anthropology at University of Hamburg. Her main research areas are the culturality of information science and the empirical analysis of digital cultures. The latter also will be the topic of her edited book *Digitisation: Theories and Approaches for Empirical Cultural Analysis* (forthcoming, 2018).

Conny Kristel is a senior researcher at NIOD Institute for War, Holocaust and Genocide Studies, which is part of the Royal Netherlands Academy of Arts and Sciences (KNAW). From 2014 to 2016 she served as a Director of DARIAH-EU. Dr Kristel holds a Ph.D. in History from the University of Amsterdam. Before joining NIOD she worked among other things as a freelance journalist and writer, while preparing her dissertation. She also held several part-time positions in academic research (project) management. Her principal research interests cover the study of war and society. She is the coordinator of the DARIAH-affiliated EHRI project (www.ehri-project.eu).

Simon Mahony is Associate Director for Teaching at the UCL Centre for Digital Humanities and Principal Teaching Fellow at the Department of Information Studies, University College London, where he is Programme Director for the MA/MSc. in Digital Humanities. He has research interests in the application of new technologies to the study of the ancient world, using web-based mechanisms and digital resources to build and sustain learning communities, collaborative and innovative working, and in the development of education practice, and the use of new tools and technologies to facilitate this. He is also an Associate Fellow at the Institute of Classical Studies (School of Advanced Study, University of London) and one of the founding editors of the Digital Classicist.

Aileen O'Carroll is Policy Manager at the Digital Repository of Ireland. She is based at Maynooth University where she also manages the Irish Qualitative Data Archive (IQDA). As well as policy development, she advises researchers on best practice in managing and archiving research projects, both to ensure that ethical commitments are met and that the data gathered is of the highest standard to facilitate optimal reuse by a variety of audiences.

Julian D. Richards is a Professor of Archaeology at the University of York. He is Director of the Archaeology Data Service, the e-journal *Internet Archaeology*, and York's Centre for Digital Heritage. His direct involvement in archaeological computing began in 1980 when he studied pre-Christian Anglo-Saxon burial ritual using the computing power of an ICL mainframe and an early Z80 micro-computer for his Ph.D.

Apart from computer applications, his research interests focus on Anglo-Saxons and Vikings, and he has directed numerous excavations. He is author of *Viking Age England*, now in its third edition.

Laurent Romary is Directeur de Recherche at Inria, France, and Director General of DARIAH. He received a Ph.D. in computational linguistics in 1989 and his Habilitation in 1999. He carries out research on the modelling of semi-structured documents, with a specific emphasis on texts and linguistic resources. He has coordinated various EU projects, in particular MLIS/DHYDRO, IST/MIAMM and eContent/Lirics projects. He has been active in standardization activities with ISO, as chair of committee ISO/TC 37/SC 4 and the Text Encoding Initiative chairing the technical council in 2008–2011). He has been involved in the definition of the scientific information policy of CNRS (2005–2006), the Max-Planck Digital Library (2006–2008) and Inria.

Seamus Ross, Visiting Professor at School of Information Sciences and Technology, Athens University of Economics and Business (Athens), and Visiting Scientist at the Digital Curation Unit, IMIS-Athena Research Centre, during 2016 and 2017, is Professor in the Faculty of Information at the University of Toronto and HATII at the University of Glasgow. At Toronto he served as Dean (2009–2015) and subsequently Interim Director of the McLuhan Centre for Culture and Technology (2016). Before joining Toronto, he was Professor of Humanities Informatics and Digital Curation and Founding Director of HATII (Humanities Advanced Technology and Information Institute) at the University of Glasgow (1997–2009). He served as Associate Director of the Digital Curation Centre in the UK (2004–2009) and was Principal Director of ERPANET (2001–2004). Dr Ross's scholarly research has focused on digital preservation/curation, digital humanities, digitization, digital repositories, emulation, digital archaeology, semantic extraction and genre classification, cultural heritage informatics and knowledge representation and reasoning.

Veerle Vanden Daelen holds a Ph.D. (2006) from the University of Antwerp and has held fellowships at the University of Michigan (Frankel Institute for Advanced Judaic Studies) and at the University of Pennsylvania (Herbert D. Katz Center for Advanced Judaic Studies). Her Ph.D. work concerned the return and reconstruction of Jewish life in Antwerp after the Second World War (1944–1960). Her post-doctoral research has enlarged the scope of this research on Antwerp Jewish History. She currently coordinates the work package 'Data Identification and Integration' for the European Holocaust Research Infrastructure (EHRI) at CegeSoma. She is also affiliated to the University of Antwerp, where she has taught courses on Migration History, Jewish history and other topics. Since 2007, Karin Hofmeester and Veerle Vanden Daelen have organized the annual 'Contact Day Jewish studies on the Low Countries' at the Institute of Jewish Studies at the University of Antwerp.

Claire Warwick is Pro-Vice-Chancellor (Research) at Durham University. Previously, she was a Professor of Digital Humanities and a member of the Department of Information Studies at UCL. From 2009 to 2013, she was Vice-Dean for Research for the Faculty of Arts and Humanities, and Director, then Co-Director of the UCL Centre for Digital Humanities. Her research is concerned with the way that digital resources are used in the humanities and cultural heritage; in the use of social media in these areas; and in reading behaviour in physical and digital spaces. She has led or collaborated on several digital humanities research projects – for example, the INKE project and the QRator project. Her Ph.D. from Cambridge University was in English Literature.

Sharon Webb is a Digital Humanities Lecturer in the Sussex Humanities Lab and the School of History, Art History and Philosophy. Previously, Sharon was the Requirements Analyst for the Digital Repository of Ireland and specified the functional and non-functional requirements to develop DRI's trusted digital repository. She played a key role in informing policy related to the building of a trusted digital repository as well as developing guidelines for metadata, copyright and intellectual property, and business sustainability.

Preface

The genesis for this edited book was a two-day workshop entitled 'Cultural Heritage Creative Tools and Archives' (CHCTA), hosted at the National Museum of Denmark, Copenhagen, 26–27 June 2013. The workshop was organized by DIGHUMLAB Denmark, and the Digital Curation Unit (DCU), IMIS-Athena Research Centre.

There were approximately two-dozen presentations from around a dozen European research organizations and infrastructures (such as DARIAH, DASHISH, ARIADNE, CENDARI, MUBIL, Europeana Cloud, 3D-icons, the European Holocaust Research Infrastructure, Serious Games Interactive, The Digital Repository of Ireland, LARM and NeDiMAH), two invited international speakers (Professor Seamus Ross and Professor Julian Richards), and a final panel, which explored how research infrastructures dealing in digital cultural heritage could work more closely together. Various future projects were kick-started from this workshop, which also saw the launch of the Europeana Researchers Coordination Group (ERCG). Some of the presenters later organized a workshop at Digital Humanities 2014 conference in Lausanne.

The intention behind both the CHCTA2013 and DH2014 workshops was to foster a shared critical understanding of the current state of digital infrastructures and the potential of digital archives, tools and services to contribute to humanities scholarship, particularly in cultural heritage. Both successes and failures were discussed and debated, for the workshops hoped to reach widespread agreement between the various presenters on how and where we could better integrate infrastructure projects, develop a richer understanding of user needs and foster more collaboration among researchers. This book carries on these aims, but the journey is long and challenging.

I would like to thank The European Association for the Digital Humanities (www.allc.org/) for providing us with a small grant award, Lorna M. Hughes of NeDiMAH who provided travel funding for NeDiMAH members, the National Museum for hosting and refreshments, and DIGHUMLAB for additional funding. DARIAH and other infrastructures helped us with publicity. Costis Dallas provided much needed academic overview, and

helped with both hosting and with programme scheduling, while Agiatis Benardou was instrumental in bringing in Europeana and Europeana Cloud representatives.

Lastly, I would like to thank Lene Elsner and Marianne Ping Huang of DIGHUMLAB. Their support and organizational skills were invaluable.

Erik Champion

1 Introduction: a critique of digital practices and research infrastructures

Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes

Digital Humanities might appear a recent phenomenon. Yet almost seventy years have gone by since Father Roberto Busa initiated his Digital Humanities project: the computer-assisted lemmatization of the complete Thomistic corpus (www.corpusthomisticum.org/). Although Busa first conceived of this project in 1946, it took him nearly four decades to realize it, leveraging the power of the digital computer as an ordering machine capable of processing and listing potentially infinite amounts of textual data. The development of the first computational analysis of archaeological materials, a numerical classification of Eurasian Bronze axes conducted by Jean-Claude Gardin and Peter Ihm in the late 1950s (Cowgill, 1967; Huggett, 2013) introduced a different aspect of computer-based research: one that brought to the fore the possibilities afforded by digital methods for dimension reduction, discovery and visualization of latent structures of complex data.

Fast-forwarding to the present day, two surprisingly distinct communities have already emerged in digital arts and humanities research. On one hand, Digital Humanities, at least until very recently, appeared preoccupied with transforming the traditions of text-based humanities computing, drawn directly from library collections and scholarly practice. Digital Heritage, on the other hand, has drawn more from theories and practices in digital archaeology and the digital representation of material culture, but has often gained attention for its adoption of cutting-edge visualization and virtual reality technology. While driven by the traditions of custodian institutions such as museums, galleries, libraries, and archives and special collections, Digital Heritage leverages the capabilities of contemporary technologies in visualizing and representing cultural objects beyond text, and occasionally borrows ideas from the entertainment industry.

Digital Heritage might influence Digital Humanities in terms of lessons learnt from visualization, scanning/recording, 3D photorealistic modelling, GPS and mapping technologies, and possibly even instructional design and serious game development. But Digital Heritage could also learn from developments and strengths of Digital Humanities: community-based collaboration of scholars, virtual research environments, critical debates,

university-linked makerspaces, flipped classroom teaching, THATcamps and Digital Humanities Unconferences.

The contemporary landscape, mapping the use of digital resources, methods and tools for scholarly research, extends to most of the disciplines under the scope of the arts and humanities, including those disciplines relying on the fast-advancing capabilities of contemporary digital technology to represent cultural phenomena through increasingly accurate visual reproduction, audio, video and 3D photorealistic modelling. In addition, information and communication technologies are now routinely adopted for the more mundane aspects of research work – from information seeking and searching to note-keeping, bibliographic citation management, organizing personal research resources and, last but not least, preparing scholarly work for publication, by the overwhelming majority of researchers, far outside the core communities of the Digital Humanities and Digital Heritage.

A considerable body of work in the humanities is often differentiated from research in the natural sciences by its interest in the particular: a concrete work or corpus, a historical event or period, a culture, an artefact, or an artist, to name some examples. In this light, humanities research can, in many disciplines, be characterized as often being *idiographic,* aiming to capture an adequate account and provide understanding of a particular phenomenon, rather than *nomothetic,* aiming to produce generally applicable (and replicable) laws, or law-like generalizations (Dallas, 1999). It is also distinctive in the higher degree of subjectivity, and lower degree of repeatability and falsifiability of research findings. A related consideration, crucial to the construction of knowledge in humanities research, concerns the centrality of recorded information, exemplified in its reliance on the construction and study of homogeneous corpora (of texts, archival resources, visual representations, etc.) and a variety of other, often complex and heterogeneous collections of information objects representing the record of human experience and knowledge.

It is therefore no accident that a major application of digital technology in the arts and humanities has been in the construction of scholarly databases and digital collections of humanities resources. As early as the 1990s, the Perseus Digital Library used the early SGML version of the Text Encoding Initiative standard for the structuring and conversion of a canon of ancient Greek texts and their English translations. Together with a broad collection of digitized photographs of Greek art and architecture, architectural plans and drawings, and even animations, this project allowed cross-referencing and analysis of sources for the benefit of research and academic education (Crane, 1998). In contrast to the monolithic research database or digital collection, such as the digital processing of text-based corpora drawn from datasets (for example, the Thesaurus Linguae Graecae, TLG), Perseus prefigured a different approach: the integration between digital information resources, a diversity of analytical and visualization tools, and an active community of researchers sharing knowledge and co-developing

research practices. This integration between a community of researchers, digital tools and organized digital resources underlies the major developments in the last twenty years, leading to the current era which has seen the establishment of a number of successful Research Infrastructures across a variety of disciplines, several of which are described in this volume.

As noted by Erik Champion, '[i]n the case of Digital Humanities, what is missing is the notion of a scholarly eco-system' (Champion, 2014). Like Perseus, contemporary Research Infrastructures have the aspiration of being not merely collections of research resources or tools to conduct research: they are energized by a community of research institutions and individual researchers, and become living environments of evolving, synergistic but also often competing research, education and communication practices. It may be argued, therefore, that an ideal digital Research Infrastructure today should be conceived of primarily as a *scholarly ecosystem*: one that supports ongoing scholarly development and use of research resources, tools and methods, and the outputs they enable, through the application of digital technologies. Viewed as an ecosystem, a digital Research Infrastructure can thus be viewed as consisting of interdependent parts, which make up a whole that should be greater than the sum of its parts. Given the dynamic and evolving nature of research, such a digital research ecosystem should provide for the survival and evolutionary development of 'traditional research' in new and more effective ways, but also enable the conceptualisation of important new research questions and the birth of entirely new forms of research tools, methods and approaches.

As an ecosystem, a digital Research Infrastructure can only be effective if it addresses the abilities and needs – not to mention the lifecycle – of its diverse 'resident species' and the attributes of their environment. To engineer an infrastructure as a sustainable and effective ecosystem calls, therefore, for an understanding of the practices and needs of scholars, archivists, technical specialists as well as other end users of the knowledge production, reproduction and dissemination process. This enquiry goes beyond instrumentality: to take the example of research data creation and capture using digital means, it should not just allow us to know how data capture happens through digitization, but to contribute to a greater and more rigorous critical understanding of the whole process of digital source creation *and critical use* (van Peursen, 2010):

> The creation of digital objects – be it images of inscriptions or manuscripts, electronic versions of ancient corpora, or collections of secondary literature – is a crucial part of humanities research. It is more than just preparation for research. This is a fundamental difference between databases as they are used in the humanities and those that are used in the natural sciences. The way in which inscriptions are photographed or in which text corpora are transcribed and encoded, is crucial for the way in which these research objects will be studied in the future.

The creation, curation and use of digital objects in scholarly work hinges on the development and adoption of a wide range of digital methods at the intersection between research in the disciplines and the management of scholarly collections. Such methods span the full lifecycle of scholarly research, from the inception of a research idea to publication and knowledge translation. They include text analysis and mining, image analysis, moving image capture and analysis, and quantitative and qualitative data analysis. The centrality of research methods for scholarly practices is reflected, as early as 2005, in the work of the AHRC ICT Methods Network (http://methodsnetwork.ac.uk) on documenting the use of ICT methods in the UK, and the methodological initiative of the art-and-humanities.net project, based at King's College, London, to develop and disseminate knowledge on 'computational methods used by artists and humanists'. From 2011 to 2015, the European Science Foundation Network for Digital Methods in the Arts and Humanities (NeDIMAH.eu) expanded this work within the European context (and it was NeDIMAH that funded the 2013 workshop in Copenhagen that was the genesis for this volume). The major output of NeDIMAH was a resource documenting how the digital humanities research lifecycle can be represented as a process, showing dependencies and relationships, and showing how it can, in fact, provide a framework for the creation, enhancement and use of digital cultural heritage.

This need for an explicit model of the research process, capturing the interplay of all the important elements of the scholarly ecosystem is central in the process of defining the scope and affordances of scholarly infrastructures. Inspired by the influential notions of John Unsworth's 'scholarly primitives (Unsworth, 2000), and Willard McCarty and Harold Short's 'methodological commons' (McCarty, 2003), as well as on emerging work on scholarly information behaviour (Borgman, 2007; Palmer *et al.*, 2009), researchers at the Digital Curation Unit, Athena Research Centre, proposed a Scholarly Activity Research Model, grounded on empirical evidence for researcher practices and needs within DARIAH and EHRI, the European Holocaust Research Infrastructure project (Benardou *et al.*, 2010, 2013). The SRAM model, compliant with the CIDOC CRM ontology of cultural heritage (ISO standard 21127), was intended to support the elicitation of requirements, and the design and development of information repositories and services in digital humanities infrastructures. The confluence between this ontological approach to scholarly activity modelling and NeDIMAH's initiative of NeDIMAH to establish a formal framework for the conceptualization of research methods in the arts and humanities, led to the development of NeMO: the NeDiMAH Methods Ontology (http://nemo.dcu.gr).

NeMO was established as an ontology of digital humanities that formally documents the practice of digitally based scholarship as a sociotechnical knowledge activity, explicitly addressing the interplay of conceptual dimensions of agency (actors and goals), process (activities and methods) and resources (information resources, tools, concepts) in the scholarly process:

showing the dependencies of *content, tools* and *methods* (Hughes *et al.*, 2016; Pertsas and Constantopoulos, 2016). Researchers at Glasgow University and the Digital Curation Unit, Athena Research Centre in Athens, are currently using NeMO as a conceptual framework to describe the use of digital methods and content for research. NeMO is a tool for semantic linking in an environment of interoperable resources and services for discovering, understanding, selecting, linking and contributing content, tools and methods.

The development of the NeMO ontology incorporated existing research that had attempted to understand digital humanities projects, methods or tools by expressing them through taxonomies (e.g. Borek *et al.*, 2016): analysis of these indicated an ontology, intellectually and technically, was a missing piece of the digital humanities research infrastructure.[1] By providing a formal framework for critique and debate about the contexts and dependencies within the use of digital content for research, it facilitates much needed methodological and epistemological reflexivity within the digitally based humanities, and accommodates within an overarching conceptual framework the workings of digital infrastructures, tools and services in humanities research and digital heritage, and the processes and methods adopted by researchers, stewards and users of cultural heritage information resources.

The intimate interdependence between the affordances of Research Infrastructures and the scholarly methods and practices they enable is confirmed by work in the broader domain of practice studies (Schatzki *et al.*, 2001), and in the more focused area of infrastructure studies (Edwards *et al.*, 2009; Edwards *et al.*, 2007; Jackson, 2007).[2] In her work on the 'relational undergirding of epistemic practice', and the characterization of research as a creative and constructive 'objectual practice' relying on evolving sociocultural arrangements around knowledge objects, Karen Knorr-Cetina (2001) has demonstrated how infrastructures are not just the site of routinized enactments of established research protocols and methods, but a site of dynamic reinvention and change for research through the discovery and invention of new approaches to intellectual enquiry – in our terms, evolving ecosystems.

From the perspective of those creating digital archives and resources, this ecological approach to digital Research Infrastructures can form the basis for a theoretical reflection concerning the mode of production of scholarly knowledge in the arts and humanities. Developing maker spaces, drop-in data labs, open-sourced data and online review communities as part of a critical feedback process that informs and helps improve the role and function of Research Infrastructures could be vital components in the development of this scholarly ecosystem. However, there are still too few complete and coherent examples. Too many Research Infrastructures lack clear evidence of impact and engagement by the wider public, too many Research Infrastructures live and die based on short-term funding of technology rather than on meaningful usage, too many research groups are divided by institutional or national political or historical diversions that

prevent them from benefiting from the potential synergy of different backgrounds, interests, experiences, skills and data sets.

The development of digital Research Infrastructures in the last decade was in many respects anticipated by the publication of the Atkins Report on cyberinfrastructure for e-science (Atkins *et al.*, 2003), the 'cultural commonwealth' report of the American Council of Learned Societies (ACLS, 2006) and the European Strategy Forum on Research Infrastructures (ESFRI) roadmap (http://ec.europa.eu/research/infrastructures/pdf/esfristrategy_report_and_roadmap.pdf). Benefiting from European Commission funding, such infrastructures in Europe include CLARIN, the European Research Infrastructure for Language Resources and Technology (www.clarin.eu), DARIAH, the Digital Research Infrastructure for the Arts and Humanities (www.dariah.eu), both recommended in the context of the ESFRI roadmap, and also ARIADNE, the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (www.ariadne-infrastructure.eu). A further number of specialized Research Infrastructures are expected to emerge in the context of new rounds of infrastructural European funding made available in the foreseeable future.

The expectations of digital Research Infrastructures are high, in accordance with the funding they have received. Ideally, such infrastructures should address the complex nature of Digital Humanities data sets, research methods and collaborative work practices, offering humanities scholars new and productive ways to explore old questions and develop new ones, even addressing some of the 'grand challenges' in the humanities, linking data and researchers through support for digital research practice. They should also offer more specific opportunities for data-driven and quantitative humanities research. Finally, they must provide a platform to address institutional and social issues, such as strengthening higher education programmes, as well as the recognition of digital research and the implications of this for scholarship and research careers. To meet these expectations, Research Infrastructures may offer a range of complementary opportunities and challenges: researcher input and engagement; preservation and sustainability; the evaluation of digital research and its outputs; communities of practice; cultural and linguistic variety (transnational Research Infrastructures); and education and training.

These considerations on the emergence, traits and requirements of Digital Heritage Research Infrastructures, viewed as living ecosystems, set the stage for this volume. Inspired by a Digital Heritage workshop in Europe (Cultural Heritage, Creative Tools and Archives, Copenhagen, 2013), this collection of chapters is predominantly European-focused and discusses European Research Infrastructures, but the findings may be extrapolated to other countries and regions. The current range of initiatives in much of the digital cultural heritage research presented in this volume points to salient challenges and prospects for further work in shaping the future scholarly ecosystem. They underscore the importance

of ambitious, long-ranging ('blue skies') research on the affordances and specifications of digital infrastructures sustainable in the long term that will anticipate what scholars need in the future, working in collaboration with holding institutions (such as libraries, archives, museums and galleries) as well as the technical disciplines.

Since the late 2000s, European research institutions have benefited from European Union funding to develop transnational Research Infrastructures in different disciplines, the funding intended to increase the development and competitiveness of the European research space. Initial projects funded under a European Commission grant programme have subsequently been encouraged to form a particular form of transnational structure, a European Research Infrastructure Consortium (ERIC), to provide a variety of research services to researchers. As noted by Erik Champion in his chapter in this book, 'the ERIC status is reserved for state-of-the-art Research Infrastructures that will create unique opportunities to carry out advanced research, attract the best researchers from across the world and train highly qualified students and engineers' (Champion, 2014). It is possible to map these requirements to the sources and methods for digitally enabled cultural heritage research: content, tools, methods, technical infrastructures and communities of practice, both researchers and users, as well as collaborative and open publishing and dissemination routes. Underpinning these infrastructures are national and international networks of co-operation, which bear the promise of bringing together institutions and individuals from research in the disciplines, holders of research data and resources, and technology specialists involved in Research Infrastructure design and development. Much of this thinking has informed the development of Research Infrastructures in the sciences, where many sophisticated supporting infrastructures have been developed, benefiting from the relatively stronger funding environment for scientific and technological research. Infrastructure initiatives in the arts and humanities are fewer and funded at a smaller scale, but nonetheless still significant.

While Research Infrastructures in the sciences often take the form of highly specialized physical laboratories and sophisticated experimentation set-ups, the reliance of humanities research on tangible resources that can be made accessible through information technology, as well as the erosion of disciplinary boundaries and the growing importance of public discourse and feedback, privileges a different kind of Research Infrastructure for the arts and humanities, centring on digital services that are built around communities of epistemic practice characterized by shifting research agendas and diverse theoretical and methodological orientations. The challenge is brought in focus by comparing CLARIN, addressing primarily the needs of literary and linguistic computational research, with DARIAH-EU, addressing the far more complex and less focused field of 'the digital arts and humanities' – a vast and moving target. Both infrastructures seek to address complex issues of governance associated with scholarly practice in a digital

age, such as policies associated with data ownership and preservation, ownership and licensing of tools and services, IPR issues. But DARIAH-EU, presented in this volume by Tobias Blanke, Conny Kristel and Laurent Romary in their capacity as its directors, rather than being a centralized infrastructure, acts more as an effective umbrella organization for successful national infrastructure initiatives (like DARIAH-DE in Germany), providing broad support at a highly strategic data management level, as well as the ability to facilitate the sharing of data and tools through partner organizations and affiliated projects. One identifiable result of DARIAH-EU to date has been the initiatives of its Virtual Competence Centre on Research and Education (VCC2) on researching the information practices and digital needs of European humanities researchers, on mapping the landscape of courses and learning resources on Digital Humanities available in Europe, on examining the applicability and preferred characteristics of Virtual Research Environments (VREs) for humanities research, and on mobilizing an active community of interest including researchers from the arts and humanities, as well as information scientists and computer scientists active in the design and development of digital infrastructures.

In Chapter 4, Blanke, Kristel and Romary argue that DARIAH-EU 'focus[es] on Research Infrastructures rather than (digital) library and archive integration projects such as Europeana, because Research Infrastructures share the ultimate aim to action research. Europeana on the other hand aims to primarily fulfil the needs of a culturally interested public rather than a research community'. Yet, in the work carried out under the auspices of the Europeana Cloud project, described by Benardou and Dunning in their chapter, we find a clear focus on understanding the research needs of users of digital content, and the tools that can support humanities research, that goes beyond serving just the users of cultural heritage content. This exploration of 'deeper engagement' with primary sources in digital format and the tools for their analysis is now the focus of the Europeana Research initiative, which draws also on research such as that conducted by Christina Kamposiori, Simon Mahony and Claire Warwick, who, in their chapter, analyse the transformation of scholarly practices in a specific discipline (in this case study, art history) afforded by increased access to digital resources, specifically examining how scholars approach, create and manage information. As the digital resources and related tools and methods for using these resources expand, there is an ongoing need for this development to be informed by the needs of researchers, if the digital turn is truly to effect transformative research in the arts and humanities.

Another benefit of digital Research Infrastructures is that they can become the hub for nurturing an interdisciplinary community of researchers working on focused and discrete research themes or topics. Based on the presentation of one of these initiatives, the European Holocaust Research Infrastructure (EHRI) project, Veerle van der Doelen (this volume) presents an intriguing view of how user requirements may ensure that the data

collected by large infrastructure projects remain appropriate to the needs of its target audience of researchers spanning different disciplines in the humanities and the social sciences, as well as of important communities beyond professional researchers.

Similarly, in their account of the Digital Repository of Ireland's research into tools for Digital Heritage in Ireland, Sharon Webb and Aileen O'Carroll (this volume) reveal that an understanding of the needs of researchers can also be shaped by considerations of what services and resources can be provided at the national level for preserving, curating and sustaining digital cultural heritage. Webb and O'Carroll raise a very pertinent question that underpins so many discussions of sustainability of digital collections and research outputs: 'Why save a million objects if users cannot usefully engage with those objects?' The reliance of the future use of Digital Heritage on concrete measures ensuring its sustainability has been documented by a number of research projects[3] and, in this process, a pertinent question emerges: what is the degree to which the developers of Digital Heritage should encourage their reuse for new and unforeseen purposes in order to justify the investment in the creation of resources?

In their chapter, Alexandra Angeletaki and Marcello Carrozzino also address this issue as they explore how libraries can improve the integration of digital technologies with their archive material to promote better engagement with their audiences. This is an important consideration as we move beyond the idea of digitization being driven solely by the prerogative of information access: users increasingly demand enriched access to heritage, and greater engagement with sources, rather than just accessing digital resources as passive consumers of information. Introducing a 3D space for reading and studying in the Norwegian University Library of Trondheim, the authors further demonstrate the importance of a user-centred approach to this kind of innovation.

Wider issues of digital anthropology and ethnography are addressed by Gertraud Koch (Chapter 5, this volume), who looks at one important articulation between Digital Humanities and cultural anthropology through the notion of the 'ethnography of infrastructures'. Koch raises an important question: why has Digital Humanities turned to information technology to solve questions of use, and what might be the use of tools and methods developed over decades of research and practice in cultural anthropology in the field of Digital Humanities?

An ongoing concern in this context remains the need to develop enhanced, open publication models to communicate research in the arts and humanities, and the field of cultural heritage, to the widest possible audience. In the light of the fact that many digital publications do not go beyond replicating the culture of print, Julian Richards's chapter on the challenges and opportunities for a much more enriched understanding of online publishing is especially timely. His insights originate from his experience of publishing *Internet Archaeology*, a pioneering journal bringing

together scholarly articles with interactively accessible data publication. The move towards open access, not just for research outcomes, but also for research data, is now a requirement by many funding agencies, so an exemplar of good practice from archaeology, a discipline that relies on a rich and comprehensive variety of complex digital sources, is particularly valuable. While Richards largely focuses on the challenges associated with developing open access approaches, there are wider issues associated with publishing innovative research online, specifically the ability to integrate interactive and experimental approaches to working with data. There is a great deal of expertise in open and innovative publishing within Digital Humanities, and more creative and sustained knowledge transfer between Digital Humanities and publishing is required. Initiatives like *Internet Archaeology* are excellent exemplars for this debate.

The volume is concluded by Seamus Ross's reflection on the future of digital infrastructures for humanities research and cultural heritage at a time of huge intellectual, technological and sociocultural challenges. Grounded on a historical account of advances and setbacks in the digitization of information resources and scholarly communication, and drawing from insights in the domain of digital preservation and curation, digital humanities scholarship and publishing, Ross advances the view that the future of digital infrastructures for the digital humanities and digital heritage lies in coordinated work on several interconnected areas, including advocacy, understanding of the needs and mobilization of research and cultural heritage communities, and 'intelligence at the level of the digital object'. He recognizes the huge new challenges faced by cultural heritage institutions at a time of increased commodification of cultural information, and increased risks to ensure the integrity and authenticity of cultural objects, and argues for alternative, post-custodial approaches to both preservation and access, possibly leveraging new mechanisms of ensuring a 'web of trust' such as blockchain technology. In his view, collaboration in underlying mechanisms for scholarly resource curation, access and publication are inextricably linked to forging a common vision that unites the fields of humanities research and cultural heritage.

This diverse collection of chapters introduces perspectives on a number of initiatives (many funded by the European Commission) that have developed resources, tools, services and methods for digital research engagement with cultural heritage content. Authors span a diverse community of stakeholders in digital infrastructures in the arts and humanities that ranges from Digital Humanities and digital archaeology scholars to information scientists studying scholarly work, museum studies researchers engaged with questions of learning and engagement based on cultural heritage resources, archivists and data managers tasked with the curation of databases and collections of cultural materials useful for scholarly research, and computer scientists involved in the specification, design and development of digital infrastructures. It represents a snapshot of emerging practice around

sharing and using resources useful for cultural research, manifested within officially sanctioned Research Infrastructures under custodial control, but also 'in the wild' by researchers and data curators employing the capabilities of pervasive networked digital technologies (Dallas, 2015, 2016), and exemplifying practices of collaboration and innovation that push the boundaries of what can be achieved in the digital mediation of heritage.

If significant investment in Research Infrastructures is to have value for scholarship, there needs to be a clear role for scholars to contribute to a greater scholarly investigation and critique of the digital content lifecycle, and, more generally, to nurture a deeper theoretical reflection concerning the role of the digital in humanities research. Reviewing elements of a research ecosystem mapped in the chapters presented in this volume is an important opportunity for a praxis-based critical engagement, the key to understanding how digital is actually affecting knowledge production. Lessons drawn from research presented in these chapters will inform the future development of Research Infrastructures with affordances that address more closely the practices and needs of humanities scholars, drawing us into new collaborations, leading us to encounter new methods for engaging with content, and assisting in developing new insights into cultural heritage as a field of intellectual inquiry and social engagement.

Alan Liu (2012) and others have explored Digital Humanities as a vehicle of cultural critique (e.g. Berry, 2012; Gold, 2012; Terras *et al.*, 2013), but this collection of chapters aims also to extend critique to the role and effectiveness of Research Infrastructure *in practice*. By looking at the insights of those who have been involved in actually crafting digital tools and infrastructures, we can see that maker perspectives and theorist perspectives are not incompatible. For it is in the developing and building of digital projects in the humanities that we can conduct cultural and critical analysis more effectively. It is through questioning many of the assumptions on which digital resources are built and communicated that we can develop a better framework for understanding how working with digital content and digital infrastructure transforms our practices of production and consumption of knowledge. By building and using digital collections for research, we can develop a better understanding of their role in the humanities research lifecycle, and start to address questions about how digital content is not just helping us to do research more effectively, but can act as a disruptive, transformative intervention that unsettles epistemic paradigms and allows the emergence of new kinds of intellectual enquiry.

Notes

1 This conclusion was reached through an assessment of the complexity of the multidisciplinary landscape of digital research in the humanities, involving a combination of digital content, tools, and methods and research practices from a range of disciplines and traditions: making practice seem fragmented and hard to define. Recent debates about the nature of Digital Humanities exemplify how this

lack of transparency inhibits a shared understanding of digital research methods, their contexts, dependencies and affordances, and prevents effective peer review of digitally enabled research outside one's 'home' discipline. Similarly, the role of cultural heritage organizations and collections can be opaque: by expressing the dependencies within the 'methods/tools/content' triad, NeMo provides a 'layer' that allows arts and humanities researchers to express the work they have done to develop, refine and share digital research.

2 As argued by Geoff Bowker in his study of corporate information infrastructures for oil- drilling research at Schlumberger (Bowker, 1994), the institutional arrangements and affordances of an information infrastructure have a huge effect on research practice, in what he calls 'infrastructural inversion' – examining the mundane workings of infrastructure becomes therefore central to understanding practice. Conversely, his study on the development and use of the International Classification of Diseases – a major component in the information infrastructure of medical research – demonstrates how the infrastructure itself is shaped by the contingency of practice and the sociocultural norms and pragmatic implications of medical practice (Bowker, 1998).

3 See, for example, Hughes, L. M. (2014) 'Live and Kicking: The Impact and Sustainability of Digital Collections in the Humanities', in Proceedings of the Digital Humanities Congress, 2012, eds Mills, C., Pidd, M. and Ward, E. Special edition of Studies in the Digital Humanities. Sheffield: HRI Online Publications; and Hughes, L.M. (ed.) (2008), The AHRC ICT Methods Network. Office for Humanities Communication, London.

References

ACLS. 2006. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. American Council of Learned Societies. Available at: www3.isrl.illinois.edu/~unsworth/sdl.html.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P. and Wright, M. H. 2003. Revolutionizing Science and Engineering through Cyberinfrastructure. *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*.

Benardou, A., Constantopoulos, P. and Dallas, C. 2013. An approach to analyzing working practices of research communities in the humanities. *International Journal of Humanities and Arts Computing* 7(1–2): 105–27. doi:10.3366/ijhac.2013.0084.

Benardou, A., Constantopoulos, P. and Dallas, C. and Gavrilis, D. 2010. 'A conceptual model for scholarly research activity. In *iConference 2010: The Fifth Annual iConference*, John Unsworth, Howard Rosenbaum and Karen E. Fisher (eds), 26–32. Urbana-Champaign, IL: University of Illinois. Available at: http://nora.lis.uiuc.edu/images/iConferences/2010papers_Allen-Ortiz.pdf.

Berry, D. M., ed., 2012. *Understanding Digital Humanities*. Basingstoke: Palgrave Macmillan.

Borek, L., Dombrowski, Q., Perkins, J., Schöch, C. 2016. TaDiRAH: A Case Study in Pragmatic Classification. *Digital Humanities Quarterly* 10(1). Available at: www.digitalhumanities.org/dhq/vol/10/1/000235.html

Borgman, C. L. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA; London: MIT Press.

Bowker, G. C. 1994. *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920–1940*. Cambridge, MA: MIT Press.

Bowker, G. C. 1998. The History of Information Infrastructures: The Case of the International Classification of Diseases. *Historical Studies in Information Science*, 81.

Champion, E. 2014. Researchers as Infrastructure. In *Proceedings of the Digital Humanities Congress 2012: Studies in the Digital Humanities*. Sheffield. HRI online publications. Available at: www.academia.edu/6414838/Researchers_as_Infrastructure.

Cowgill, G. L. 1967. Computer Applications in Archaeology. *Computers and the Humanities*, 2(1), 17–23. doi:10.1007/BF02402460.

Crane, G. 1998. The Perseus Project and Beyond: How Building a Digital Library Challenges the Humanities and Technology. *D-Lib Magazine* 1: 18. Available at: http://dlib.org/dlib/january98/01crane.html.

Dallas, C. 1999. Humanistic Research, Information Resources and Electronic Communication. In *Electronic Communication and Research in Europe*, eds Jack Meadows and Heinz-Dieter Boecker, 209–39. Luxembourg: European Commission.

Dallas, C. 2015. Curating Archaeological Knowledge in the Digital Continuum: From Practice to Infrastructure. *Open Archaeology* 1(1): 176–207.

Dallas, C. 2016. "Digital Curation beyond the 'Wild Frontier': A Pragmatic Approach." Archival Science 16 (4): 421–57. doi:10.1007/s10502-015-9252-6.

Edwards, P. N., Bowker, G. C., Jackson, S. J. and Williams, R. 2009. Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems*, 10(5): 6.

Edwards, P. N., Jackson, S. J., Bowker, G. C. and Knobel, C. P. 2007. Understanding Infrastructure: Dynamics, Tensions, and Design. Report of the workshop, *History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures* Ann Arbor, MI: Deep Blue. Available at: http://hdl.handle.net/2027.42/49353.Gold, Matthew K. 2012. *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press.

Huggett, J. 2013. Disciplinary issues: Challenging the research and practice of computer applications in archaeology. In *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications* and Quantitative Methods in Archaeology (CAA), Southampton, 26–29 March 2012, edited by Graeme Earl, Tim Sly, Angeliki Chrysanthi, Patricia Murrieta-Flores, Constantinos Papadopoulos, Iza Romanowska, and David Wheatley, 13–24. Amsterdam: Amsterdam University Press.

Hughes, L., Constantopoulos, P. and Dallas, C. 2016. Digital Methods in the Humanities: Understanding and Describing Their Use across the Disciplines. In *A New Companion to Digital Humanities*, eds Susan Schreibman, Ray Siemens and John M. Unsworth, 150–70. Chichester: Wiley-Blackwell.

Jackson, S. J., Edwards, P. N., Bowker, G. C. and Knobel, C. P. 2007. Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy. *First Monday* 12(6). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/1904/1786.

Knorr-Cetina, K. 2001. Objectual Practice. In *The Practice Turn in Contemporary Theory*, eds T. R. Schatzki, Karin Knorr-Cetina and E. Von Savigny, 184–97. London: Routledge.

Liu, A. 2012. The State of the Digital Humanities: A Report and a Critique. *Arts and Humanities in Higher Education*, 11(1–2): 8–41. doi:10.1177/1474022211427364.

McCarty, W. 2003. Humanities Computing. In *Encyclopedia of Library and Information Science,* 1224–35. Basingstoke: Palgrave Macmillan.

Palmer, C. L., Teffeau, L. C. and Pirmann, C. M. 2009. Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development. Report commissioned by *OCLC Research*. Available at: www.oclc.org/programs/publications/reports/2009-02.pdf.

Pertsas, V., and Panos Constantopoulos. 2016. Scholarly Ontology: Modelling Scholarly Practices. *International Journal on Digital Libraries*, 1–18. doi:10.1007/s00799-016-0169-3.

Schatzki, T. R., Knorr-Cetina, K. and von Savigny, E. eds, 2001. *The Practice Turn in Contemporary Theory*. London: Routledge.

Terras, M., Nyhan, J. and Vanhoutte, E. eds, 2013. *Defining Digital Humanities: A Reader*. Farnham: Ashgate Publishing.

Unsworth, J. 2000. Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This? In *Humanities Computing: Formal Methods, Experimental Practice Symposium, King's College London*. King's College London. Available at: www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html.

Van Peursen, W. 2010. Text Comparison and Digital Creativity: An Introduction. In *Text Comparison and Digital Creativity,* eds W. Van Peursen, E. Thoutenhoofd and A. Weel. Brill: Leiden, pp. 1–28.

2 The role of 3D models in virtual heritage infrastructures

Erik Champion

Introduction

The success of virtual heritage projects, through the careful inspection, contextualization and modification of 3D digital heritage models with virtual reality technology, is still problematic. Models are hard to find, impossible to download and edit, in unusual, unwieldy or obsolete formats. Many of the freely available models are stand-alone 3D meshes with no accompanying metadata or information on how the acquisition of the data. Few have information on if or how the models can be shared (and if they are editable). Fewer still quantify the accuracy of the scanning or modelling process, or make available the scholarly documents, field reports, photographs and site plans that allowed the designers to extract enough information for their models.

Where there are suitable models in standard formats that are available from repositories, such as in Europeana library portal, they are likely to be in unwieldy 3D Formats. For example, 3D models encased in the proprietary PDF format cannot be extended, altered or otherwise removed from the PDF. Part of the problem has been with the development of virtual heritage; part of the problem has been due to a lack of necessary infrastructure. In this chapter, I will suggest another way of looking at virtual heritage, and I will promote the concept of a scholarly ecosystem for virtual heritage where both the media assets involved and the communities (of scholars, shareholders and the general public) are all active participants in the development of digital heritage that is a part of *living* heritage.

Virtual heritage

Nearly two decades ago, experts defined virtual heritage as a fusion of virtual reality technology with cultural heritage content (Addison, 2000; Addison *et al.*, 2006; Roussou, 2002). Stone and Ojika (2000) defined virtual heritage as:

> The use of computer-based interactive technologies to record, preserve, or recreate artifacts, sites and actors of historic, artistic, religious, and cultural significance and to deliver the results openly to a global audience in such a way as to provide formative educational experiences through electronic manipulations of time and space.

Many of the most famous and ambitious virtual heritage projects are digital simulations of UNESCO world heritage sites, yet the definitions do not explicitly address UNESCO's definition of intangible heritage. UNESCO defined intangible heritage as: 'practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their cultural heritage' (UNESCO, 2003). I suggest extending the definition to virtual heritage, to address the concerns of intangible heritage, because if we do not create simulations of our understanding as to how artefacts and sites (buildings and landscapes) were used, treasured and cared for, the public may not appreciate the cultural significance of the heritage site or object.

Virtual Heritage Environments (VHEs), in particular, can display content in new and more experientially appropriate ways, bridging disparate collections and remote sites with the intangible heritage of the original shareholders (such as oral history, mythology, and other cultural beliefs and traditions) beyond the physical constraints of the real world. Ideally, VHEs help the public to:

- Create, share and discuss hypothetical or counterfactual places.
- Meet virtually in these places with colleagues to discuss them.
- Contextually understand limitations forced on the original inhabitants of the simulated environment.

The technology also helps content experts and scholars develop experiential ways to entice a new and extended audience to both admire the content and the methods of their area of research, while providing them with feedback mechanisms and community input that does not require physical visitation.

Despite recent technological advances, simulating intangible heritage with digital media is not trivial. Creating a visual and photo-realistic simulation can be misleading; people did not live in photographs, they *inhabited* space. Hence, I take a wider and more inclusive view of visualization than the visual, even if important virtual heritage charters like the *London Charter* (Denard, 2009) define computer-based visualization as '[t]he process of representing information visually with the aid of computer technologies'. Virtual heritage is *not* simply the re-creation of what used to be there, a collection of objects. Cultural objects had specific and situated meaning in terms of the cultural perceptions of the land's traditional inhabitants, so digital reproducing how they may have looked as pristine artefacts is not enough – it only conveys what was left, not what was used and why it was valued. We must also convey the specific and situated importance of that cultural heritage to the public because UNESCO World Heritage awards heritage status for objects and sites of unique value.

In a book chapter that I wrote in 2008, I suggested that virtual heritage is 'the attempt to convey not just the appearance but also the meaning

and significance of cultural artefacts and the associated social agency that designed and used them, through the use of interactive and immersive digital media'. I defined new media (Champion, 2008) as 'the act of reshaping the user experience through the innovative use of digital media'. New media is thus a part of but not directly synonymous to digital heritage, for new media is constantly changing, so the way it is used and appreciated by the end user is also constantly changing.

If we are examining the end-user experience of virtual heritage and, in particular, how the end user is affected by changing experiences predicated by evolving technology, we could call this phenomenon new heritage. If new heritage is the application of new media to digital heritage, then by extension, the purpose of new heritage is to 'examine the user experience that digital media can provide for the understanding and experiencing of tangible and intangible cultural heritage'.

However, it is not just the user experience of virtual heritage that requires constant review; the development of virtual heritage as an academic discipline and particularly as scholarly infrastructure is seldom discussed – the aims and objectives are assumed to be shared and understood. In an article I co-authored with Laia Tost (2011), we identified six aims for the scholarly development of virtual heritage:

> Firstly, it should aim to carefully capture objects and processes of scientific, social or spiritual value. Secondly, it should present this information as accurately, authentically, and engagingly as possible. Thirdly, it should attempt to distribute the project in a sensitive, safe and durable manner to as wide and long-term an audience as possible. Fourthly, it should aim to provide an effective and inspirational learning environment that best communicates the intended pedagogical aims. Fifthly, it should allow the possibility to participate in its construction. Finally, it should attempt to carefully evaluate its effectiveness with regard to the above five aims in order to improve both the project and virtual heritage in general.

The above aims are more pedagogically focused than the earlier definitions, because there is no educational impact gained from records and collections hidden from the public. Yet galleries, libraries, archive and museums (GLAM industries) can only display a fraction of the collections that they own and have access to. Many museums lack the space to display the majority of their collection (Bradley, 2015) and there are a myriad of other problems to be faced when preserving cultural heritage in physical museums (Baio, 2015; Barsanti *et al.*, 2014; CSIRO, 2014; Lepore, 2015; Michaelis *et al.*, 2012; Widdowson, 2014).

There are logistic challenges with physical collections, but also conceptual challenges in how heritage collections can be maintained, disseminated, improved upon and expanded in the face of changing markets, budgets

and technology. Digital technologies hold obvious promise for expanding the public dissemination of knowledge associated with these collections even when physical access of the artefacts is restricted or their physical constitution is fragile, but the potential of digital media is limited when designers do not maximize the pedagogical impact of virtual heritage projects by selecting appropriate interaction. Digital models become meaningful simulations when they convey the culturally contextual ways in which they may have been used by past and distant cultures.

Why is infrastructure required for virtual heritage?

Previous definitions of virtual heritage have emphasized the criterion of preservation, but I suggest (with reluctance) that virtual heritage has been focused more on communication than on preservation; it showcases new uses and potential of technology for cultural heritage, yet the funding models and composition of project teams have had minimal usability evaluations and preservation strategies. Unfortunately, while virtual heritage has had muted success in showing how digital technology can provide insight into past cultures, as *digital heritage* it has been remarkably unsuccessful at saving its own showcase projects. Lost or inaccessible examples include *Rome Reborn* (Dylla *et al.*, 2008), *Beyond Space and Time* (IBM, 2008), or the 1996–1998 *SGI Teotihuacan VRML 2.0* model, (partially online but with missing links at www.kith.org/logos/things/VRML/handbook/).

Hal Thwaites (2013) is even more damning in his assessment of the current situation:

> In the very near future some critical issues will need to be addressed; increased accessibility to (and sharing of) heritage data, consistent interface design for widespread public use and re-presentations of work, the formalization of a digital heritage database, establishment of a global infrastructure, institutionalized, archival standards for digital heritage and most importantly the on-going curation, of work forward in time as the technology evolves so that our current digital, heritage projects will not be lost to future generations. We cannot afford to have our digital heritage disappearing faster than the real heritage or the sites it seeks to 'preserve' otherwise all of our technological advances, creative interpretations, visualizations and efforts will have been in vain.

There is a global imperative to collate and store digital heritage models of heritage sites (Reinhard, 2014; Thwaites, 2013). We also lack a way to provide access to the models, sites and paradata (which the London Charter (Denard, 2009) defines as 'Information about human processes of understanding and interpretation of data objects'). Despite initiatives such as the *London Charter* (Denard, 2009) and the *Seville Charter* (Lopez-Menchero

and Grande, 2011) as there are few publicly accessible models (Barsanti *et al.*, 2014) shared standardized evaluation data are hard to find. Scholars have complained about user experience issues and a scarcity of suitable pedagogical material (Economou and Pujol, 2008). There is also a myriad of practical and technical problems, such as how we are to gauge the accuracy of the recording and modelling process from a single 3D mesh, or how we are to judge the relative authenticity of the simulated material (De Reu *et al.*, 2012; Pitzalis *et al.*, 2010).

Discussions of virtual heritage models are found in journals such as *Journal of Computing and Cultural Heritage,* but actual virtual heritage models are much more difficult to find. There are very few online and library-accessible depositories for virtual heritage models, and many of the academic research projects lack long-term infrastructure and preservation strategies. Yet infrastructure is critical if we are to sustain scholarly communication, enrich public involvement and consolidate the currently promised – rather than proven – *heritage* component of *virtual heritage.*

If we are serious in helping the public to understand and participate in virtual heritage, then the public needs to understand the potential and limitations of the technologies as well. Workshops on 3D tools and software are required, which will allow communities, heritage groups and classrooms to learn from developing their own models and artefacts using free and open source game engines and 3D modelling tools.

Desirable features of virtual heritage projects

3D models

Virtual heritage projects are typically composed of 3D models. Key features of the models should be that they engage the audience, are formative (allowing the audience to create test and share hypotheses), can be recycled and reconfigured, and are amenable to preservation. It might seem opportune then to digitalize everything one can afford but the reliability of 3D data for long-term preservation is an ongoing issue: 'The possibility exists for precious and costly data sets to be lost on failed hard-drives, destroyed in floods or fires, or simply thrown out' (Greenop and Barton, 2014).

Formats are another issue (http://pvw.illinois.edu/pvw/) in the field of computer games. Researchers have even gone so far as to preserve the entire original game by wrapping it as an executable inside another programme (Carroll, 2012). Sven Havemann (2012) went so far as to complain:

> The file format problem is maybe today the most annoying obstacle for a further spread of 3D technology. Most of the aforementioned sophisticated shape representations can simply not be stored due to a lack of a common file format. The menace of file format degradation makes sustainable 3D only an illusion.

I find it difficult to disagree. A serious technical obstacle is the absence of a shared, secure, feature-rich format for 3D models (Koller *et al.*, 2009). Although there are at least 140 file formats for 3D models (McHenry and Bajcsy, 2008) almost all have major issues in either access, reliability, longevity or range of features (Koller *et al.*, 2009). Most research projects and publications that examine the usefulness of 3D file formats for virtual heritage appear to focus on .obj, .3ds, .u3d (which allows a 3D model to be embedded inside the .pdf file format), .o3d (a less well-known Google format which is now Open Source), .x3d (the successor to VRML), or Collada's .dae format (originally a file format designed for easy transfer between different modelling applications). Luckily, there are powerful and free file format converters like MeshLab (http://meshlab.sourceforge.net/).

Which 3D format holds the most promise for virtual heritage projects? Choosing a format that is robust, durable, well supported, free, highly interactive, cross-platform, and easy to create or export to or export from is a serious challenge. Major 3D formats such as .unity, .dae, .3ds, .obj, .blend and .x3d all have their advantages and disadvantages, but it is risky to offer only one format. At the time of writing, I suggest that the most promising file formats for archiving of 3D file formats and for simple web-based viewing of the 3D models would be obj, .x3d, or .dae. However, we need to distinguish between a format to store models in an archive and a format that allows people to immerse themselves in an online browser-based virtual environment. For example, .x3d (related to and a successor to VRML) offers a stable environment, is truly cross-platform, works well on the Internet and is free, but its functionality is limited and there are still relatively few exemplars and showcases.

To make interactive 3D models available via the Internet, various commercial and open source game engines have a range of features, 3D model libraries, examples and shortcuts to avoid extensive programming. Major common game engines that feature accessible editing and mudding for communities include Unity, Blender, CryEngine and Unreal. Most of these game engines can run as stand-alone applications, as web plugins, and across a range of devices (desktop and surround computers, specialized stereoscopic surround display screens, smartphones and game consoles).

Another relatively recent option is WebGL, a JavaScript Application Programming Interface (API) that allows 3D interactive graphics (and 2D graphics) to work inside any major web browser requiring a plugin, three.js. The plugin will load .obj models into WebGL (http://threejs.org/examples/webgl_loader_obj.html) without requiring advanced programming (and there are methods to export to Collada format). X3d models can run natively in HTML pages and Blender models will export directly to WebGL using Blend4web (www.blend4web.com/).

Arguably, the most popular interactive game engine for virtual heritage is Unity. It is both relatively stable and very powerful and flexible, but the Pro version is considerably more expensive. Unity formats might not

support such formats in the future. Other commercial game engine editors like CryEngine and Unreal 4-UDK have free versions but they also have cost/profit requirements based on the revenue from games sold, and they are under no obligation to ensure that older versions are still being maintained. Therefore, a proprietary 3D format is best avoided. If a game engine (a real-time rendering engine) is required, then a solution would be to have the game engine or application add the components (assets) dynamically, requiring the model to be broken up into subcomponents and then the computer would stream and connect to these subcomponents (packages) at runtime. There may also be a compromise solution that allows both a robust but limited 3D format for archived models and a more interactive format available either via a browser or as a downloadable application. An example of such a solution would be to archive models in, e.g., .x3d, but also provide an online converter or reformatter that can export from .x3d into, e.g., blender files (.blend) or text files that can be read by game engines.

Metadata

A complementary issue to that of selecting appropriate formats and solutions for 3D models is how will we even find these models? We require metadata in the 3D models so we can find and classify them, an ontology of model components so we can find and label individual parts, a storage and retrieval system for the 3D models and a way of linking the models with external assets (other media assets as well as publications and papers).

Metadata is essential for virtual heritage to establish itself as a long-term research area, but metadata has to help the objectives of virtual heritage, which are arguably as much or more about education as they are about preservation. In a previous section, I mentioned six aims for virtual heritage projects: care, accuracy, sensitivity, effective and inspirational pedagogical features, and they should be collaborative and evaluation-orientated (Tost and Champion 2011). Extrapolating from these aims, I suggest that the following features are desirable for designing 3D virtual heritage models or for developing an infrastructure that can support virtual heritage models for the purpose of classroom teaching and public dissemination:

1 Data accuracy: the level of accuracy and type of data capture method should be documented and associated with the model (and, if possible, the geographical location).
2 Format limitations: any known limitations or required conditions due to the digital format or way in which the data was created should also be included with the model.
3 Provenance: the record of ownership and scholarship and community input should be recorded and accessible (the source and the ownership rights).

4 Community protocols: social, cultural and institutional protocols that guide who accesses the sourced cultural heritage and how that should regulate the transmission, distribution and dissemination of the digitally simulated model.
5 Authenticity: the known, extrapolated, omitted, simplified and imagined areas and components of the model should be identified in some form of thematic (and preferably standardized) schema.
6 Cultural presence: models should aim towards explaining the cultural significance of the original site and give an impression of the situated cultural value of the place as experienced by the original inhabitants.
7 Evaluation data: these aims should be clearly explained and any evaluation data of participants should be linked to (or otherwise associated with) the models.
8 Purpose: the generic ways in which original creators and shareholders intended the models to be edited or otherwise modified could be described in accompanying text.

The first three considerations (and possibly all of the others, especially 5: Authenticity) require the careful and appropriate use of metadata. As Wise and Miller (1997) have noted, metadata ('data about data'), allows users to be informed without having to access the entire body of data; it helps them find information and it helps them to group and link 'bodies of information' together. In 2008, Addison (2008) proposed the following list of virtual heritage metadata elements:

Type	#	*Data encoding/format*
What	i	HeritageID (a superset of existing WorldHeritageID)
	ii	Title/brief description
	iii	Heritage type/classification (e.g. cultural: archaeological. . .)
	iv	Heritage time period (e.g. geologic or historic time)
	v	Heritage time span
Why	vi	Purpose (reason recorded/produced)
How	vii	Recording device parameters (type, sample rate, precision . . .)
	viii	Secondary device(s) (data manipulation)
	ix	Environmental conditions
Whom	x	Submitter and date of submission
	xi	Rights given/withheld
	xii	Author/copyright holder
	xiii	Sponsor/funder/client
When	xiv	Date (of recording, manipulation)
Where	xv	Location (latitude/longitude + compass direction if applicable)

Although I disagree with their initial selection of 3D-PDF (.pdf) as a file format, a notable success of the CARARE consortium (D'Andrea and Fernie,

2013) was the definition of CARARE Schema (Fernie, 2013). This metadata schema was inspired in part from CIDOC-CRM, which is arguably the best-known cultural heritage ontology framework (Geser and Niccolucci, 2013) and inspired in part by the MIDAS UK metadata standard. It is interesting to compare the list of elements in CARARE Schema (Heritage Asset Identification Set, Digital Resource, Activity and Collection Information) with Addison's proposed metadata. Although the CARARE metadata schema includes a separate Global Information element that holds additional information (record information, appellation, rights, temporal and spatial information, actors, contacts, addresses and a publication statement), I would suggest one further addition: part of the metadata should record the culturally significant cultural heritage features noted above, and the reasons why that heritage environment or artefact deserves to be preserved, simulated and communicated (Dappert and Farquhar, 2009; Hockx-Yu and Knight, 2008; Knight and Pennock, 2009).

Connecting to text and other digitalized resources

Even if we agree on a suitable 3D format supported by a robust and open infrastructure, we also need to leverage the potential of digital media to create new synergies between traditional forms of media. Many of the historic strengths of print-based publishing have now become cumbersome liabilities. Even digitized scholarly articles rarely allow interactivity, they are typically distributed in the PDF format and are plagued by its limitations (pdf files are slow, can crash the computer, take up valuable screen space, are confusing to annotate, and the only application to take full advantage of their features is expensive), and underlying data is seldom conveniently retrievable. Added to these issues is the undeniably messy reality of archaeological excavation and recording. For instance, Reinhard (2014) wrote:

> Archaeology is messy, and it deals with three-dimensional artifacts in four-dimensional space-time. Its publications should reflect that . . . Our new publications must incorporate all of these elements to create a record and interpretation of what we have discovered, leaving that data and interpretation open to criticism, dialogue, and growth over time . . .
>
> There are two major issues that all publishers of archaeology (and of scholarship generally) must address now: 1) how to publish archaeology online, moving away from a traditional, two-dimensional, print-informed model, toward a multi-dimensional, interactive one that accepts that archaeological data is messy and continues to grow and change over time, and 2) how to publish archaeology in an open fashion that makes content easily discoverable and immediately accessible, promoting linking from external sources while linking itself to other open online resources.

How can we link 3D models to library and archival systems holding scholarly literature and multimedia resources that communicate important historical and cultural aspects of the simulated heritage site? Current journals that feature scholarly papers and 3D models typically lack the capability to integrate with text resources, and have limited interactivity and immersion (Elsevier, undated). If, on the other hand, we create dynamic links between 3D models and 2D assets (text and other media), then it may be possible to develop evaluation mechanisms to understand how the viewed and downloaded heritage models and simulations are used and critically reflected on.

Do we have feasible options to achieve this? Consider a publication system which is actually a framework dynamically drawing on various media components through assigned URIs – in other words, a library of Linked Open Data. It could be a journal-publishing framework (like http://scalar.usc.edu/) or a communal blogging and publication framework (for example, www.openeditions.com/), which would dynamically link to URIs of 3D projects. Archaeology also requires a way of updating and augmenting information (Dallas, 2015, 2016), while heritage studies research requires a way of fostering and including community engagement. I am also convinced that scholars would appreciate a way of creating visual scholarly arguments that allows feedback from their colleagues and from the public.

Repositories

I think it is fair to say that we now have many institutional repositories for academic publications and scholarly collections. Despite recent European and North American moves to create archives and digital humanities infrastructures 3D models have not yet been fully incorporated into these new infrastructures while allowing full public access (Huggett, 2012). For example, a major EU project, CARARE (discussed earlier), created a common library format of 3D models, but they were trapped inside the Adobe PDF format so people could not modify and develop their own content, and the model did not dynamically link to the scholarly information that made the model possible. Commercial model repositories offer very consistent formats and protocols for disseminating downloadable models, but these models are either trapped inside a proprietary format that is designed to prevent flexible use, are expensive, prohibit modification and future commercial use, or their accuracy and quality cannot be verified before purchase. Further international efforts to remedy the above issues include work by 3D Icons (3D HOP) in CIDOC CRM, Europeana (discussed in the final chapter), Smithsonian Institute X3D BETA, Fraunhoefer (X3DOM ON GITHUB), Ariadne, EU EPOCH and V-MUST.

While in Europe, ARIADNE and 3D-ICONS are developing standards and archives that may help provide some of the answers; in many other regions, there are very few accessible 3D models of heritage sites that use a common, stable format. Although there are interesting prototypes and

selective web-based prototypes (such as http://vcg.isti.cnr.it/3dhop/and www.3dicons.ie/3d-content) and online commercial suppliers of 3D models of varying quality and accuracy, there is no standard thematic research repository or national data service for 3D heritage models that I have been able to find in the Pacific region or relevant heritage content from overseas websites.

For example, in Australia we simply do not have a shared data infrastructure service that provides specialized support for 3D archaeological and heritage data. Yet the Commonwealth Scientific and Industrial Research Organization of Australia (CSIRO) has released a report (CSIRO 2014), stating: 'Australia's cultural institutions risk losing their relevance if they don't increase their use of digital technologies and services.' Michael Brünig (2014) noted that while the Australian GLAM industry is worth AUD2.5 billion a year, roughly only a quarter has been digitized. This is why we look to Europe for inspiration.

Portals

Recent European trends are to create archives and digital humanities infrastructures, but 3D models have not kept up with the progress achieved for other formats of cultural heritage – they are still silos. In the Europeana portal (www.europeana.eu/portal/) one can search by media type – in this case, 3D. As Europeana is both a portal and a platform, the website visitor can be taken to an external website without knowing if the 3D model can run in the browser or requires specialist equipment (for example, please view the artefacts at http://public.cyi.ac.cy/starcRepo/explore). There are other interesting 3D model websites for cultural heritage institutes such as the Smithsonian, but they do not clearly allow for downloadable usage or explain carefully any cultural protocols that need to be associated with the ways in which the 3D models can be used (http://3d.si.edu/). The Smithsonian http://3d.si.edu/ website is in a sense also a platform: it provides 3D tools to edit and build with, but it is not a complete platform in terms of infrastructure – it is one way. Users can play with 3D models, sometimes edit them, sometimes download them, but they cannot permanently alter what stays online in the website database and the digital 3D model does not provide full archival records or other information resource links.

Portals, unlike platforms, merely collect or link to other sites and resources; they don't provide their own tools and resources to build things with. They still have their uses: portals can assemble disparate information conveniently and in a more useful standardized format; they attract more visitors than individual sites and provide larger amounts of web-traffic statistics. They can also allow shareholders (individual website owners) the ability to retain original assets while appearing as part of a greater digital collection.

Augmented reality

Twenty years ago, Ron Azuma (1997) published a survey on augmented reality (AR), contrasting it with Virtual Environments (VE) and virtual reality (VR). In that paper, Azuma declared that unlike VR, AR supplements the real world, superimposing virtual objects on the real world or compositing virtual objects on the real world. Today's AR phone applications do not really qualify. They do not use computer vision to merge 3D data, but simply rely on the GPS tracking of the camera. Even with this less sophisticated and less spatially immersive technology, the potential for virtual heritage is obvious (Billinghurst *et al.*, 2015; Chung *et al.*, 2015; tom Dieck and Jung, 2015). Augmented reality does not have to create or re-create an entire scene and it can now be carried on consumers' mobile phones because it does not require the same graphic rendering as a full virtual reality environment. Also, it can range in complexity from augmented avatars that appear on your phone or head-mounted display (HMD) to text labels on the screen of a phone camera, appearing to float above real-world objects (Dredge, 2011). There are even research projects examining how AR can integrate with museum collections (Kraemer and Kanter, 2014) and community-based projects (Speiginer *et al.*, 2015).

In 2016, I was invited to UCLA for the second National Endowment of the Humanities (NEH) *Advanced Challenges in Theory and Practice in 3D Modeling of Cultural Heritage Sites* week-long workshop. At the first workshop in 2015 at the University of Massachusetts Amherst, some of the participants decided to start or continue augmented reality history and heritage projects; the technology looked fascinating. Many AR applications could be used on Android or iOS phones; some were free, some had clear and easy-to-use examples, and the increasing power and ease of phone-based cameras along with their increasing computational power and GPS accuracy. The recent craze of Pokémon Go (Moskowitz, 2016) may have also convinced heritage professionals of a huge potential education and cultural tourism market. In terms of technological promise, the rise of software such as Layar (www.layar.com/), Aurasma (www.aurasma.com/), Vuforia (www.vuforia.com/), along with suitable archaeology, museum, heritage and history case studies, seems to bear this out.

Nevertheless, in terms of infrastructure, augmented reality is a quagmire. When I visited the NEH workshop in 2016 (I could only attend the 2015 workshop virtually), I was surprised by the number of educators and designers who had run into major issues with augmented reality heritage projects. In this field, we are all used to strange formats, unreliable software, bad user documentations and the collapse of certain software. What was new to me was the scale of the augmented reality company failures. In some cases, the software would store all digital media assets on the Cloud (on distant servers). However, when the software disappeared, so did the original data. As free software, the companies had not made the transfer and secondary

storage of digital media projects or the digital media assets convenient, or, in some cases, even possible. Obscure formats, hidden files, incompatible version updates: these are all common when taking up new software and the hype cycle of new technology can be seductive (Gaudiosi, 2016; Gilbert, 2015; Mainelli, 2016). But not being able to access the data created by the user? This was clearly a failure of infrastructure. While the online press predicted a future where AR made screens vanish in the real world (Chapman, 2016), the AR products or even AR companies were the ones doing the vanishing (Miiler and Constine, 2015). As Ogden (2015) notes, commercial AR products are walled gardens, with little in the way of shared standards.

The lesson learnt here: compose your content online, have it stored in the cloud, and when the software changes or the company goes bankrupt or is sold to an even larger company, be prepared not only to lose your project, but also the digital media assets that made up that project. It is always desirable to acquire software from a proven, reliable company that allows the user to store the project and assets offline, provide and accept standard formats, can be linked to other media and other referencing systems, and has the ability to export in different formats.

A digital scholarly eco-system for 3D digital heritage

Even if we find robust technology and a good range of heritage models that are reliably stored, have provenance data and allow the community to edit and modify them, we have a further problem. Current examples lack initial meaningful context, audience feedback, and updated and maintained content. Despite a plethora of web archives of digital tools and models, there is simply no consistent way for the community to provide feedback. We need to develop ways of linking 3D models back to the scholarly resources that created them and inform them.

In the proceedings of the 2015 *Computer Applications and Quantitative Methods in Archaeology (CAA2015)* conference, I noticed an explosion of papers on Linked Open Data. Before too long, we should see systematic ways for Linked Open Data to connect to the text files associated with 3D models. However, we also need to develop ways of dynamically linking models and subcomponents of models to dynamic but stable documentation on the Internet (Haslhofer and Isaac, 2011). This should be a dynamic two-way link that would link text, 3D models, other media and community feedback (scholarly reviews, classroom projects that expand, review and comment on the material).

Furthering the development of a digital heritage journal requires the support of a community of specialists. For example, it could necessitate incorporating the taxonomy research of CIDOC and NeDiMAH, the 3D tools directory of DiRT Bamboo, the projects and community of DHCommons and centerNet, and international networks of digital humanities centres. What would or could such a system be? Could tools, methods, projects,

scholarly communities and an open access online journal-publishing system exist to communicate between and beyond digital research infrastructures, versed in text or in 3D models? Could active communities be encouraged to adopt and extend this 'ecosystem'? More specifically, how could it benefit research and practice into digital heritage? What are the specific needs and challenges of digital cultural heritage? Which components, user requirements and test beds should be addressed if such a scholarly ecosystem is to benefit digital cultural heritage?

Given that I had already proposed a definition of virtual heritage, why have I returned to the concept of digital heritage here? Much of today's heritage resources are text and two-dimensional. The development of low-cost and accessible virtual heritage equipment is still around the corner; hence, it makes more sense to develop digital heritage resources that will be able to be used as content and context when tomorrow's immersive technologies finally take hold in the general public. Digital heritage can provide dynamic content to virtual heritage, but it should be separated in name and in practice from the latter, for the latter undergoes constant change.

A virtual heritage repository

Three major thematic issues could prove to be of great import to a virtual heritage repository. First, VR equipment is moving towards the consumer level, based on the notion of a component-based system whereby your smartphone is both the stereoscopic viewer and the computer (such as in the case of the Samsung Gear). Such consumer technology frameworks will help VR technology and related content become far more accessible. Second there are research groups so concerned at the silo mentality of earlier virtual heritage projects that they are developing technology solutions that allow people to create their own content using free and open source technology such as the EU CHESS project (Pujol *et al.*, 2012), or they are providing technical exemplars using free software that others can download, modify and learn from. Third journals are beginning to provide technology that allows authors to add 3D models inside or next to text-based articles. Two journals that come to mind are *Internet Archaeology* (http://intarch.ac.uk/), discussed in this volume, and *Digital Applications in Archaeology and Cultural Heritage* (www.journals.elsevier.com/digital-applications-in-archaeology-and-cultural-heritage/).

In 2015, Cameron Neylon wrote:

> What should a shared infrastructure look like? Infrastructure at its best is invisible. We tend to only notice it when it fails. If successful, it is stable and sustainable. Above all, it is trusted and relied on by the broad community it serves. Trust must run strongly across each of the following areas: running the infrastructure (governance), funding it (sustainability), and preserving community ownership of it (insurance).

I agree with Neylon and would distinguish between hard infrastructure (equipment) and soft infrastructure (people), as both are necessary (Gotbaum, 2011). I previously suggested (Champion, 2014) that a digital humanities network will not survive for long if it does not create effective synergies between equipment and people. While some scholars in the digital humanities have argued that research infrastructures are not research per se (Rockwell, 2012), I would like to point to the *European Research Infrastructure Consortium* (ERIC) practical guidelines. The guidelines clearly state (European Research Infrastructure Consortium, 2015) that 'the ERIC status is reserved for state-of-the-art research infrastructures that will create unique opportunities to carry out advanced research, attract the best researchers from across the world and train highly qualified students and engineers'.

Quality research infrastructure is measured by research, by the quality of the contributors, by the impact of their contribution and by the effect of the research infrastructure on them as researchers. Following these premises, a useful research infrastructure in the new digital age is a scholarly ecology, an ongoing scholarly appraisal and reconfiguration of all media assets and outputs. It would be *digital* in order to leverage the specific benefits of digital media and digital-savvy audiences. Changes over time, different input mechanisms and learning mediums allow counterfactual exploration, log user responses, track user preferences, and share insights and personal feedback from distributed audiences. It would be *scholarly*, providing associated tools, interpretative mediums and careful references as well as usage data that could also provide evidence for solid scholarly arguments. Above all, it would be an *ecosystem*. All its parts would be interdependent and it would hopefully be greater than the sum of its parts. A review community could be summoned to discuss and add to the models via publications and related links. Future publications could in turn integrate the community feedback into new research findings, improved critiques and an enhanced research base.

Technical obstacles and potential answers

We require a cross-platform long-term robust solution that has ongoing community support and a flexible, rich feature list. Unfortunately, virtual heritage projects are typically walk-around interactions only, possibly with pre-recorded text or voice-based information. There are a few instances of Kinect-based camera tracking or data-driven environment enrichment, but these projects are still relatively rare and the preservation strategy of these interactive approaches is seldom discussed.

Overall, the projects are monolithic, not traceable, not reconfigurable, not easily preserved, and do not link to external 2D material either statically or dynamically. Because of their inherent resistance to reconfiguration or to added audience feedback, these projects do not have strong ongoing community reach, support and input. Despite a shift to open access models

and greater collaboration with the public, Brünig (2014) cautioned that there is an ongoing need to explore new approaches to copyright management that stimulate creativity and support creators.

I propose the adoption of a component-based system that can load a robust file format and add links and media assets to create a dynamic and interactive online environment that can be taken apart and further modified by the public. Ideally, the web model can include specific camera angles that can be triggered by scripts or other cues in case the viewer wishes to be guided through the simulation. Both the holding page of the archived model and the web model should provide suitable metadata and includes provenance metadata (Huggett, 2014). The models stored in the system would also link dynamically to external scholarly repositories. If there are shareholder requirements that stipulate copyright ownership of the high-resolution model, an agreement could be negotiated so that a lower resolution model or a model with reduced interaction features is provided to the public.

Shareholders and communities

It may appear that the overall number and difficulty of technical issues is the major problem to resolve, but if there is no public involvement, understanding and appreciation, the virtual heritage project has failed *despite* any technical brilliance or infrastructure support. Infrastructure that is not used is not really infrastructure, it is merely equipment. Previous writers have written convincingly of the importance of archives (Limp *et al.*, 2010), but there is also another important requirement – ensuring that the archive is effectively used. As Garnet and Edmond (2014) have declared: 'Building an API is not enough!' The success of virtual heritage projects is dependent on community involvement, which includes scholars, students, the wider public, but also the original shareholders and owners of the cultural content simulated.

As we develop models, frameworks and infrastructure, we may need to consider how to approach indigenous heritage individuals and groups regarding access to the recording and dissemination of specific cultural heritage content. These considerations may require:

1 Involvement with indigenous shareholders and experts in the development of guides and protocols and the sensitive development of digital heritage knowledge.
2 Exploring digitally filtered ways of creating accessible layers and levels of cultural knowledge.
3 Researching and testing a method for providing on-demand 3D model formats where the level of access can determine the accuracy and resolution of the generated model to suit the copyright and ownership requirements of the owners and creators while providing a predetermined level of public knowledge.

4 Tailoring digital ontologies, indigenous record-collection metadata and folksonomies to specific aspects of heritage simulations.
5 Developing new forms of copyright permissions that are relevant to the cultural significance and guardianship of the heritage objects depicted.

For example, in Australia, rock art and other heritage artefacts may have cultural taboos placed on how they are presented or viewed, whether the audience can be outside the 'mob' or even be of a certain gender. Luckily, in Australia there are guidelines both at national level (Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS, 2012) and at international level (UNESCO, 2015; United Nations, 2008) and there is a great deal of scope to help communities establish how these policies and guidelines can work with the capture, augmentation and dissemination of sensitive cultural heritage data, indigenous or non-indigenous. I suggest that adherence to these guidelines is indicated in the metadata.

Where tutorials, tools and training materials are to be developed for indigenous communities or for using with indigenous content, they should be developed after consultation with relevant research conduct policies and ethical guidelines. In Australia, this may mean following documents like the *Guidelines for Ethical Research in Australian Indigenous Studies* (GERAIS) and the *UN Declaration on the Rights of Indigenous Peopl*es (UNDRIP) with awareness of and commitment to free, prior and informed consent protocols that will address Intellectual Property issues, copyright requirements and other related permissions (such as for multimedia and other media content).

Conclusion: a new virtual heritage infrastructure

I sought to advance three major points in this chapter. First, virtual heritage will not succeed as digital heritage if it cannot even preserve its own models and it will not be effective if it cannot implement the most important advantages of digital technologies (real-time reconfiguration to suit the learner, device and task at hand, individual personalization, increased sense of agency, automatic tracking and evaluation mechanisms and filtered community feedback). My suggestion is to implement not so much a single file format but to agree upon a shared relationship between assets. For want of a better word, I have described the overall relationship of components of virtual heritage infrastructure as a digital scholarly ecosystem.

Second, in this new age of digital communication the 3D model must be recognized as a key scholarly resource (Di Benedetto *et al.*, 2014). As a core part of a scholarly ecosystem the 3D model should be traceable; it should link to previous works and to related scholarly information. I suggest that the model should be component-based so that parts can be directly linked and updated. Web models could be dynamically created at runtime. The model should be engaging, thus requiring extensive play-testing and evaluation to ensure that it actually does engage its intended audience. As part of a

scholarly infrastructure, the 3D model format (and all related data formats) should be easy to find and reliable. It should not require huge files to download, or it should at least provide users with enough information to decide whether and what to download. Metadata can also help record the completeness, measurement methodology and accuracy of the models, and Linked Open Data can help connect these media assets in a sensible and useful way.

Third, the community of scholars, students and the wider public should be involved, and we must endeavour to incorporate their understanding, feedback and participation (in line with the fact that this is a core requirement of UNESCO World Heritage status for physical monuments). Community involvement is necessary for scholars as well, so I suggest that virtual heritage projects dynamically link to journals and refereed conference papers, and to the list of tools and methods that were used. A robust feedback system could help continually improve the system. Other shareholder issues such as varying levels of learning skills and varying levels of knowledge required or cultural knowledge that needs to be hidden (privacy and ownership issues) should also be incorporated into the project.

3D heritage models are here to stay, but that does not mean they are addressing the aims and objectives of heritage. They are too essential and important to be locked away in proprietary, unsupported file formats. As designers, practitioners and educators, we need to provide both public access and robust preservation infrastructures to support them. Technology is only one part of the solution; we must also develop incentives, guidelines and frameworks.

References

Addison, A. C. (2000) Emerging trends in virtual heritage, *Multimedia, IEEE* 7: 22–5.

Addison, A. C. (2008) The vanishing virtual: Safeguarding heritage's endangered digital record, in Kalay, Y. E., Kvan, T. and Affleck, J. (eds) *New Heritage: New Media and Cultural Heritage*. Routledge, Oxford, pp. 27–39.

Addison, A. C., Refsland, S. and Stone, R. (2006) Special issue: Virtual heritage guest editors' introduction, *Presence: Teleoperators & Virtual Environments* 15: iii–iv.

Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) 2012 Guidelines for Ethical Research in Australian Indigenous Studies (GERAIS).

Azuma, R. T. (1997) A survey of augmented reality, *Presence: Teleoperators and Virtual Environments* 6: 355–85.

Baio, A. (2015) Never trust a corporation: As Google abandons its past, Internet archivists step in to save our collective memory, *The Message*. The Message, Online.

Barsanti, S. G., Remondino, F., Fenández-Palacios, B. J. and Visintini, D. (2014) Critical factors and guidelines for 3D surveying and modelling in Cultural Heritage, *International Journal of Heritage in the Digital Era* 3: 141–58.

Billinghurst, M., Clark, A. and Lee, G. (2015) A survey of augmented reality, *Foundations and Trends® Human–Computer Interaction* 8: 73–272.

Bradley, K. (2015) Why museums hide masterpieces away, *BBC: Culture*. BBC.

Brünig, M. (2014) Historic collections could be lost to 'digital dinosaurs', *the conversation*. The Conversation, Online. https://theconversation.com/historic-collections-could-be-lost-to-digital-dinosaurs-31524

Carroll, C. (2012) *News Brief: Carnegie Mellon Awarded Grant To Preserve Executable Content*. Carnegie Mellon University News, www.cmu.edu/news/stories/archives/2012/october/oct4_executablecontent.html (accessed 31 January 2017).

Champion, E. (2008) Explorative shadow realms of uncertain histories, *New Heritage: New Media and Cultural Heritage*. Routledge, Oxford, pp. 185–206.

Champion, E. (2014) Researchers as infrastructure, *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. Sheffield*. HRI Online Publications, Sheffield, p. Online.

Chapman, G. (2016) Augmented reality looks to future where screens vanish, *Phys.org*, Online.

Chung, N., Han, H. and Joun, Y. (2015) Tourists' intention to visit a destination: The role of augmented reality (AR) application for a heritage site, *Computers in Human Behavior* 50: 588–99.

CSIRO (2014) *Australian museums must innovate or risk becoming 'digital dinosaurs'*, www.csiro.au/Portals/Media/Australian-museums-risk-becoming-digital-dinosaurs.aspx (accessed 27 October 2014).

D'Andrea, A. and Fernie, K. (2013) CARARE 2.0: a metadata schema for 3D Cultural Objects, in Addison, A. C., De Luca, L., Guidi, G. and Pescarin, S. (eds) *Digital Heritage International Congress (Digital Heritage 2013)*. IEEE, Marseille, France, pp. 137–43.

Dallas, C. (2015) Curating archaeological knowledge in the digital continuum: From practice to infrastructure, *Open Archaeology* 1.

Dallas, C. (2016) Jean-Claude Gardin on archaeological data, representation and knowledge: Implications for digital archaeology, *Journal of Archaeological Method and Theory* 23: 305–30.

Dappert, A. and Farquhar, A. (2009) Significance is in the eye of the stakeholder, *International Conference on Theory and Practice of Digital Libraries*. Springer, pp. 297–308.

De Reu, J., Plets, G., Verhoeven, G., De Smedt, P., Bats, M., Cherretté, B., De Maeyer, W., Deconynck, J., Herremans, D., Laloo, P., Van Meirvenne, M. and De Clercq, W. (2012) Towards a three-dimensional cost-effective registration of the archaeological heritage, *Journal of Archaeological Science* 40: 1108–21.

Denard, H. (2009) *The London Charter for the Computer-Based Visualisation of Cultural Heritage*, www.londoncharter.org/ (accessed 18 July 2016).

Di Benedetto, M., Ponchio, F., Malomo, L., Callieri, M., Dellepiane, M., Cignoni, P. and Scopigno, R. (2014) Web and mobile visualization for cultural heritage, *3D Research Challenges in Cultural Heritage*. Springer, pp. 18–35.

Dredge, S. (2011) What is mobile augmented reality for? Overlaying information onto the real world looks cool, but its real value could be as an interface rather than as a killer app, *The Guardian. The Guardian*, p. Online.

Dylla, K., Frischer, B., Müller, P., Ulmer, A. and Haegler, S. (2008) Rome reborn 2.0: A case study of virtual city reconstruction using procedural modeling techniques, *Computer Graphics World* 16.

Economou, M. and Pujol, L. (2008) Educational tool or expensive toy? Evaluating VR evaluation and its relevance for virtual heritage, *New Heritage: New Media and Cultural Heritage*, London, Routledge: 242–60.

Elsevier (undated) *Three-dimensional U3D Models in Online Journal Articles*, www.elsevier.com/about/content-innovation/u3d-models.

European Research Infrastructure Consortium (2015) ERIC practical guidelines: Legal framework for a European Research Infrastructure Consortium. European Commision, Brussels, Belgium: European Commission.

Fernie, K. (2013) CARARE: Connecting archaeology and architecture in Europeana, *Uncommon Culture* 3: 85–92.

Garnett, V. and Edmond, J. (2014) Building an API is not enough! Investigating reuse of cultural heritage data. LSE: Maximising the impact of academic research, *LSE–London School of Economics BLog*.

Gaudiosi, J. (2016) 7 Ways AR and VR will change tech in 2016, *Fortune Tech*, Online.

Geser, G. and Niccolucci, F. (2013) Virtual museums, digital reference collections and e-science environments, *Uncommon Culture* 3: 12–37.

Gilbert, B. (2015) *It's Going to be at Least 5 Years Before Virtual Reality Goes Mainstream*, www.businessinsider.com.au/virtual-reality-on-gartner-hype-cycle-2015-8 (accessed 30 January 2017).

Gotbaum, R. (2011) *The Difference Between Soft And Hard Infrastructure, And Why It Matters*, http://stateimpact.npr.org/new-hampshire/2011/10/26/infrustructure-soft-and-hard/ (accessed 31 January 2017).

Greenop, K. and Barton, J. (2014) Scan, save, and archive: how to protect our digital cultural heritage. *The Conversation*.

Haslhofer, B. and Isaac, A. (2011) data. europeana. eu: The europeana linked open data pilot, *International Conference on Dublin Core and Metadata Applications*, pp. 94–104.

Havemann, S. (2012) Intricacies and potentials of gathering paradate in the 3D modelling workflow, in Bentkowska-Kafel, A., Denard, H. and Baker, D. (eds) *Parada and Transparency in Virtual Heritage*. Ashgate, London, pp. 146–60.

Hockx-Yu, H. and Knight, G. (2008) What to preserve?: Significant properties of digital objects, *International Journal of Digital Curation* 3: 141–53.

Huggett, J. (2012) Lost in information? Ways of knowing and modes of representation in e-archaeology, *World Archaeology* 44: 538–52.

Huggett, J. (2014) Promise and paradox: Accessing open data in archaeology, in Mills, C., Pidd, M. and Ward, E. (eds) *Proceedings of the Digital Humanities Congress 2012*. HRI Publications, University of Sheffield, Sheffield.

Knight, G. and Pennock, M. (2009) Data without meaning: Establishing the significant properties of digital research, *International Journal of Digital Curation* 4: 159–74.

Koller, D., Frischer, B. and Humphreys, G. (2009) Research challenges for digital archives of 3D cultural heritage models, *Journal on Computing and Cultural Heritage (JOCCH)* 2: 7.

Kraemer, H. and Kanter, N. (2014) Use and re-use of data how Collection Management Systems, Transmedia and Augmented Reality impact the future of museum, *Virtual Systems & Multimedia (VSMM), 2014 International Conference on*. IEEE, pp. 214–16.

Lepore, J. (2015) The cobweb–Can the internet be archived?, *new yorker. The New Yorker*, New York. http://www.newyorker.com/magazine/2015/01/26/cobweb

Limp, W. F., Payn, A., Winters, S., Barnes, A. and Cothren, J. (2010) Approaching 3D digital heritage data from a multitechnology, *38th Conference on Computer Applications and Quantitative Methods in Archaeology*, Granada, Spain.

Lopez-Menchero, V. M. and Grande, A. (2011) The principles of the Seville Charter, in Pavelka, K. (ed.) *CIPA Symposium Proceedings*. Czech Technical University in Prague, Faculty of Civil Engineering, Prague, pp. 2–6.

Mainelli, T. (2016) *Tapping the Brakes on the Virtual-Reality Hype Machine, recode*, Online.

McHenry, K. and Bajcsy, P. (2008) An overview of 3d data content, file formats and viewers, *National Center for Supercomputing Applications* 1205.

Michaelis, N., Jung, Y. and Behr, J. (2012) Virtual heritage to go, *Proceedings of the 17th International Conference on 3D Web Technology*. ACM, Los Angeles, CA, pp. 113–16.

Miiler, R. and Constine, J. (2015) *Apple Acquires Augmented Reality Company Metaio*. techcrunch.

Moskowitz, B. (2016) *Pokémon GO in Museums: Gotta visit 'em all*, https://museumhack.com/pokemongo-museums/ (accessed 30 January 2017).

Neylon, C. (2015) *Principles for Open Scholarly Infrastructures*, http://figshare.com/articles/Principles_for_Open_Scholarly_Infrastructures_v1/1314859.

Ogden, H. (2015) *Augmented Reality: Believe The Hype (Cycle), techcrunch*, techncrunch.com.

Pitzalis, D., Niccolucci, F., Theodoriou, M. and Doerr, M. (2010) LIDO and CRM dig from a 3D cultural heritage documentation perspective, proceedings of the 11th International Conference on Virtual Reality, Archaeology and Cultural Heritage. Eurographics Association, pp. 87–95.

Pujol, L., Roussou, M., Poulou, S., Balet, O., Vayanou, M. and Ioannidis, Y. (2012) Personalizing interactive digital storytelling in archaeological museums: the CHESS project, 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology. Amsterdam University Press.

Reinhard, A. (2014) *Publishing Archaeological Linked Open Data: From Steampunk to Sustainability*. NYU Library's Ancient World Digital Library, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/reinhard/ (accessed 31 January 2017).

Rockwell, G. (2012) As transparent as infrastructure: On the research of cyberinfrastructure in the humanities, *Connexions*: Online.

Roussou, M. (2002) Virtual heritage: From the research lab to the broad public, *Bar International Series* 1075: 93–100.

Speiginer, G., MacIntyre, B., Bolter, J., Rouzati, H., Lambeth, A., Levy, L., Baird, L., Gandy, M., Sanders, M. and Davidson, B. (2015) The evolution of the argon web framework through its use creating cultural heritage and community-based augmented reality applications, *International Conference on Human–Computer Interaction*. Springer, pp. 112–24.

Stone, R. and Ojika, T. (2000) Virtual heritage: What next?, *IEEE Multimedia* 7: 73–4.

Thwaites, H. (2013) Digital heritage: What happens when we digitize everything?, in Ch'ng, E., Gaffney, V. and Chapman, H. (eds) *Visual Heritage in the Digital Age*. Springer, London, pp. 327–48.

tom Dieck, M. C. and Jung, T. (2015) A theoretical model of mobile augmented reality acceptance in urban heritage tourism, *Current Issues in Tourism*: 1–21.

Tost, L. P. and Champion, E. (2011) Evaluating presence in cultural heritage projects, *International Journal of Heritage Studies* 18: 83–102.

UNESCO (2003) Convention for the safeguarding of the Intangible Cultural Heritage, in UNESCO (ed.). UNESCO, Paris, pp. 1–15.

UNESCO (2015) The operational guidelines for the implementation of the World Heritage Convention, *World Heritage Centre*, UNESCO, Paris.

United Nations (2008) UN Declaration on the Rights of Indigenous Peoples, United Nations. https://www.humanrights.gov.au/publications/un-declaration-rights-indigenous-peoples-1

Widdowson, J. (2014) *Can Libraries Survive Disruptive Technologies?* www.cilip.org.uk/cilip/blog/can-libraries-survive-disruptive-technologies (accessed 9 December 2014).

Wise, A. and Miller, P. (1997) Why metadata matters in archaeology, *Internet Archaeology*, p. 81.

3 *Internet Archaeology* and digital scholarly communication

Julian D. Richards

Introduction

The e-journal *Internet Archaeology* was first published in 1996. At the time, it was the first fully online peer-reviewed e-journal in any discipline. At the time of writing, over 20 years later, it is in its forty-second issue and still publishing rich interactive content, including monograph length multi-layered articles, online databases and Geographic Information System (GIS) interfaces, virtual reality models, and multimedia sound and movie files. All content in archived by the UK's Archaeology Data Service, and the journal has won several awards for its creative exemplars of linked e-publications and archives.

When *Internet Archaeology* was established, the Internet itself was still in its infancy. The journal has had a transformative effect on scholarly communication in archaeology and a significant impact on the humanities more broadly. The goals of this chapter are to review the challenges faced in developing a fully online journal and the particular difficulties in establishing a sound publishing business model, setting them in the wider context of developments in electronic publication, Open Access and Open Data. The chapter will try to demonstrate how *Internet Archaeology* has adapted to a rapidly changing external environment.

Online publication

Acceptance of electronic publication has been slower than many predicted and we have always felt that *Internet Archaeology* has been ahead of the curve. As early as 1978 Wilfrid Lancaster wrote:

> In my opinion, there is no real question that completely paperless systems will emerge in science and in other fields. The only real question is 'when will it happen?' We can reasonably expect, I feel, that a rather fully developed electronic information system . . . will exist by the year 2000, although it could conceivably come earlier.
>
> (Lancaster, 1978, 355)

That we have still not, 15 years later than Lancaster predicted, completed the transition to a full electronic system is down to a combination of resistance from both publishers and readers (Richards, 2006). Many publishers have simply sought to migrate their traditional business model to the Internet, setting up web-based services that give access to electronic versions of existing printed journals, often using the Portable Document Format (PDF) to maintain the typography and layout of the printed version (Day, 1999). For many, the shift actually allowed them to decrease costs, reducing printing costs and using the Internet as a means of distribution. They found common cause with librarians who, with budgets under pressure from spiralling journal inflation, found they could cut shelving and staffing costs if they moved from a collections policy to an access policy.

Open access publication

However, for the more radical proponents of electronic communication, such as Stevan Harnad, the move to electronic subscriptions did not go far enough. They asked why the status quo in paper journals should simply duplicate itself in the new medium (Harnad and Hemus, 1997; Harnad, 2001). They saw self-publishing (sometimes called self-archiving) through the Internet as 'a means of returning the responsibility of ownership and distribution of scholarship to its creators' (Day, 1999).

One of Harnad's basic assumptions was that when scholars and scientists publish in peer-reviewed journals they are not primarily interested in monetary reward – which would, in any case, be unlikely – but in having their work read, used and referenced (Harnad and Hemus, 1997). In the 'Gutenberg era', authors had to maintain what Harnad called a 'Faustian bargain' with commercial publishers, whereby they handed over their copyright in return for having their research published (Harnad and Hey, 1995). Harnad argued that this made sense when publishing remained an exclusive and expensive domain, but that it had no relevance in the electronic era when scholars can publish their own papers at little or no personal cost. He argued that authors should make the texts of their papers freely available on the Internet and that readers would then access the free electronic version of a paper rather than a more expensive paper version published much later (Harnad and Hey, 1995, 114–15). One of the first exemplars of the 'subversive-proposal' in action was the 'e-print archive' for high-energy physics set up by Paul Ginsparg at the Los Alamos National Laboratory in 1991 (Ginsparg, 1994). It very quickly became the primary means of scholarly communication in its subject area and has since expanded to cover the whole of physics, mathematics and computer science.

However, as institutions have sought to promote the research undertaken by their own scientists and researchers, and to increase its impact, many have developed their own institutional eprint repositories, creating a fragmented hybrid publication landscape, split between discipline-based

data centres, generic institutional e-repositories and traditional publishers. Harnad's suggestion that eprints should be made freely available prior to formal publication has evolved into what is now often referred to as 'Green' open access (OA), whereby authors publish in any journal and then self-archive their pre-publication text in their institutional repository or on some other open access website. By contrast, 'Gold' open access is provided by authors publishing in an open access journal that provides immediate OA to all of its articles on the publisher's website (Harnad, 2005). In order to be free at the point of use, these journals are generally funded by subventions drawn from authors, and ultimately research-funding bodies, known as Author Processing Fees (or APCs). In addition, in this transitional era, there are also what are known as 'hybrid' open access journals which provide Gold OA only for those individual articles for which their authors (or their author's institution or funder) pay an OA publishing fee.

In the UK, the development of institutional repositories and the pressure for OA has been accelerated by the periodic Research Excellent Framework (REF) whereby the research outputs of academic staff are subject to peer assessment, with those institutions achieving higher scores gaining additional research funding. In 2012, the report from the National Working Group on Expanding Access to Published Research Findings (the 'Finch Group') was published (Finch, 2012). The report recognized the need for different channels to communicate research results, but recommended support for the 'gold' route in particular. UK Research Councils have used the findings of the group to further develop their own policies, and the UK Higher Education funding bodies have introduced a new policy for open access in relation to research assessments after the 2014 REF. The policy states that, to be eligible for submission to the post-2014 REF, authors' final peer-reviewed manuscripts must have been deposited in an institutional or subject repository on acceptance for publication (HEFCE, 2014). Deposited material should be discoverable, and free to read and download, for anyone with an Internet connection. In turn, most UK academic institutions now require staff to deposit such copies in their institutional repository. Although the current REF requirement applies only to journal articles and conference proceedings with an International Standard Serial Number and not to monographs, some institutions have already extended it to all publications.

The concept of Open Access publication has gained considerable momentum in archaeology, but concern has also been expressed for the future of learned societies and their journals. Publication still carries a cost, even if it can continue to depend on the goodwill of editors and reviewers, or their employers, who are willing to contribute their time for free. The publication infrastructure, whether it is a traditional journal, an ejournal, or an eprint repository, still costs money to staff and maintain. The Open Archiving movement simply transfers the cost from the publisher on to the university libraries that maintain the repositories.

Open data

Within archaeology, the debate on openness has typically focused on 'open access' publication, and has been particularly focused on its impacts on the 'traditional' outputs of research and grey literature (Lake, 2012). Yet, as the benefits of openness within archaeological publication have been recognized, its expansion to the structured data produced during archaeological research and fieldwork seems logical. Within archaeology, we have long recognized the benefits and potential impact that the sharing and reuse of data can bring. The benefits of increased accessibility, and the messages of open access and open data, are especially relevant for archaeology, given the primary and unrepeatable status of most data sets. Indeed:

> within a discipline that relies upon destructive research methods, lack of information sharing not only inhibits scholarship, but also represents a tragic loss of irreplaceable cultural and historical knowledge. The discipline urgently requires a more professional approach if researchers are to make credible and replicable knowledge claims and act as better stewards of cultural heritage.
>
> (Kansa and Whitcher Kansa, 2013, 88)

As a profession, archaeologists have sometimes been reluctant to share their primary research data with others. For some, this is attributed to the technical barriers associated with providing access to data (Condron *et al.*, 1999; Kansa and Whitcher Kansa, 2013), or more practical restrictions on the dissemination of data imposed by publishers or data providers. Yet by far the greatest hurdle to overcome is conceptual; while Pratt has observed that 'archaeologists are eager to find ways to publish these data sets' (2013, 101), some remain unconvinced about the benefits that open data promotes. Others may be reluctant to expose perceived deficiencies in primary data recording to the critical scrutiny of their peers, or may believe that there is a risk that their data will be published by others before they have the opportunity to do it themselves. An awareness of the academic, symbolic and economic 'capital' of archaeological data streams has hindered the sharing of data (Porter, 2013), while potential misuse and misappropriation of data have always been concerns. For Kansa, 'the discipline should not continue to tolerate the personal, self-aggrandizing appropriation of cultural heritage that comes with data hoarding'; indeed, data withholding 'represents a clear threat to preserving the archaeological record' (2012, 507).

Such cultural reluctance is not new to archaeology; these issues have not precluded the sharing of data in the past, but have simply constrained the scale of dissemination. Within the current climate with disparate groups and communities conducting related research, where the scale of research and the data produced has increased exponentially, such an approach

is unsustainable. Open data offers researchers a mechanism to improve disciplinary interaction and, as a consequence, enhance research. Increased accessibility has the potential to allow others to test the validity of our interpretations, allowing them to examine and reanalyse the original data. As Lake contends, these 'approaches to knowledge have the potential to bolster scientific rigour by increasing transparency' (2012, 473). At the same time, this transparency can serve to illustrate the professionalism of data creators by highlighting good research practice (Kansa, 2012).

While increased accessibility and reuse has done much to raise awareness of the intrinsic value of research data, official recognition of its importance has served to encourage data creators to share these outcomes. The UK Government, for example, has stated that:

> The work of researchers who expend time and effort adding value to their data, to make it usable by others, should be acknowledged as a valuable part of their role. Research funders and publishers should explore how researchers could be encouraged to add this value.
>
> (UK Government, 2011)

Despite this change in mindset, the data outputs of archaeological research can still be treated with some diffidence – an incongruent outcome of less significance than the final interpretation or synthesis. Costa *et al.* propose that in order to overcome this perception, archaeological data needs to be treated as 'a more relevant part of the archaeological publication, research, management, curation and policy process, and not merely an afterthought' (Costa *et al.*, 2013; Atici *et al.*, 2012; Pratt, 2013). The solution advocated by many is to treat the dissemination of data as a form of publication, one that should employ established practice found within text-based publishing, included citation and editorial control (Kansa *et al.*, 2010; Kansa and Whitcher Kansa, 2011). This, it is believed, will instil a sense of familiarity to the process of disseminating and citing digital resources. This move towards what is termed 'data sharing as publication' is intended make the dissemination of data 'a more regular and integral part of professional practice' (Atici *et al.*, 2012, 161).

Internet Archaeology: adventures in e-publishing

Internet Archaeology (intarch.ac.uk) was established in 1995 with initial funding until 1998 (later extended until 2001) from the Joint Information Systems Committee (JISC) eLib electronic libraries programme (Rusbridge, 2001). Its first paper – a searchable visual catalogue of Roman amphorae in Britain by Paul Tyers – was published in September 1996, just three years after the release of Mosaic, the first windows-based web browser. *Internet Archaeology* was the first refereed online e-journal in archaeology and has been very successful in gaining international recognition as a high-quality

academic journal. It is still unique in archaeology in that it is a multi-media journal available exclusively on the web; it has no print equivalent. It includes elements that would be impossible in a paper publication, such as searchable database and map interfaces to analyse online; full-colour, interactive multimedia; video footage; virtual reality models and access to related digital archive material. The idea from the outset was that the articles would enable readers to drill down into the data, to test interpretations and to put forward rival hypotheses (Heyworth *et al.*, 1996).

The proposal to develop the journal was put forward by a consortium which comprised the Council for British Archaeology, the British Academy and the University of York. The founding editor was Alan Vince, who worked part-time on the journal until 1999, supported by Judith Winters, who then took over as Editor. The running of the journal is overseen by two co-directors, Julian Richards and Mike Heyworth, representing the host institution (University of York) and the publisher (the Council for British Archaeology) respectively. The project also had an Advisory Committee, which has subsequently evolved into an Editorial Advisory Board.

Rusbridge himself, the programme leader for JISC, saw it as one of the great success stories of the Electronic Libraries programme: 'More successes? From the Electronic Journals area, a journal which aimed from the start to get value from the technology rather than just using it to carry images of printed pages – this was Internet Archaeology' (2001).

Reviewing Issue 1, Costis Dallas (1997) described it as 'an intriguing glimpse of the potential of electronic media for scholarly publication'. The *New Scientist* noted that 'for anyone studying or working (or wishing to publish) in the subject, this is an important online resource'. Writing in the *Times Higher Educational Supplement* in November 2000, Professor Steve Mithen described *Internet Archaeology* as:

> a flagship e-journal . . . providing a fine balance between the more creative use of new technologies and traditional publishing formats. The editors of *Internet Archaeology* are making an invaluable contribution to the discipline, one that goes far beyond the provision of their journal alone.

To a large extent 'publication of data' has always been part of the *Internet Archaeology* publication model. From the outset, the journal has endeavoured to promote links between the traditional outputs of research and supporting datasets. The award-winning Linking Electronic Archives and Publications (LEAP) project set out explicitly to provide a series of exemplars of linked publications in *Internet Archaeology* with archives held by the ADS, including the projects of Merv, Silchester, Troodos and Whittlewood (Richards *et al.*, 2011). Of course, this relationship is not exclusive and *Internet Archaeology* has also published articles linked to data sets held in other data archives, including tDAR in the United States (Holmberg 2010).

A more open archaeology and the dissemination of increasing quantities of data necessitates the development of new techniques and tools to deal with the proper referencing and citation of digital resources; indeed, without this there is a very real possibility of becoming 'lost in information' (Huggett, 2012). At the same time, a common concern among data creators is the lack of accreditation for data. Both concerns could be addressed through improved citation. Traditionally, digital resources have utilized the URL to reference digital resources; however, the durability of this method of citation has begun to be questioned (Jeffrey, 2012). A number of schemes have attempted to address this issue; one of these is the DOI system which 'allows collections of data or individual data files to be allocated a URL that will not change irrespective of changes to the physical location of the files in question' (Jeffrey, 2012, 564).

The 'minting' and subsequent management of DOIs is handled by a conglomerate of organisations, working as part of the International DOI Foundation, who guarantee the sustainability of the citation system (Datacite, n.d.). As an adopter of the DOI system, the ADS creates persistent identifiers that consistently and accurately reference digital objects and collections. This serves to address one of the principal concerns of the PUNS report (Jones *et al.*, 2001) by formalizing associations between digital resources and printed outputs. An important outcome of the DOI system is that it also allows citations to be tracked, meaning that data creators, users and repositories can track the use and impact of specific data sets or publications (Hole, 2012).

In 2013, *Internet Archaeology* introduced another publication model to encourage researchers to provide access to their data sets: the data paper. The concept of the data paper was developed in the physical sciences and has been extended to archaeology via the new *Journal of Open Archaeological Data*, established at University College London under the auspices of Ubiquity Press (JOAD, n.d.). A data paper is generally a short paper which simply describes and summarizes a research data set and outlines how it might be re-used. It is generally a condition of publication that the dataset must have been deposited in an archive and have been allocated a Digital Object Identifier.

Thus, for example, a paper by Bevan and Conolly on the Antikythera survey project (2012) references a dataset held by the ADS (2014). *Internet Archaeology* has developed the concept of the data paper further, adding a published review of the dataset, by a named external reviewer (e.g. Williams *et al.*, 2014). In conjunction with the ADS, *Internet Archaeology* has also introduced an annual digital data reuse award to encourage archaeologists to undertake their research in the digital archive, rather than expensively destroying more primary data in the field or laboratory (ADS, 2014).

When *Internet Archaeology* was founded, the Open Access movement was in its infancy, but from the outset it was what would now be regarded as an Open Access journal, supported by grant funding from the JISC. Registration was required, but no charge was made for access. By July 1997, over 3,400 readers had already registered to use the journal, and this was growing at the rate of 270–470 a month (Heyworth *et al.*, 1997, 1041).

At that stage, 39 per cent of the readership came from the UK, 37 per cent from North America, just under 15 per cent from the rest of Europe and about 7 per cent from the rest of the world. By the time the journal ceased requiring readers to register, the number of registered readers had grown to over 25,000 (Richards, 2006, 218).

With the ending of eLib funding in July 2001, *Internet Archaeology* was encouraged to adopt what was then the standard journal business model, based upon subscription payments. Although the preference of the Editor and Co-directors would have been to remain free at the point of use, the journal was costly to produce and demanded the work of a full-time editor and part-time copy-editor to prepare articles for online publication. The view of JISC at that time was that the success of a traditional subscription model would demonstrate the viability of ejournals more widely. *Internet Archaeology* therefore sought to establish itself as a sustainable electronic journal, with a business plan based upon several income streams, comprising publication subventions, advertising (or 'sponsorship') and journal subscriptions, including pay-per-article and pay-per-volume options. By 2006, it had an established subscription base of over 40 UK universities, and an equivalent and growing number of overseas subscriptions, but was still struggling to break even.

By this stage, the academic publishing climate was also beginning to change and JISC was supporting site licences to make e-content freely available to Higher and Further Education Institutions. In 2006, we agreed with JISC Collections that we would put forward a three-year plan, renewable for a further three years to continue the transition to Open Access, dependent upon performance over the first period. This comprised immediate purchase of the journal backfile for 1996–2006, making ten years of rich multimedia scholarly content immediately available to the entire UK Higher and Further Education (HE/FE) domains, followed by the introduction of Open Access to issues 22–25, published from 2007 to 2008.

The agreement involved a declining subsidy, while the journal generated a growing proportion of its revenue from publication subventions from research councils, commercial developers and state funding archaeological agencies. The aim was to allow free access to the growing number of universities that teach Archaeology, but also to encourage use in the FE post-16 education sector, as well as extending usage of the journal to a much wider range of disciplines (History, History of Art, Creative and Performing Arts, Geography, Biological and Earth Sciences) and making it freely available for its technological and methodological interest to those in Information Science and Librarianship disciplines.

The new model also suited public sponsors of archaeological research, such as Historic England and Historic Environment Scotland, which had traditionally subsidized the publication of the research results, and the journal negotiated subvention deals to publish the results of some key excavations conducted within the UK over the last ten years. Similarly, in the academic sector, the Arts and Humanities Research Council (AHRC) was promoting

electronic dissemination, and the journal was successful in winning one of the AHRC ICT strategy grants, for the LEAP project (see above).

From 2007, *Internet Archaeology* was enabled by JISC to become an Open Access journal for UK HE/FE, although it continued to charge a subscription to overseas institutions and individuals and organisations not based within UK HE/FE. By 2008, 89 institutions had signed up to the JISC licence, a 300 per cent increase on the 29 UK HE/FE subscribing institutions in 2006. From 2010, we began to publish specific articles for which global Open Access was provided via APCs. The proportion of this content has gradually increased and in 2014 the tipping point was reached. In October 2014, *Internet Archaeology* announced that it was to become a full Gold open access journal. The hybrid phase was concluded and subscriptions will no longer be charged for access to any past and future content. Several things spurred this decision.

Over the preceding years, the journal made active efforts in this direction, by switching to a default CC-BY licence, by opening up back issues with an annual rolling wall and by adjusting subscription charges accordingly. During this time, there has also been a marked increase in quality, funded submissions, including several themed issues. By the start of 2014, over 50 per cent of articles published in the journal were open access, so it was becoming difficult to justify charging for a declining proportion of the journal's content. Other hybrid journals will face a similar dilemma, as the proportion of OA content increases. For commercial publishers, this will be particularly challenging as in archaeology many may have set the level of their APCs as a 'loss leader', at what they think the market will bear, rather than what they would need to recoup to cover all their publication costs, should all their content be OA.

There is no doubt that being fully OA increases usage of journal content. It also facilitates an international readership. In 2016, there were over 87,000 visits to the journal and over 225,000 unique page views. Over that period, 64.4 per cent of readers were based in Europe (including 42.2 per cent in the UK), 23.2 per cent in North America, 6.2 per cent in China, 3.2 per cent in Oceania, 1.3 per cent in South America and 1.2 per cent in Africa. The move to Gold OA has also stimulated a significant increase in article submissions, demonstrating that many authors prefer a full OA route. Many – whether in the UK or overseas – appear to have access to funding to cover the cost of APCs, although the issue of whether to cross-subsidize the publication of papers by those who have no access to APC funding is a difficult one. Currently, we have set up a donations fund to help raise funds to cover their APC costs.

In conclusion, *Internet Archaeology* has always tried to be more than 'just a journal'. It explores the possibilities of the web and has delved into many different publication formats. It has often felt as if we have been 'before our time' and that our publication model would not survive in the current harsh publication climate. Nonetheless, being a small operation has meant

that we could be responsive to changes in the wider scholarly landscape and adapt to fast-changing times. It will be interesting to see what the next twenty years bring.

References

ADS 2014. *Digital Data Re-use Award*. Retrieved 18 December 2014 from http://archaeologydataservice.ac.uk/blog/2014/10/digital-data-re-use-award/#more-1641.

Atici, L., Whitcher Kansa, S., Lev-Tov, J. and Kansa, E. C. 2012. Other people's data: A demonstration of the imperative of publishing primary data, *Journal of Archaeological Method and Theory 4*, 149–67.

Bevan, A. and Conolly, J. 2012. Intensive survey data from Antikythera, Greece, *Journal of Open Archaeology Data* 1(1), doi:10.5334/4f3bcb3f7f21d.

Bevan, A. and Conolly, J. 2014. *The Antikythera Survey Project* (data-set). York: Archaeology Data Service (distributor) doi:10.5284/1024569.

Condron, F., Richards, J., Robinson, D. and Wise, A. 1999. *Strategies for Digital Data: Findings and Recommendations from Digital Data in Archaeology: A Survey of User Needs*. Retrieved 18 December 2014 from http://ads.ahds.ac.uk/project/strategies.

Costa, S., Beck, A., Bevan, A. and Ogden, J. 2013. Defining and advocating open data in archaeology, in G. Earl , T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska and D. Wheatley (eds), *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology* (CAA), Southampton 26–29 March 2012. Amsterdam: University Press, 449–56.

Dallas, C. 1997. A step beyond reading in archaeological publication, *Archives & Museum Informatics* 11, 55–64.

Datacite. n.d. Retrieved 18 December 2014 from www.datacite.org.

Day, M. 1999. *The Scholarly Journal in Transition and the PubMed Central Proposal,* Ariadne *21*. Retrieved 18 December 2014 from www.ariadne.ac.uk/issue21/pubmed.

Finch, J. 2012. *Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications*. Report of the Working Group on Expanding Access to Published Research Findings.

Ginsparg, P. 1994. First steps towards electronic research communication, *Computers in Physics* 8, 390–96.

Harnad, S. 2001. The self-archiving initiative: Freeing the refereed research literature online, *Nature* 410, 1024–5.

Harnad, S. 2005. *Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold,* Ariadne *42*. Retrieved 18 December 2014 from www.ariadne.ac.uk/issue42/harnad.

Harnad, S. and Hemus, M. 1997. All or none: No stable hybrid or half-way solutions for launching the learned periodical literature into the post-Gutenberg galaxy, in I. Butterworth (ed.), *The Impact of Electronic Publishing on the Academic Community*. London: Portland Press, 18–27.

Harnad, S. and Hey, J. 1995. Esoteric knowledge: The scholar and scholarly publishing on the Net, in L. Dempsey, D. Law and I. Mowat (eds), *Networking and the*

Future of Libraries 2: Managing the Intellectual Record. London: Library Association Publishing, 110–16.

Heyworth, M.P., Ross, S. and Richards, J. D. 1996. *Internet Archaeology*: An international electronic journal for archaeology in H. Kamermans and K. Fennema (eds), *Interfacing the Past: Computer Applications and Quantitative Methods in Archaeology CAA95*. Leiden: University of Leiden, Analecta Praehistorica Leidensia 28, 517–23.

Heyworth, M.P., Garside-Neville, S., Richards, J.D. and Vince, A.G. 1997. *Internet Archaeology*: A quality electronic journal, *Antiquity* 71, 1039–42.

Higher Education Funding Council for England (HEFCE) 2014. *Policy for Open Access in the Post-2014 Research Excellence Framework*. Retrieved 18 December 2014 www.hefce.ac.uk/media/hefce/content/pubs/2014/201407/HEFCE2014_07.pdf.

Hole, B. 2012. A call for open scholarship in archaeology, in C. Bonacchi (ed.), *Archaeologists and the Digital: Towards Strategies of Engagement*. London: Archetype, 114–26.

Holmberg, K. 2010. Placing immateriality: Situating the material of highland Chiriquí, *Internet Archaeology 28*. http://dx.doi.org/10.11141/ia.28.5.

Huggett, J. 2012. Lost in information? Ways of knowing and modes of representation in e-archaeology, *World Archaeology* 44, 538–52.

Jeffrey, S., 2012. A new Digital Dark Age? Collaborative web tools, social media and long-term preservation, *World Archaeology* 44, 553–70.

JOAD. n.d. *Journal of Open Archaeology Data*. Retrieved 18 December 2014 from http://openarchaeologydata.metajnl.com.

Jones, S., MacSween, A., Jeffrey, S., Morris, R. and Heyworth, M. 2001. *From the Ground Up: The publication of archaeological projects: A user needs survey*. York: Council for British Archaeology.

Kansa, E. 2012. Openness and archaeology's information ecosystem, *World Archaeology 44*, 498–520.

Kansa, E. C. and Whitcher Kansa, S. 2011. Towards a do-it-yourself cyberinfrastructure: Open data, incentives, and reducing costs and complexities of data sharing, in E. C. Kansa, S. Whitcher Kansa and E. Watrall (eds), *Archaeology 2.0: New Approaches to Communication and Collaboration*. Los Angeles, CA: The Cotsen Institute of Archaeology, 57–91.

Kansa, E. C. and Whitcher Kansa, S. 2013. Additional thoughts on sustaining and promoting open data in archaeology, *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 1, 102–3.

Kansa, E. C., Whitcher Kansa, S., Burton, M. M. and Stankowski, C. 2010. Googling the grey: Open data, web services, and semnatics, *Archaeologies: Journal of the World Archaeological Congress 6*, 301–26.

Lake, M. 2012. Open archaeology, *World Archaeology* 44, 471–8.

Lancaster, F.W. 1978. Whither libraries? Or, wither libraries, *College and Research Libraries* 39, 345–57.

Porter, B. W. 2013. Sharing data is hard! But worth it, *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 1, 100–1.

Pratt, D. 2013. Not an either/or proposition: Combining interpretive and data publication, *Journal of Eastern Mediterranean Archaeology and Heritage Studies 1*, 101–2.

Richards, J. D. 2006. Electronic publication in archaeology, in T.L. Evans and P. Daly (eds) *Digital Archaeology: Bridging Method and Theory*. Oxford and New York: Routledge, 213–25.

Richards, J. D., Charno, M. and Winters, J. 2011. Making the LEAP: Linking Electronic Archives and Publication, in J. E. Jerem, F. Redo and V. Szeverényi (eds), *On the Road to Reconstructing the Past: Proceedings of the 36th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Budapest, Hungary, 2008*. Budapest: Archaeolingua, 141–6.

Rusbridge, C. 2001. After eLib, *Ariadne* 26. www.ariadne.ac.uk/issue26/chris. Retrieved 18 December 2014.

Tyers, P. 1996. Roman amphoras in Britain, *Internet Archaeology* 1. http://dx.doi.org/10.11141/ia.1.6.

UK Government 2011. *Science and Technology Committee – Eighth Report. Peer Review in Scientific Publications*. Retrieved 18 December 2014 from www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/85602.htm.

Williams, A. N., Ulm, S., Smith, M. and Reid, J. 2014. AustArch: A Database of ^{14}C and Non-^{14}C Ages from Archaeological Sites in Australia – Composition, Compilation and Review (data paper). *Internet Archaeology* 36. http://dx.doi.org/10.11141/ia.36.6.

4 Crowds for clouds: recent trends in humanities research infrastructures

Tobias Blanke, Conny Kristel and Laurent Romary

Introduction

The humanities community has been remarkably successful in applying for research infrastructure funding in Europe. For instance, in the first call for the Horizon 2020 programme on Integrating Activities 'to open up key national and regional research infrastructures to all European researchers . . .' (Commission, 2014), from over 60 proposals 10 were finally chosen for the first round of funding. Out of these selected proposals, two proposals came directly from humanities communities, which makes them one of the most successful communities in this call. Humanities seemed to have convincingly argued that they need transnational research opportunities and through the digital transformation of their disciplines also have the means to proceed with it on a previously unknown scale, leveraging facilities and researchers across Europe.

The digital transformation of research and its resources means that many of the artefacts, documents, materials, etc. that interest humanities research can now be combined in new and innovative ways. A recent project in the UK links documents about court cases in the Old Bailey system with the fate of deportees in Australia (www.digitalpanopticon.org/). Such projects are emerging because there is a need in the corresponding communities to expand their research environments. Digital transformations have brought about new actors and practices in all areas of researching culture. More and more cultural objects are integrated into the digital space through processes of datafication (Kitchin, 2014; Mayer-Schönberger and Cukier, 2013), while infrastructures working with these digital objects are now better able to foster a sense of stability and continuity (Edwards *et al.*, 2009).

At the same time, digital transformations offer new possibilities for humanities research to reassemble new and emerging socio-technical methods and devices (Ruppert *et al.*, 2013) in order to explore society and culture. Due to these digital transformations, (big) data and information have become central to the study of culture and society. Big data is not limited to the sciences and to large-scale enterprises. With more

than 7 billion people worldwide, large amounts of data are produced in social and cultural interactions, while we can look back on several thousand years of human history that have produced vast amounts of cultural records.

Humanities research infrastructures manage, organize and distribute this kind of information and many more data objects as they become relevant for social and cultural research. Edwards has explored infrastructures as global socio-technical systems and as characteristics of modern society, where one lives within and by means of infrastructures (Edwards *et al.*, 2009). Research infrastructures, in particular, helped disciplines to redefine themselves around a shared set of devices that support their research. Humanities research infrastructures have been theorized as digital ecosystems without a centre and constituted through heavily interconnected online platforms (Anderson and Blanke, 2012).

Along these lines, the European Commission defines research infrastructures as 'facilities, resources or services of a unique nature that have been identified by research communities to conduct top level activities in their fields. They may be single sited, distributed or virtual' (ESFRI, 2010). They 'often produce large amounts of data requiring data management'. In the case of humanities research infrastructures, much of the 'data' for integration are not a product of the infrastructure itself but are the primary source materials, produced as a result of the activities of cultural heritage institutions, mostly in archives and libraries. Large-scale digitization efforts have recently begun to create digital surrogates for human history. The European Union seems to be especially committed to digitize and present its cultural heritage online. For example, by the end of 2014, its Cultural Heritage aggregator, Europeana, made available over 30 million digital objects through its portal (Bernipe, 2014).

In this chapter, we concentrate on recent trends in humanities research infrastructures. We observe the common practices that have emerged in various large-scale transnationally operating infrastructure projects. We focus on research infrastructures rather than on (digital) library and archive integration projects such as Europeana, because research infrastructures share the overall final aim to action research. Europeana, on the other hand, primarily aims to fulfil the needs of a culturally interested public rather than the needs of a specific research community.

Hidden in this broad distinction are, of course, many commonalities between these initiatives. They all make use of similar technologies for integrating digital collections, and the primary user community of libraries (and in particular archives) has been humanities researchers. Finally, they are concerned with ongoing access to cultural collections. Nevertheless, the differences between these initiatives are defined enough to merit a focus on research infrastructures, as a distinct undertaking in the humanities.

Humanities research infrastructures

The link between humanities and their infrastructures has been strong in the past although sometimes hidden. Some, like the British historian Marina Warner, have even defined humanities themselves 'as infrastructure' for critical thinking about culture and how 'people connect' (Preston, 2015). Furthermore, literary scholars and philosophers are regulars in research libraries, while historians lead the development and maintenance of many archives.

The UK's *Times Higher Education* summarizes the view that the humanities too need firm foundations and a new culture of sharing, both of which are signature elements of research infrastructures (Reisz, 2014). There have been several attempts to define what should interest the humanities with regard to infrastructures. In order to discriminate humanities infrastructures from scientific ones, the digital humanities scholar Patrik Svensson famously asked for a 'conceptual cyberinfrastructure' that can be 'seen as a set of underlying ideas that provide the ideational grounding of a particular instance of research infrastructure' (Svensson, 2011). Svensson wants a humanities infrastructure to be first and foremost about humanities, even though he clearly sees an interest from those in the sciences in humanities research infrastructures: 'Dan Atkins, then head of the US NSF Office of Cyberinfrastructure, [demanded] that the humanities and social sciences step up and show leadership in relation to the issue of future cyberinfrastructure' (Svensson, 2011). While the humanities have not managed to sustain larger projects of integrated infrastructures in the US, the Europeans have undertaken various successful projects.

A couple of years ago, Manfred Thaller from Cologne brought together a number of digital humanities researchers for a seminar on the current state of affairs in the digital humanities (Thaller, 2012). The controversial discussions included an exchange on the nature of research infrastructures and the relationship of digital humanities to them. The Dutch digital humanities scholar Joris van Zundert asked: 'If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities' (Van Zundert, 2012). He held up his own experience in the international network InterEdition against three examples of 'big all-encompassing all-serving digital infrastructures', which he considered meaningless for the development of humanities. He cited Bamboo in the US as well as the European initiatives DARIAH (which brings together digital humanities initiatives in Europe) and CLARIN (which develops language resources).

Bamboo's funding has in the meantime run out (Dombrowski, 2014). We thus focus on DARIAH (www.dariah.eu) and CLARIN (www.clarin.eu/) and we will also cover their related more domain-specific infrastructures such as ARIADNE (www.ariadne-infrastructure.eu/), which focuses on archaeology; CENDARI (www.cendari.eu/) working on medieval and First World War resources; EHRI (www.ehri-project.eu), concerned with Holocaust

research; and finally, IPERION (www.iperionch.eu), which is about material research on cultural heritage. The latter four projects are all associated with DARIAH, which has through them become a platform for transnational humanities collaborations, while CLARIN has continued to have a transforming impact on language research in Europe.

While van Zundert's criticism was not based on the above associated infrastructures, it could have also applied to them. ARIADNE follows the assumption that archaeologists have a need to integrate their research data, while van Zundert seems to imply (Van Zundert, 2012) that digitized humanities resources are available through libraries and in particular their own large-scale collaborations like Europeana. EHRI, on the other hand, was focused, in its first phase (from 2010 to 2015) to integrate archival material on the Holocaust from relevant sources across Europe. Thus, the project made a strong effort to integrate three overlapping but not always collaborating communities: history, digital humanities and archives. The direct work with archivists in Holocaust institutions was considered especially important, because most of the material on Holocaust is not yet digitally available. The Dutch NIOD, the coordinator of both EHRI and a well-funded Dutch national archive on Holocaust and genocides, has archives of approximately 2.5 kilometres long, which is a considerable amount of data. Only 2 per cent of it, however, is even available in a digital format, although it is not always accessible online. By 2016, they aim to increase the amount digitized to 7 per cent. Any kind of research infrastructure project based on data that aims to link across archives therefore needs to find ways of linking both analogue and digital information.

The final two domain infrastructures IPERION and CENDARI concentrate on addressing what van Zundert criticizes as 'snowballing IT-based methodological innovation into a humanities domain' (Van Zundert, 2012). IPERION organizes access to humanities research to support the preservation of material cultural heritage. The use of IT-based instruments to produce 3D visualizations of artworks and other heritage objects has enhanced the capacities of curators. IPERION brings their knowledge together. CENDARI has the unusual task by the Commission to develop tools for medieval and First World War historians. It has already demonstrated that while these historical communities are distinct, they also share a needs for a methodological commons (Anderson *et al.*, 2010) for researching archives.

Van Zundert's main concern, however, is, of course, CLARIN and DARIAH and especially CLARIN, which was further advanced at the time of the workshop in Cologne. He stated: 'Being standards-driven, institutionally bound, and at worst enforcing specific implementations, they are platforms of exclusiveness' (Van Zundert, 2012). Against them, he demands 'open platforms' based on 'agile processes'. In a contribution to the same workshop, Anderson and Blanke (2012) promoted an idea of open platforms based on agile processes they named 'digital ecosystems'. The concept of ecosystems was belittled in the past, but in our experience it is now more

widely accepted and is typically used to describe a loosely coupled organization of services and activities.

Blanke (2014) analysed how the concept of digital ecosystems encapsulates decentralized digital work, precisely because of its contested origin in biological sciences. According to Briscoe and Sadedin (2009), a natural environment consists of ecosystems, which in turn are inhabited by habitats and communities. It is easy to form an analogy here, so that populations described in ecological theory are analogous to crowds forming on the Internet, or to the collaboration of large numbers of humans on a common task. The habitats are the platforms or the 'clouds' crowds work on. Together, communities and habitats build niches or, in our sense, applications and services that are built around them. Mark Zuckerberg (Blanke, 2014) has pinned his hope on digital ecosystems, because the smaller Facebook can only compete with the giants of Google and Microsoft if it manages to organize an ecosystem, with which it can effectively integrate outside innovation.

According to the IEEE Digital Ecosystem conference, digital ecosystems are 'loosely coupled, domain-specific . . . communities which offer cost-effective digital services and value-creating activities' (IEEE Digital Ecosystem, 2007). In this definition, digital ecosystems are derived from communities, or crowds with a set purpose, rather than mere technologies. For example, consider the Facebook case, where Mark Zuckerberg (Blanke, 2014) first associated developers and users with digital ecosystems, because in the digital ecosystem they define their services and take control. Therefore, the technologies of the digital ecosystem need to be thought of from the perspective of the crowds and not vice versa; they do not define what crowds are. This requires a commitment to an open platform, which can only be developed using open standards.

To think of research infrastructures as digital ecosystems, we should think of them as services that are built around communities. Infrastructures are then the sum and integration of these services that are shared through a platform. Communities as crowds work together on a common goal and become the most important resource for the sustainability of the infrastructure. The arts and humanities research communities are characterized by high diversity as well as by limited overall funding opportunities; no one community can achieve a meaningful investment over a sustained period of time. Software is commonly co-developed with open-source licences, while code is shared so that any improvement goes directly back into the software. The knowledge in existing software is reverse-engineered and made future-proof.

European humanities research infrastructures as digital ecosystems

In this section, we will investigate how, for the above-cited six European larger humanities research infrastructures, a digital ecosystem is emerging as the

collaboration between crowds and clouds. We begin with CLARIN, which is concentrated on the particular community using language resources. Many of its partners have worked together for decades. CLARIN is organized around Centres, which provide services to each other (www.clarin.eu/clarin-eric-datatables/centres). These are mature and certified through a formal procedure, registered in a central database (https://centres.clarin.eu/) and compliant with a range of requirements in order to be allowed into the CLARIN network.

With its Centre structure, CLARIN is the most formally structured infrastructure of the ones discussed here. It has managed to bring together many parts of the European community that work with language resources. CLARIN's key paper (Váradi *et al.*) has mainly been cited by the language resources community according to Google Scholar.[1] CLARIN delivers a central repository of language resources – a cloud, so to say – together with basic technologies to manage this cloud. This way, CLARIN fulfils one of van Zundert's central demands: 'We all have email, why not let us all have access to an academic computing cloud?' (Van Zundert, 2012: 18). CLARIN's strength is clearly its concentration on a cloud of language resources that serve researcher crowds with specific needs.

If CLARIN's strength is its depth, then DARIAH's is its breadth (Blanke *et al.*, 2011). DARIAH's main objective is to organize national initiatives in digital humanities in its various member states, which are all more or less independent. DARIAH's second aim is to develop a platform (cloud) for other transnational European digital arts and humanities initiatives, some of which we cited above and all of which are independent entities. Its central form of organization is the working group (https://dariah.eu/library/resources.html), where researchers come together to collaboratively develop activities and services. The working group model is clearly copied from successful Web 2.0 architectures that often succeed in realizing a collective platform by working together in special interest communities (Blanke *et al.*, 2011).

DARIAH working groups (WG) are based on a three-step approach:

1 Conceptualization: Members of DARIAH develop a concept for the WG, aggregate participants and means as well as define an action plan. The primary outcome is a proposal to the DARIAH community that meets the WG requirements.
2 Implementation: At this stage, efforts are integrated and the connection to further communities outside DARIAH is established.
3 Service: Finally, the service is enabled with guaranteed hosting and sustainable funding through DARIAH. The service is published to the community.

There are no specific criteria for working groups, but they need to be collective efforts and need to comply with the DARIAH mission (http://dariah.eu/about/mission.html). Current working groups include visual media,

education, a Digital Humanities course registry, digital annotations, service registries, a digital methods observatory, etc.

DARIAH's organization is the most bottom-up we could find. While CLARIN's strength is its organization across many stable centres, DARIAH is organized around distinct collaborations of its crowds. This makes the construction less stable but allows for a stronger dynamic and better flexibility to adjust to new research areas. CLARIN and DARIAH can implement these structures, as they are stable membership organizations within the European Strategy Forum on Research Infrastructures framework (ESFRI, 2010). The rest of the observed initiatives are in comparison traditional projects, funded mainly and exclusively by the European Commission. Nevertheless, these projects have shown remarkable innovations of how services and their platforms can be developed around communities.

CENDARI and EHRI both work with communities that are engaged in archival research. Both assemble historians, archivists and specialists in the Digital Humanities to make archival documents (which were often collected under difficult circumstances) and ensure they are available and accessible for research, and to develop innovative techniques and methodologies for the analysis and interpretation of such documents. Both projects concentrate in particular on those archives that are not part of larger infrastructures and/or are 'hidden' to most researchers. CENDARI has a stronger technical focus with more work dedicated to developing Digital Humanities solutions (Gartner and Hedges, 2013), while EHRI concentrates mainly on archival integration (Speck *et al.*, 2014). CENDARI collaborates with large clusters of special collections in The European Library. EHRI's focus is on individual archives. Both, however, share the commitment that exploration of archives in the digital age is only possible if historians form collaboration crowds with archivists, and these crowds should include digital experts. This way, the archive becomes a new research space for historical discovery. CENDARI's corresponding aim is thus a discovery space, which can be considered as its cloud (www.cendari.eu/description-of-work/).

Both projects develop a virtual observatory for research collections as a key component of their work. EHRI has already released its 'virtual observatory' for Holocaust collections (Blanke and Kristel, 2013), which is a cloud of collection descriptions. The observatory is based on an integration of descriptions from partner sites and new descriptions arising from EHRI's own investigation work. The main objective was to allow researchers to retrieve information about archival sources that pertain to a particular research theme across repositories such as 'find all information about the departure of Dutch prisoners to the Terezín Ghetto'.

Through the EHRI Virtual Observatory (https://portal.ehri-project.eu/), researchers are now able to query disparate archives. To this end, a wide range of different types of institutions holding relevant archival material was identified, including national and regional archives, memory institutions, museums, and local and private collection holders.

Each type of archive poses a unique challenge in terms of integrating their material, harmonizing the metadata and publishing it in an integrated portal. As new collections and other material are continuously discovered, the identification and integration work needs to continue all the time. EHRI thereby also sets an example for other domains engaged in historical research, showing how 'big data' on human history can be developed. Both EHRI and CENDARI thus develop a research environment that will allow their respective communities to cope with the big research data they are faced with.

The technical design challenge (Blanke *et al.*, 2013) for archival research infrastructures such as EHRI and CENDARI is to innovate a dynamic, research-driven collection cloud, where new material is permanently discovered, added and analysed. Their work has to rethink some of our assumptions that were based on traditional research with relatively stable cultural heritage collections. Instead, they must develop an environment that is technically flexible enough to allow for the integration of heterogeneous material and that is social enough to allow researcher crowds to discover and analyse their material, and make new connections.

IPERION works with curators and other researchers concerned with the material aspects of heritage, while ARIADNE develops devices for archaeologists. Both share their commitment to developing data infrastructures that go beyond CENDARI's and EHRI's connection to archives and concentrate on the full sharing of research data sets. Both projects are committed to long-term preservation and dissemination. Contrary to earlier cited community concerns about standards as limiting research freedom, ARIADNE standards enable the free exchange between crowds. For ARIADNE, '[t]he main processes for making data understandable and shareable are standardization and registration' (Aloia *et al.*, 2014). Standards have to be part of any research data cloud. The standardization and sharing of data sets cannot simply be handed over from the communities to professional providers such as libraries. While they need to work together, for ARIADNE there remains a distinct need for subject-based repositories. As Julian Richards (one of the leaders of the ARIADNE project), has put it, few of the libraries are currently equipped with the knowledge or the platform to deal with complex data sets that incorporate many discipline-specific assumptions. 'Several studies have recognised the value of discipline-based repositories in developing stakeholder communities, avoiding fragmentation, and establishing discipline-specific data preservation expertise' (Richards, 2012).

Working with complex data sets that are developed by highly diversified research projects to investigate cultural objects poses particular challenges to the 'conceptual' integration of data in clouds. For researchers, the grand promises and challenges with regard to the integration of data sets lie in their 'linkability'. This is independent whether the data is collected in small independent projects, using unorganized spreadsheets and word documents, or in large joint projects such as IPERION and ARIADNE.

All are research databases and contain interpretations and expressions of uncertainty. We simply do not know enough from the past to exactly determine when Queen X died or Volcano Y erupted. These datasets are part of what Buneman from the Digital Curation Centre has called 'curated databases' (Buneman *et al.*, 2008), as they are human created and therefore full of inconsistencies.

ARIADNE works on a Linked Data cloud that conceptually integrates archaeological curated databases (Aloia *et al.*, 2014), which will realize the promises of linkability. A number of earlier experiments at King's College London with integrating traditional humanistic data sets (Blanke *et al.*, 2012) tried to find out whether with a Linked Data cloud, uncertainties can be reduced by linking the information in one data source with the information in another. The results were mixed. While the production of links is already an issue because of the heterogeneous nature of curated databases, it can be even more difficult to define how to sensibly consume information that is highly interlinked. Links are good but also confusing for human researchers. A possible solution will be a community effort like ARIADNE that takes into account the production as well as consumption needs effort.

In this section, we have discussed how all the humanities infrastructures attempt to develop service and activities clouds around communities. For CLARIN, this has meant to concentrate on dedicated centres as a core stable form of organization. DARIAH is an umbrella organization that develops its platform collectively in working groups. The four other projects we discussed share a concern for the specifics of humanities data whether it is held externally from the communities in archives or is produced directly by the community. Those concentrated on archives are mainly concerned with how to translate the cultural heritage institutions' holdings into something that fits into the needs of digital research. Those projects concerned with data sets that are created by researchers themselves (or at least for them) need to overcome the limitations of research data sets, as they are common not just to the humanities, but many other research disciplines. For humanities, this means a focus on concepts and Linked Data clouds as supporting integration and overcoming heterogeneity that are the result of how these data sets are produced and curated by research crowds.

Enhancing and strengthening the crowd

All our discussed examples demonstrate how important it is for humanities infrastructures to create crowds with their data. As computers fail with the complexity of knowledge in curated databases of culture, crowds become more important. Research infrastructures need to connect human brains to perform complex reasoning on data. 'Networks connect people as well as devices, and when they are cheap and easy to use it means that those intellectual tasks more efficiently performed elsewhere by other people can be broken out and distributed' (Zittrain, 2008).

Following the idea of ecosystems, crowds stand next to clouds as equal components of an infrastructure (Blanke, 2014). This idea goes back to an early conceptualization of Amazon to use its crowds to perform tasks where its clouds were unable to help. It set up its Mechanical Turk system to achieve this. Taking the Amazon infrastructure view (Blanke, 2014), crowds collaborate in the Mechanical Turk platform as part of an emerging larger infrastructure that supports the new kinds of production and consumption of digital value. Amazon has chosen to offer its crowd-sourcing functionalities through the same interface by which its other services are accessible. The substitution of computer intelligence by human intelligence is hidden from the outside world. If the crowds work smoothly, the service appears to be as seamless as a computer service. 'Hidden away under the appearance of computer-generated work, crowds have been increasingly rendered visible only through the design of new infrastructures' (Aradau and Blanke, 2013: 38).

Of course, in the Amazon view and its Mechanical Turk implementation, crowds are paid to contribute for often not very satisfying tasks. Amazon's crowd is not based on volunteer contributions, but each participant gets paid a small amount for each task completed. In Amazon's terminology, these are Human Intelligence Tasks (HITs), and requesters define tasks and upload data, while workers (aka Turkers 'Turkers', which is the name Amazon Mechanical Turk workers are known by) complete tasks and get paid for doing so. Typical tasks include the identification of email addresses in texts or the labelling of images. Workers' rights do not seem to be available to the Turkers. There is, for instance, no guarantee of payment after the job and there are no benefits. A hall of shame of worst jobs offered to Turkers can be found on http://turkernation.com.

For non-commercial research infrastructures' interest in crowds, financial benefits cannot be an option. Malone *et al.* (2009) have mapped the 'genome of collective intelligence' from crowds. They define collective intelligence as 'groups of individuals doing things collectively that seem intelligent' (Malone *et al.*, 2009: 2) and give three reasons why people would like to collaborate to appear intelligent: money; love or enjoyment of an activity; and glory when recognition is achieved among peers. All these are opposed to a hierarchical distribution of labour. In the following section, we will investigate how the future of research infrastructure development in the humanities might be inspired by these crowd-based collective intelligence initiatives.

Humanities need to recognize that its infrastructure needs are specific and realize that humanities scholars need to take computing innovations seriously, as long as computers are as they are, and as long as the funding for collections in the humanities will be less than funding for the sciences (and this situation does not look like improving). As long as humanities collections are too complex and analogue-based for computers to effectively deal with, much more focused work on crowds is needed, which thus require — according to the ecosystem model — a recognized form

of infrastructure. In the past, crowds and humanities were concentrated on enriching existing heritage collections (Holley, 2010), which can be a successful model to help cash-strapped heritage and humanities communities to add value to their data. However, there are other innovations in the larger crowd economy that we should recognize as particularly suited for addressing humanities computing challenges.

We have already discussed earlier that such challenges generally include a focus on working with data. Because in the humanities this data comes in many different formats, has complex semantic relationships and generally does not comply with the needs of algorithms, there is detailed work involved in preparing the data. OCRing (Optical Character Recognition), for instance, nowadays works well with standard print collections of newspapers, but it continues to struggle with hand-written documents, and this is a problem as they are common to many historical recordings. Therefore, many heritage institutions have begun to set up large-scale crowd-sourcing projects to transcribe their collections (Holley, 2010). One of the most successful heritage crowd-sourcing projects was the digitization of the Australian newspaper archives. In 2007, the National Library of Australia began to digitize out-of-copyright newspapers. It used crowds to help correct OCR mistakes, and the public followed in large numbers and analysed millions of lines of text (Holley, 2010). The Digital Humanities have taken up efforts to digitize collections with some standout projects of their own, such as Transcribe Bentham (http://blogs.ucl.ac.uk/transcribe-bentham/) or Diaries of First World War soldiers (www.operationwardiary.org/).

However, we believe we can learn about the infrastructure future in the humanities from other crowd-sourcing innovations as well, because of the specific type of computation Digital Humanities projects entail. We have just discussed the complexities of humanities data that require human intelligence tasks rather than computational analysis. Another typical characteristics of Digital Humanities is that humanities computation is, generally speaking, quite small and well contained. The tasks involved are often similar to each other such as setting up a website but not similar enough for computers to completely take over. Finally, and most importantly, individual organizations in the Digital Humanities are not big enough to have the in-house resources that would enable them to exploit cultural resources sufficiently.

CrowdSPRING (www.crowdSPRING.com) is a typical example of an online marketplace, designed to address the kinds of challenges just discussed. It offers 'creative services' such as logo design, website development, etc. The clientele is anybody from large and small businesses to private people. They choose from the offerings of 'creatives' from over 200 countries, who present examples of their work. The model has been criticized as undermining traditional design work (Hyde, 2008) but has proved to be effective in distributing small well-contained creative tasks. 'Because buyers on crowdSPRING select from actual designs, designers

on crowdSPRING submit work on spec. 'Spec' is a short name for doing any work on a speculative basis . . . (Basecamp, 2008). The advantage for designers is that crowdSPRING offers project management services and militates against payment and legal risks.

Online marketplaces are booming (Laumeister, 2014) by offering effective management and allocation of small tasks for work that can only be done by humans. They would therefore be perfectly suited models for humanities computing tasks. Blanke *et al.* (2011) develop a model for such marketplaces in DARIAH and demonstrate that the challenges involved are social. The marketplace platform has no value to its users unless both those who need to define tasks as well as those willing to offer their work are present. Both user groups need to be present from the beginning to make a successful community. The online marketplace also requires a critical mass of providers and consumers of services to be present at any moment in time. To this end, the marketplace has to make it easy to retrieve and access services. Finally, trust needs to be developed so that the usual risks involved in transactions on the online marketplace can be reduced. In order to cultivate trust, consumers and producers of tasks can often review each other's performances.

Another resource almost completely missing from Digital Humanities organizations is analytical expertise for advanced specialized computing tasks. Kaggle (www.kaggle.com/) is a crowd-based infrastructure specializing in delivering such skills. It aims to develop models to solve advanced data problems and make data analytics a 'sport'. Those in need of data analytics can pay Kaggle to host a competition with their data. Kaggle in turn offers connections to existing analytical talent. Anyone who wants to develop a competition with Kaggle needs to submit their data, with Kaggle offering services to help publish the data on its platform in an effective and ethical way. Participants in a competition then use the data sets, trying their models against them. At the end of the competition, the winner is announced and given the prize. 'Kaggle is a way to organize the brainpower of the world's most talented data scientists and make it accessible to organizations of every size', according to Google's Varian (Reuters, 2011).

Kaggle's crowd-sourcing approach uses the fact that data problems have many possible solutions, while there are also established techniques to find out which models describe the data problem best. But there remains a mismatch, as few organizations that collect data have the in-house analytical skills to also exploit this data. The situation is thus similar to many humanities organizations; they have collections but lack the skills to gain insights from the data. Again, the commercial model of Kaggle would have to be shifted to one based on voluntary contributions to make Kaggle's crowd-sourcing work in the humanities. Nevertheless, both online marketplaces and online competitions offer new exciting opportunities for humanities and should thus be part of research infrastructures. We are sure that the next couple of years will see new projects develop here that will fill this gap.

Conclusion

This chapter has introduced a number of ideas behind major humanities research infrastructure initiatives in Europe. It started off by investigating how they address existing criticisms through a new innovative model that can be described as a loosely coupled ecosystem of services and activities. Humanities research infrastructures have been very successful and led to high-profile projects in the last decade in Europe. They have opened up completely new funding schemes to the involved communities and followed a specific need for transnational research. This has been necessary because of the highly diverse and heterogeneous research landscape of humanities in Europe.

Humanities research infrastructures have successfully developed new models of integrating and developing services around communities. It has become clear during these developments that the infrastructures need to entail next to cloud components that offer a digital platform for research but also for also crowd-based participation. This corresponds to the specific ways of working with data in the humanities and the challenges of the underlying complexities of data, which go beyond what is currently possible with computational means. We have described in this chapter how this interaction of crowds and clouds works for six successful large-scale European initiatives. A major distinction here was whether the research communities have to deal with data produced during their own research processes when analysing cultural objects or data that is offered to them in traditional archives.

Finally, we tried to predict the future of other crowd-sourcing innovations that could enhance digital humanities research practices. Especially promising are crowd-based marketplaces and competitions. Humanities research infrastructures have been so successful because they can answer a particular community need. We will see many more types and structures develop in the next years that will make use of crowd-based innovations.

Note

1 Google Scholar search 14/10/15: https://scholar.google.co.uk/scholar?cites=99035008761397907 36&as_sdt=2005&sciodt=0,5&hl=en

References

Aloia, N., Papatheodorou, C., Gavrilis, D., Debole, F. and Meghini, C. Describing research data: A case study for archaeology. On the move to meaningful internet systems: OTM Conferences, 2014. Springer, 768–775.

Anderson, S. and Blanke, T. 2012. Taking the long view: From e-science humanities to humanities digital ecosystems. *Historical Social Research/Historische Sozialforschung,* 147–164.

Anderson, S., Blanke, T. and Dunn, S. 2010. Methodological commons: Arts and humanities e-Science fundamentals. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 368, 3779–3796.

Aradau, C. and Blanke, T. 2013. The politics of digital crowds. *Lo Squaderno,* 33, 31–38.

Basecamp. 2008. *The NO!SPEC Campaign vs. CrowdSPRING* (Online). Available: https://signalvnoise.com/posts/1253-the-nospec-campaign-vs-crowdspring.

Bernipe, P. 2014. *European's Cultural Heritage Online* (Online). Available: http://epthinktank.eu/2014/04/09/europes-cultural-heritage-online.

Blanke, T. 2014. *Digital Asset Ecosystems: Rethinking Crowds and Cloud*, Elsevier.

Blanke, T., Bodard, G., Bryant, M., Dunn, S., Hedges, M., Jackson, M. and Scott, D. Linked data for humanities research—The SPQR experiment. Digital Ecosystems Technologies (DEST), 6th IEEE International Conference on, 2012. IEEE, 1–6.

Blanke, T., Bryant, M. and Hedges, M. Back to our data—Experiments with NoSQL technologies in the Humanities. Big Data, IEEE International Conference on, 2013. IEEE, 17–20.

Blanke, T., Bryant, M., Hedges, M., Aschenbrenner, A. and Priddy, M. Preparing DARIAH. E-Science (e-Science), IEEE 7th International Conference on, 2011. IEEE, 158–165.

Blanke, T. and Kristel, C. 2013. Integrating Holocaust Research. *International Journal of Humanities and Arts Computing,* 7, 41–57.

Briscoe, G. and Sadedin, S. 2009. Digital Business Ecosystems: Natural science paradigms. *arXiv preprint arXiv:0910.0646*.

Buneman, P., Cheney, J., Tan, W.-C. and Vansummeren, S. Curated databases. Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2008. ACM, 1–12.

Commission, E. 2014. *Research Infrastructures Programme* (Online). Available: http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/64-infraia-1-2014-2015.html.

Dombrowski, Q. 2014. What ever happened to Project Bamboo? *Literary and Linguistic Computing,* 29**,** 326–339.

Edwards, P. N., Bowker, G. C., Jackson, S. J. and Williams, R. 2009. Introduction: An agenda for infrastructure studies. *Journal of the Association for Information Systems,* 10, 6.

ESFRI. 2010. *ESFRI Roadmap* (Online). Available: https://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf.

Gartner, R. and Hedges, M. CENDARI: Establishing a digital ecosystem for historical research. Digital Ecosystems and Technologies (DEST), 7th IEEE International Conference on, 2013. IEEE, 61–65.

Holley, R. 2010. Crowdsourcing: How and why should libraries do it? *D-Lib Magazine* (Online), 16. Available: www.dlib.org/dlib/march10/holley/03holley.html.

Hyde, A. 2008. *Spec Work Is Evil/Why I Hate CrowdSpring* (Online). Available: http://andrewhy.de/spec-work-is-evil-why-i-hate-crowdspring.

IEEE Digital Ecosystem. 2007. *Digital Ecosystem* (Online). Available: www.ieee-dest.curtin.edu.au/2007/ (accessed 2 January 2013).

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*, London, Sage.

Laumeister, G. 2014. *Why Online Marketplaces Are Booming* (Online). Available: www.forbes.com/sites/groupthink/2014/08/20/why-online-marketplaces-are-booming.

Malone, T., Laubacher, R. and Dellarocas, C. 2009. Harnessing Crowds: Mapping the Genome of Collective Intelligence. *Working Paper Series* (Online). Available from: http://cci.mit.edu/publications/CCIwp2009-01.pdf.

Mayer-Schönberger, V. and Cukier, K. 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think,* Boston, MA, Houghton Mifflin Harcourt.

Preston, A. 2015. *The War Against Humanities at Britain's Universities* (Online). Available: www.theguardian.com/education/2015/mar/29/war-against-humanities-at-britains-universities.

Reisz, M. 2014. *Humanities Research 'Needs Firm Foundations'* (Online). Available: www.timeshighereducation.com/news/humanities-research-needs-firm-foundations/2014810.article.

Reuters. 2011. *Kaggle Raises $11 Million in Series A Financing Led by Index Ventures and Khosla Ventures* (Online). Available: www.reuters.com/article/idUS58636+03-Nov-2011+BW20111103.

Richards, J. D. 2012. Digital Infrastructures for Archaeological Research: A European Perspective. *CSA Newsletter* XXV.

Ruppert, E., Law, J. and Savage, M. 2013. Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society,* 30**,** 22–46.

Speck, R., Blanke, T., Kristel, C., Frankl, M., Rodriguez, K. and Daelen, V.V. 2014. The past and the future of Holocaust research: From disparate sources to an integrated European Holocaust research infrastructure. *arXiv preprint arXiv:1405.2407.*

Svensson, P. 2011. From optical fiber to conceptual cyberinfrastructure. *Digital Humanities Quarterly,* 5.

Thaller, M. 2012. Controversies around the digital humanities: An agenda. *Historical Social Research/Historische Sozialforschung*, 7–23.

Van Zundert, J. 2012. If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities. *Historical Social Research/Historische Sozialforschung*, 165–186.

Váradi, T., Wynne, M. and Koskenniemi, K. CLARIN: Common Language Resources and Technology Infrastructure. 2008.

Zittrain, J. 2008. Ubiquitous human computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 366, 3813–3821.

5 The ethnography of infrastructures

Digital Humanities and Cultural Anthropology

Gertraud Koch

Introduction

The Digital Humanities (DH) tools and infrastructures discussed in this volume describe new applications and results, particularly in those areas of Cultural Anthropology focusing on the digitalization of ethnographic museums and archives for the storage and exhibition of the cultural, usually as material forms of expression. However, digital archiving of material culture is not just for Cultural Anthropology but is also relevant to many other disciplines. More specific to and constitutive for Cultural Anthropology is its ethnographic research methodology, which will form the basis of the reflections in this chapter. The guiding motivation here is determining how Digital Humanities tools and infrastructures in Cultural Anthropology are established in order to then develop a more informed awareness of the current state of the art and to identify crucial factors for the further development of their contribution to Digital Humanities.

This contribution picks up the thread of an area of cultural anthropological research that, because of long-standing research into ICT, has developed a heuristic approach demonstrating how infrastructures and practices of related user groups are connected. Even if this ethnography of infrastructures may not be strictly counted as Digital Humanities but rather as Science and Technology Studies (STS) research, it still constitutes an interesting set of tools that will help critical reflection on the state of the art of Digital Humanities vis-à-vis Cultural Anthropology.

The added value of this consideration lies in the fact that the ethnography of infrastructures (EoI) argues and analyses in terms of culture. In contrast to usability studies and interaction design (both of which take an approach based on individual factors such as motivation, user friendliness, and so on), the EoI provide a leverage point for understanding where and how infrastructures and the social practices of researchers intertwine. Thus, deeper insight can be gained on how and where Digital Humanities tools and infrastructures can find a starting point in order to support and supplement processes of insight in Cultural Anthropology. Such an analysis is interesting in so far as it thinks from the point of view of the *modus operandi* of the humanities and can investigate points of contact for IT development.

To this end, this contribution begins with a clarification of what is understood as the individual, but also very broad areas of work in Cultural Anthropology as an academic discipline, and of the Digital Humanities as a transdisciplinary area of research. It will become apparent that for the most part we predominantly find generic tools that are nonetheless infrequently used for Digital Humanities projects. In order to understand this low rate of penetration in a discipline that co-founded Science and Technology Studies and has been doing research on digitalization for decades, the heuristic concept of the ethnography of infrastructures (and its position on the connection between infrastructures and social practices) is presented. This is applied in order to analyse the use and also the non-use of Digital Humanities tools and infrastructures in Cultural Anthropology. These analyses turn our attention to existing, missing and potential matches between epistemological ways of working in Cultural Anthropology and Digital Humanities infrastructures and tools currently being developed or already in use.

Defining Cultural Anthropology and Digital Humanities

Cultural Anthropology is an established academic discipline, but also similar to Digital Humanities (and many other present-day developments in science) through its beginnings in the United States, where Franz Boas developed scientific research on the Anthropos from its cultural origins (in addition to the biological and the historical–archaeological) as a new paradigm in anthropology; this background contributed to its disciplinary differentiation. The cognitive, social and historical identity which characterizes Cultural Anthropology as an academic discipline[1] has since developed in different accentuations corresponding to the different national landscapes of academia.[2] The overlaps, fringes and differentiations in Cultural Anthropology, between social and cultural anthropology, ethnology, European ethnology, empirical cultural studies, cultural studies and folklore studies are revealed in the names of the European professional associations EASA (European Association of Social Anthropologists) and SIEF (Société Internationale d'Ethnologie et de Folklore). As is the case in all small disciplines, there are intensive negotiations over which orientation has the most promising future in the face of societal and university policy changes (for example, Hannerz, 2010; Herzfeld, 2001; Rabinow *et al.*, 2004).

Without tracing these different positions in detail here, and without being able to present the cognitive, social and historical identity that has grown over many decades, Cultural Anthropology can be conceived of as a humanities discipline that works empirically and hermeneutically, that argues historically and works out cultural analyses founded in social theory.

The dominating research paradigm in this discipline is ethnography, especially praxeography, which connects a broad spectrum of methodological and methodical approaches, from discourse and media analysis up to participant observation and interviews. The aim is primarily (post)

phenomenologically and social-constructivistically oriented analyses of the everyday as lived life in its specific executions and its manifold instances. Cultural Anthropology has been described as a science of how everything is connected to everything else (Schiffauer, 2011), as a science of translation (Kaschuba, 1999), a way of researching complexity and contextuality (Hannerz, 2010) or as 'zones of awkward engagements' (Tsing, 2005).

Compared to Cultural Anthropology, Digital Humanities is a relatively young and newly developing research area. Questions as to what constitutes the substance of Digital Humanities and how it should be defined cannot be answered in unanimous agreement. The topics, approaches and interests of the different agents who are part of such transdisciplinary developments are too dispersed and distinctive to let themselves be grasped in a coherent definition (Kuhn, 1976).

In a four-part essay on the state of the art and future developments of Digital Humanities, Patrik Svensson, the head of the Digital Humanities centre in Umeå, *stresses the continuing dynamic of the field of research:*

> The territory of the digital humanities is currently under negotiation. While there is no doubt that the field is expanding, it is not entirely clear what is included and how the landscape can be understood or structured. These ongoing negotiations occur on multiple levels, from an individual graduate student and local institutions to national funding agencies and international institutional networking. They are consequently situated institutionally, physically, politically and epistemically.
> (Svensson, 2010, passage 173)[3]

Digital Humanities research is widely perceived as being continuously negotiated in terms of the strength of its academic repute. The social sciences have at their disposal a range of conceptual instruments to define and describe the state of the field of Digital Humanities with empirically founded data. Hopefully, from this data they can devise a suitable classification system.[4]

Such questions, which always also have a policy aspect in academia, are, however, of less interest here than a systematization of the different currents under the Digital Humanities umbrella as undertaken by Svensson (2010) in an explicit appreciation of their differing approaches and epistemic traditions.[5] He takes as his point of departure the individual relationships between Information Technology and Humanities[6] and arrives at five ideal types of cooperative connection, while he conceives of the Humanities-oriented information sciences as a separate area:[7]

- *Paradigm 1*: Information technology as a tool. The uptake of information technology at the level of tools in the Humanities is manifold and refers to a large spectrum of applications and infrastructures. Areas of application and complexity of the different tools vary considerably here, in terms of prior IT knowledge and in terms of user friendliness.

- *Paradigm 2*: Technology as an object of study in a broad sense. Information technology itself has become an object of study in the Humanities and is studied with regard to social and cultural implications of the digital world.[8]
- *Paradigm 3*: Information technology as expressive media. The new possibilities in information technology are taken up in the Humanities to expand the repertoires of forms of representation and expression. In academia, this happens particularly in teaching, and to some extent also in scientific blogging. For the main part, however, it is the 'expressive, creative media' which are the objects of research.[9]
- *Paradigm 4*: Information technology as exploratory laboratories. The humanities laboratories or digital humanities centres foster an explorative and constructive character in the use of information technology and attempt, in physical, digital or hybrid spaces, to render the digital technologies productive in the sense and style of a 'cultural laboratory'[10] – for instance, in the shape of designated digital platforms, research environments, and so on.
- *Paradigm 5*: Information technology as activist venue. In the borderlands of art and artist practice, information technologies are taken up by the Humanities as an activist intervention in which cultural critical and aesthetic dimensions are foregrounded, and transgress, by means of provocation, the usual academic ways of work.

This very broad and inclusive discussion of different paradigms of the use of information technology is informative because different areas of academic work are mentioned. For pragmatic reasons, only aspects of research in Cultural Anthropology will be in the foreground here – that is, the use of Digital Humanities tools for data collection, analysis and modelling, and for presenting research results.

State of the art: Digital Humanities in Cultural Anthropology

In a recent essay on the routinization of research, the matter-of-fact integration of digital tools into anthropological research in line with the general development of media over the past fifty years is traced (Löfgren, 2014).[11] Much of what the Swedish anthropologist Orvar Löfgren describes as digital practices relates to the use of tools referred to as generic in Digital Humanities: word processing, email software, browser, snipping tools, reference management programs (such as Zotero, Citavi, Endnote), and database programs, presentation and graphics tools. Depending on the focus, we should also include image and video recording and editing programs, computer-supported cooperative workspaces (CSCW platforms), and so on.

Few of the tools mentioned here would today still be explicitly mentioned in the context of Digital Humanities, even if there were certainly times when the use of these applications was seen as pioneering work in

the Humanities in terms of information technology.[12] The medialization of life worlds, which, as Löfgren writes, is consented to by cultural anthropological researchers as if it is a matter of fact, has grave consequences for research practice in a discipline such as Cultural Anthropology, which banks on phenomenological and post-phenomenological points of access to its fields.

The entire setting of fieldwork changes fundamentally in its time-space relationships (Welz, 2012): people in the research field increasingly communicate via media independent of direct face-to-face contact, location and time difference, synchronously or asynchronously. The forms of interaction have multiplied. Social media and related hardware, from Facebook via email to Skype, Internet-based media and mobile phones have created new possibilities for communication which all create specific requirements for research.

The infiltration of digital media into everyday research life has permanently changed cultural anthropological knowledge production and has triggered a range of reflections on how these changed conditions and forms of communication in digital media or platforms need to be systematically considered in fieldwork (for example, Hine, 2000; Markham and Baym, 2009; Kjaerulff, 2010; Boellstorff, 2008, 2012) and others). [13] However, research in media anthropology[14] is not well covered in the Digital Humanities scientific community. At least it is not evident in citations, even though media anthropology research is primarily about the use of IT tools in the process of research as well as the reflection of related methodological problems. For example, Svensson's systematization of Digital Humanities results in insights regards Digital Humanities paradigms (IT as tool and IT as object of study) but without clear reference to research in media anthropology.

If we look at the IT applications common in Cultural Anthropology, which are not generic but constitute specific tools and infrastructures for cultural anthropological research, we find a rather modest and limited inventory. For *data collection and generation*, hardly any Digital Humanities tools are employed. At most, in quantitative research (which is the exception in cultural anthropological research), survey tools or network analysis tools are employed (Fielding, 2012). In isolated cases, such tools have also been employed in historical cultural analysis (Lemercier, 2011; Lipp, 2003). For mapping, sometimes GIS systems are used (Fielding, 2012), an occasion which remains as sporadic so far as data collection via crowd sourcing (Ridge, 2014; Griffith and Mayer, 2013) and is not part of typical academic training.

As is the case in many other disciplines, in the area of material culture, particularly in museums, archives, libraries, and so on, digital repositories – respectively, databases – have emerged in recent years, originating from the large-scale digitalization of analogue inventories *a la Europaeana*. These are of interest to research in so far as they facilitate *data storage* and provide research materials. On the other hand, the significance of such digital

archives in a discipline that primarily conceives of itself as a science of everyday life and puts the exploration of cultural processes and practices in the foreground is somewhat less central than in primarily historically oriented disciplines of the Humanities. Nevertheless, very similar questions arise (Greengrass and Hughes, 2008).

Those parts of cultural practice that are immaterial, performative and therefore tied to action are particularly relevant in Cultural Anthropology. However, they are also particularly problematic to archive, both in analogous and digital forms. A relatively ambitious approach for archiving and subsequent data storage in the sense of sustainable data (Barwick and Thieberger, 2006) was/is being developed in the context of the Australian PARADISEC project, 'a facility for digital conservation and access to endangered materials from all over the world'. Although it is primarily linguistically the focus of research, in some aspects it also offers starting points for Cultural Anthropology.[15] A primary aim here is to make gathered data accessible to a broad community of researchers, which corresponds with an increasing expectation of third- party funding bodies to the sustainability of research data.

In the area of *data analysis*, software for qualitative data analysis (MAXQDA, NVivo, ATLAS.ti) is being employed which supports the hermeneutical process of analysis. It structures multimodal data evaluation and is designed in such a way that the process of analysis can be organized according to the principles of Grounded Theory (Strauss and Corbin, 1990). Automatized or semi-automatized approaches to analysis such as are being undertaken in linguistics (e.g. Brett, 2012) are not inherent in the tool. Still, this process must be conducted entirely by head and hand of the respective researcher.

The above paragraphs largely outline the repertoire of Digital Humanities tools in Cultural Anthropology. Some generic and thus less discipline-specific tools in the area of representation and dissemination of research results could still be named, such as wikis, blogs or museum apps for mobile devices. A detailed classification of the application of Digital Humanities approaches in Cultural Anthropology becomes possible when one looks at the tool and infrastructure-related Digital Humanities activities in other disciplines of the Humanities – for instance, research undertaken in the contexts of the Europe-wide infrastructure projects DARIAH and CLARIN.

Here, a large bandwidth of IT-based tools for research becomes apparent in the aligning of objects, texts, media, and so on. With parallel data, the use of visualizing tools for data and analysis results, for instance, in the shape of tag clouds, animated graphics and digital maps, the (semi-) automatized evaluation of text corpora (linguistics and sign language), edition projects, plural annotations of objects of analysis also by non-scientists, and so on. Many more Digital Humanities tools and infrastructures can be found here that could prove useful in Cultural Anthropology if they were adapted appropriately. This raises the question as to why such options have so far not been developed.[16]

Cultural anthropological research is far more developed in research on digital developments in the sense of the second paradigm mentioned by Svensson, where it can look back on a long research tradition as a co-founder of science and technology studies. It has also generated interesting insights into the cultural implication of technology development, and specifically the implications of digital technologies (Traweek, 1992; Nader, 1996; Latour, 1987; Kelty, 2008; Pickering, 2010; Hayles, 1999; Escobar, 1994). In the face of the size of this corpus of research, these can only be discussed here approximately, even if many aspects are of interest to Digital Humanities. Particularly relevant in the context of this volume are the ethnographies of infrastructure (EoI), which are taken up in more detail in the following paragraphs.[17]

Infrastructures: entanglements with social practices of scholars

The question of how Digital Humanities tools can be designed to be more user-friendly for broad audiences is commonly and intensively discussed in the scientific community (Gibbs and Owens, 2012; Evens, 2012; Svensson, 2010) with very different avenues of explanation developed. Important feedback is expected from usability studies and the evaluation of interaction design techniques (for example, Drucker, 2013; Lövgren and Stolterman, 2004) even if there is tension between users of generic tools and tools developed for specific individual epistemic communities (Ratto, 2006).

An important criticism in this debate aims more at the cultural implications of Digital Humanities. With reference to the STS researchers Geoffrey Bowker and Susan Leigh Star (Bowker and Star, 1999), the necessity of taking seriously the impact on Humanities research created by digitalization is currently taking place. This criticism is aimed at the hegemony of US American culturality, which is said to be inscribed into the software, reproducing global hierarchies (Fiormonte, 2012). On the other hand, the epistemological limitations of the binary code in digital tools raise their own problems. Creativity is supposedly reduced to choices of possibilities selected by programs, causing independent new creations to disappear from the horizon of possibilities in the Humanities production of insight (Evens, 2012).

This addresses the fundamental area of tension between hermeneutic ways of working and the automatized processes in the relationship of IT and Humanities, and it encapsulates the less clearly specified discomfort of many humanities researchers. This fundamental criticism also brings to our attention the fact that beyond conventional usability studies more fundamental questions will play a role in the future development of useful and meaningful tools and infrastructures. An examination of fundamental epistemological and social practices of the Humanities, which have seen hardly any research in this respect,[18] could also be instructive with regard to difficulties encountered in the acceptance of Digital Humanities applications.[19]

A heuristic approach for the question and analysis of how social practices and infrastructures are connected and come together has been developed by the above-mentioned Science and Technology Studies researcher Susan Leigh Star. She based her experiences on the framework of her collaboration in software development projects with her contribution 'Ethnography of Infrastructures' (EoI) (Star, 1999), which triggered the development of wide-ranging cultural analytical infrastructure research (Lampland and Star, 2009).

Star formulated an understanding of infrastructures which stresses their entanglement with the social practices of user groups in the sense of communities of practice, and she saw these as a form and element of social organization and thus as just as fragile (Star, 1999). At the centre of effective infrastructures, in contrast to the focus of many usability and interaction studies, is a focus on the collective rather than on the individual:

> One way to explain this magnification process [of work] is to understand that in fact two processes of work are occurring simultaneously: Only one is visible to the traditional analysis of user-at-terminal or user-with-system. That is the one that concerns keystrokes and functionality. The other is the process of assemblage, the delicate, complex weaving together of desktop resources, organizational routines, running memory of complicated task queues (only a couple of which really concern the terminal or system), and all manner of articulation work performed invisible to the user.
>
> (Star, 1999, 386f.)

Star characterizes the infrastructures as relational, ecological, boring and common and thus emphasizes their close interweaving with the conduct of everyday life which – as can easily be deduced here – as a result get muddled up in the implementation of infrastructures and tools.

> It is not just the bits and bytes that get hustled into standard form in order to the technical infrastructure to work. People's discursive and work practices get hustled into standard form as well. Working infrastructures standardize both people and machines.
>
> (Bowker and Star, 2004)

The usefulness of standardizations and automatization in Humanities research, however, is not unlimited. Sorting things out (Bowker and Star, 1999), which can be standardized and automatized and for which this therefore does not apply, is always connected with epistemological questions which, in the course of IT developments for the Humanities, pose themselves anew and urgently. They are also posed emphatically with a certain research policy pressure,[20] yet the answers can hardly be the same for all disciplines.

The analysis and accommodation of epistemological new adjustments is less about dichotomies between automation and the hermeneutical, and more about gaining a more nuanced understanding of relationships and connections:

> This vision requires . . . thinking about infrastructure not only in terms of human versus technological components but in terms of a set of interrelated social, organizational, and technical components or systems (whether the data will be shared, systems interoperable, standards proprietary, or maintenance and redesign factored in).
>
> (Bowker *et al.*, 2010, 99)

If one takes seriously these cultural analytical research results which were gained in long-term empirical and theoretical work in different IT development projects by Susan Leigh Star and her colleagues, a whole assortment of social and cultural influencing factors becomes apparent. This is of critical importance to selection and acceptance of Digital Humanities tools. In the sense of an infrastructure research oriented at cultural analysis, questions arise about how the integration of IT tools and infrastructures into the process of insight in the Humanities facilitates and shapes new social forms, which social, ethical and political constellations are consolidated, disturbed or created, and which ontological changes take place in knowledge work (cf. Bowker *et al.* 2010, 105). Aside from the epistemological, there are also praxeological qualities that are crucial for whether a tool is perceived as a relevant work simplification, addition or enrichment of humanities research work.

Praxeologic here means more than mere everyday practice but, following Pierre Bourdieu, it refers to practical implementations in everyday action framed by social relationships which are inscribed into the individual and by which, conversely, the individual is incorporated into social relationships and thus characterized as an agent, as a member of a certain social group. It is therefore this action which contributes to the production and reproduction of social groups as well as the social structure originating from their entirety (Bourdieu, 1990).

Of particular relevance here, all forms of practice serve to characterize the individual humanistic academic as a member of his or her disciplinary collective and thus also to mirror the normative and structural dimensions of the systems of academia. This applies in particular to forms of practice, which are pre-reflexive and thus not immediately visible, yet nevertheless critical for research to function. From a praxeological perspective, 'the layers of silence and arenas of voice, the ecology of visible and invisible work' (Star and Strauss, 1999) are therefore important starting points for an understanding of how organizational, technological and epistemological processes intertwine in the processes of the insight of a discipline.[21]

The 'translation' and (re-)configuration of these relationships in Digital Humanities tools and infrastructures are anything but trivial,[22] even for the most mundane knowledge practices:

> Thus rather than the rhetoric of revolutionary fervor that permeates cyberinfrastructure circles, infrastructure studies take as its object change at a much more mundane scale: as forms of practice, routine, or distributed cognition associated with knowledge work. Is this position against a possibility of 'revolution?' Not at all. It is, rather, a research sensibility which seeks to make transformations of infrastructure visible relative to the everyday work of scientists, information technologists, or information managers.
>
> (Bowker *et al.*, 2010, 111)

In this sense, a detailed look at cultural anthropological research practice is necessary in order to gain information about why the state of Digital Humanities is not further developed and where possible starting points for it may lie.

Exploring the absence and the potentials for growth of Digital Humanities tools and infrastructures in cultural anthropological research

Culture has many faces; it is both persistent and highly changeable. Under the conditions of the post, late or liquid modernity (Giddens, 1990; Bauman, 2000) into which Western societies today are involved, the tendency is shifting towards the more dynamic, productive, creative side.

This is the starting point also for the analysis of culture as a hypothesis-generating research practice, which, starting from micro-analytic connections, seeks to abductively understand larger cultural developments and formations. It empirically starts from the materializations of culture, the discourses (the symbolic world with its specific horizons of interpretation), the things (the material world as an important element of social organization) and the practices (the world of action). In this, it is often the interaction of these different areas, of the contradictions and paradoxes that can be detected here.

These, however, are hard or not at all to be extracted in automatized processes. How complex and full of preconditions cultural analysis can be is shown in eminent cultural theoretical positions such as those by Michel Foucault or Pierre Bourdieu. Their insights had much to do with the social situating of researchers in the relevant societal circumstances (Bourdieu: participant observation). In conclusion, this points to the set of problems surrounding automatization in cultural analysis where the specific analytical standpoints of a researcher can only have an indirect influence and indirectly be made productive for the process of insight, at least in some respects.[23]

Epistemologically, the automatization of data analysis simply means a reversal of previous approaches in hermeneutical–empirical cultural analysis as they are undertaken in Cultural Anthropology. In the sense of hypothesis generation, work is mainly undertaken abductively, while the principles of Grounded Theory as a central analytical paradigm are, more or less explicitly, being used as a basis. The cultural analytical, hypothesis-generating way of work is committed to great openness regarding the studied field and has few predefined research parameters. In the (semi-) automatized processes of Digital Humanities, insofar as they are common – for instance, in linguistics – the research logic is diametrically linked to cultural anthropological epistemology since sets of problems, research questions and relevant research areas are already set in advance.

Wherever programs for automatized or semi-automatized analysis are employed, these must be defined, specified and determined as parameters of the field being studied beforehand. This is in contrast to the processual character of ethnographic research, with its interpretative intermediate steps and iterative adjustments of topical focus points that are substantively for the abductive, hypothesis-generating research approaches that characterize Cultural Anthropology. Also, in areas where 'learning' approaches are employed in Digital Humanities, it is only narrow partial aspects to which these adaptations can refer. Fundamental modifications of the methodological repertoire, differentiations, new adjustments or a revision of hypotheses are neither envisaged nor possible in such 'learning' settings of analysis of Digital Humanities. At the most, such changes are made in a second step when new studies are conceptualized.

Grounded theory in this sense is a powerful and rich epistemological 'tool' whose epistemological capacity cannot begin to be reached via Digital Humanities approaches – at least, not yet. The abductive, hypothesis-generating orientation has little affinity for the automatization of the processes of analysis which go beyond the use of generic tools, for one continuously enters into new cultural fields, themes and processes. In addition, these, along with their dynamics, change rapidly in current societies with their dynamics so that the sets of problems (as they become apparent in the context of dynamic annotations) here present themselves in a particularly obvious manner.

The great bandwidth, flexibility and dynamic that characterize current cultural developments make it hardly attractive to look for automatizations for analysis as it is rather 'trend scouting' which is required in order to get a glimpse of them, to detect academic ones, and as Cultural Anthropology tends to work in processes that are suitable for researching these 'near futures'. The epistemological interest thus turns to looming and potentially constitutive dispositives. As the future depends on many things and shows contingencies, this is always linked to an uncertainty as to whether such tendencies will actually become established and gain social potency (Rabinow, 2003). Pointedly formulated, from a cultural anthropological view, emerges the question as to

why a good tool which yields rich analytical insight and which has proved itself epistemologically (even though it is analogue, such as Grounded Theory), should be given up and enormous energy invested into the development of IT technologies of debatably less analytical value.

The applications predominantly employed in current Cultural Anthropology –that is, mainly generic tools for word processing, CSCW and presentation tools – can be seen as paradigmatic for the areas of the application of digital tools and infrastructures in this field. However, such a conclusion is also too quick and too superficial in the face of societal developments, which are currently driven by the processes of digitalization, inscribing themselves into everyday life. The increasing shift of social events into the digital environment and the 'emigration' of interpersonal communication into environments brokered by media are life world phenomena which must move more strongly into the centre of academic/scientific (participant) observation, which, in turn, must be reinvented.

Access to participation in and observation of email, text messages, Second Life and other applications that constitute a continuously growing share in interpersonal media-supported communication appears a mere illusion without the corresponding 'armament' of ethnographic praxeological research. The question is rather, where, from the point of view of Cultural Anthropology, specific, epistemologically compatible integration of Digital Humanities methods originate and what these might specifically look like. Conversely, it is true that if one does not want to shelve ethnographic praxeographic methods, one will either have to think of strategies or will increasingly find that more and more important spaces of social interaction (and, by extension, everyday life), will remain inaccessible to cultural anthropological research.

There are individual Digital Humanities projects (which include participation by cultural anthropologists and qualitative researchers from sociology) that point the way and are instructive in this respect, and because they partly take place in the context of recent developments in academia. In such approaches, a range of new questions and sets of problems are raised and will need to be answered. For instance, access to research data for an entire scientific community such as, for example, the PARADISEC project aims to provide for, may collide with requirements to protect involved informants or comply with ethics guidelines. Also, ways of dealing with and citing the authorship of this data and many other questions – for instance, regarding necessary metadata – must be clarified in this connection. Anthropology certainly can learn from Digital Humanities approaches in other fields.

The secondary utilization of qualitative data is a topic that will increasingly become relevant where qualitative data are 'digitally born' as a standard and thus can become an object of trans-local research practice in a relatively uncomplicated manner, which may present entirely new possibilities such as for distributive evaluation, discussion and classification of results. They have, however, so far not yet been developed.

New approaches to data maintenance and its subsequent use are also of the highest interest for each individual researcher in Cultural Anthropology because typical multi-media/multimodal research materials exhibit a great heterogeneity, and ways of dealing with them have so far been developed relatively idiosyncratically without there being any systematic, methodological concepts for this in the research literature (cf. Löfgren, 2014).

Also, one could imagine further starting points and infrastructures that support the contextualization and filing of the research data that are usually already generated with digital technologies such as photographs, sound recordings or internet documents via automatized date stamps, location coordinates, and many more. If one looks at current ethnographic praxeographic research with an interest in Digital Humanities in their course and questions existing IT applications for their potentials for facilitating, supplementing and supporting the procedures applied, a large number of further useful infrastructures and tools can certainly be identified. Even if these were tied to corresponding adaptations and effort was made in developing them, interesting and epistemologically relevant and compatible extensions of the cultural anthropological repertoire of analysis could be developed with them.[24] Decisive in this connection, however, would be to start from the standpoint of cultural anthropological approaches, rather than from developments in information technology. The question as to where one could intensify this connection, there are already existing developments in application and infrastructure – for example, the European infrastructure projects DARIAH and CLARIN.

Conclusion: a prognosis

Cultural Anthropology hardly has less of an affinity towards IT than other Humanities disciplines, even if it currently is only involved marginally in Digital Humanities developments. The digitalization of existing archives currently undertaken in the context of museums and archive inventories is probably the most important relevant activity here. With regard to cultural analytical perspectives, it is probably rather approaches that strengthen analytical capacity, and thus the epistemological repertoire of Cultural Anthropology, that holds the most promise. Increasingly, the challenge posed will include undertaking research in digitalized life worlds and also gaining access to spaces that cannot be tapped by the presence of the researcher alone in participant observation mode (because these new forms of spaces are medialized). The hermeneutic approach here meets its limits, particularly also in data generation and collection. Alternatively, powerful analysis methods such as Grounded Theory can now (and into the foreseeable future) compete with digital, automatized applications due to their pragmatism and epistemological depth.

Even now, the additional gains for Cultural Anthropology through Digital Humanities approaches can support everyday practices and processes of

research, such as in the archiving and maintaining data. These applications can be meaningfully employed to this end; however, they are still in their infancy or have not even been developed yet. Digital Humanities research and development work is necessary in order to heighten the intensity of Digital Humanities applications in Cultural Anthropology. This will also change the research processes themselves so that in the sense of an ethnography of infrastructure, IT development work will need to be flanked by cultural analytical work in order to arrive at meaningful tools and infrastructures.

As an intellectual reward for such dedicated effort, an intensified new reflection of hermeneutic methods and methodologies will help reveal implicit fundamental assumptions of epistemological principles, and also help the development of valid and useful automatized approaches. As to the area which has been largely left out here – the representation and dissemination of research – we can expect major new advances, such as in the synergistic collaboration of researcher networks and new forms of digitally augmented teaching. However, these new horizons will need to be the focus of other chapters, not this one.

Notes

1 Cf. Lepenies, 1981 on the concept of the cognitive, social and historical identity of sciences and humanities.
2 The differentiation between social and cultural anthropology (sociocultural anthropology in Germany) has become well established (cf. Barth (2005), while European ethnology marks its own tradition, with roots in German studies, the individual philologies of the country's language. Since the 1960s, European ethnology has developed into an empirically working, cultural analytical discipline with different accentuations of these research traditions with reference to cultural studies and sociology.
3 Further positions on the definition of Digital Humanities are discussed in Gold (2012); Thaller (2012); Terras (2013).
4 See, for example Galison (1997); Gorman (2010); Kanfer *et al.* (2000); Kuhn (1976)).
5 'Importantly, while there is a great deal of ongoing collaboration and identification of common goals, we need to be aware of the fact that we are concerned with several epistemic traditions and visions. It could be argued that there is a need for an expanded, multi-layered discussion of visions, collaborative possibilities, and possibly but not necessarily, an articulated convergent path forward' (Svensson, 2010, pp. Absatz 14–16).
6 The reflection and discussion of the relationship between IT and Humanities belongs to the substantial themes of the Digital Humanities and is a key question for the cognitive identity of the area of research. Therefore, it is also raised elsewhere, such as, for instance, in Rosenbloom, 2012).
7 Svensson speaks of 'modes' but not of paradigms. He also keeps an eye on other forms of systematization. On one hand, he refers to Tara McPherson (2009), who differentiates between the computing humanities, blogging humanities and multimodal humanities. On the other hand, he takes up Davidson's differentiation into Humanities 1.0 and Humanities 2.0, which traces the development of large-scale database projects to the more interactive, decentralized and participative practices of knowledge (Davidson, 2008, 711–12).

8 In his remarks on this mode of research, Svensson mainly refers to 'Internet Studies' which have organized themselves in the Association of Internet Researchers. This move indicates that there is some distance to the Digital Humanities. Internet studies are intensively and far more broadly developed further in the historical, social and cultural studies of science, which are internationally growing, interdisciplinary and organized in different associations. This limited reception in the Digital Humanities is also stated by Svensson himself and could, in this sense, also be understood as a moment in which a specification of Digital Humanities is undertaken implicitly: 'Initiatives with a significant investment in this mode often seem fairly discrete in the landscape of the digital humanities while they are rarely in fact recognized as digital humanities' Svensson (2010, pp. Absatz 130).

9 In this line of argument, Svensson refers to Schnapp and Shanks (2009, 147) and McPherson (2009).

10 The term and concept of the 'cultural laboratory' are taken on by Svensson from Janlert and Jonsson (2000), while leaving other debates on the laboratory character of the production of knowledge in the humanities and social sciences unconsidered, such as are being discussed in US cultural anthropology (Rabinow , 2008; Rabinow *et al.*, 2008).

11 However, in contrast to the technical work environments in the natural sciences, there is also a gap in research on work practices in the empirically working cultural sciences, and Löfgren points this out (Löfgren, 2014). In this sense, the following description of work practices supported by information technology in cultural anthropology is not the result of systematic research but results from para-ethnographic observations – i.e. observations collected not for research purposes but while working as a member of the scientific community in this field.

12 The manifold adaptations which these tools have undergone in terms of their user friendliness and continuous adaptations to user practices are analogous to the mutual adaptations affecting both the technologies and their users.

13 It needs to be emphasized here that media anthropological perspectives remain largely rooted in a phenomenological epistemology whose potential for insight in the face of the medialization of life worlds must at the very least be described as limited because media facilitate qualitatively new ways of perceiving the world. This is convincingly illustrated by American philosopher Don Ihde (1993).

14 The scientific community in media anthropology organizes itself in particular via the EASA Media Anthropology Net: www.media-anthropology.net/ (accessed 11 August 2014).

15 Available at: http://paradisec.org.au/home.html (accessed 12 August 2014). The project is strongly inspired by linguistic anthropology; however, it goes far beyond this field and integrates further ethnographic projects from Australia. It is of interest for cultural anthropological research in view of data storage and maintenance because audio-visual data are stored and processed. Moreover, tools for the annotation of audio-visual data via a three-dimensional annotation tool are provided (Hunter and Yu, 2011).

16 However, it is also being discussed why Digital Humanities approaches are only haltingly received in the Humanities disciplines. In this context, debates surrounding the future discipline specific development of the Digital Humanities are of importance, such as Gold, 2012; Harpold, 2012; Thaller, 2012.

17 For the other paradigmatic orientations named by Svensson, information technology as expressive media, as exploratory laboratories and as activist venue, one could also find cultural anthropological examples. As generic applications are foregrounded here, loose connections to IT are entered, and these are less relevant than the questions about infrastructures and tools that are at the centre here, so they will not be analysed further.

18 On the research literature, which indirectly provides information on this, see Löfgren, 2014.
19 A retrospective view on the examination of IT developments in art history is instructive in this connection, which went from the initial large and IT-driven goal of automatized image analysis at the end of the 1980s to the 'smaller' practical questions of the art historic community about the access to sources via digital archives (Bailey, 2010).
20 The humanities find themselves here in a dilemma. 'It appears that the digital humanities are the victims of a continuous paradox: demonstrating their ability to keep up with technologies (and with their actual and virtual protagonists) and, at the same time, not to become subject to them' (Fiormonte, 2012, 9).
21 In difference and as a contrast to tool development in Digital Humanities which, in accordance with the information technology systematics and for good reason by now tends to approach its developments more from general processes (such as editing, annotating etc.), in the connections discussed here, it is therefore also particularly the variability and bandwidth of hermeneutical approaches in the working connections of different humanities disciplines that constitute an important field of learning for a praxeological understanding of the Humanities disciplines.
22 Different qualities of infrastructures need to be considered here, which again act in relation to each other. Infrastructures function as 'material artifact', as 'trace or record of activities' and as 'veridical representation' (Star 1999, 387–388).
23 Insight into the specific and individual resources grounded in one's own biography which flow into cultural analytical work will most likely only become sound in the process of the respective research. Self-insight and scientific results here are most closely interconnected, a fact that is now being methodically systematically made available via auto-ethnography (Ehn, 2011). For other research areas where insight is less concerned about this specific praxeological moment of ethnographic research strategy, such as media or communication studies research approaches, this is more likely.
24 The development of relevant applications for CA research is most likely not of as much interest to the national funding agencies as the development projects of computer linguistics, which were funded early on with the rise of information technologies, particularly by the US American DARPA. If similar sums were available to CA and other Humanities disciplines, methodologically useful infrastructures and areas of application could certainly be imagined. From the point of view of the theory of science, more intensive praxeographic research of the Humanities would be of interest, and particularly the path of linguistics with regard to Digital Humanities would be an instructive research area.

References

Bailey, C. (2010) 'Introduction: Making knowledge visible', in Bailey, Chris and Gardiner, Hazel (eds). *Digital Research in the Arts and Humanities*. Farnham: Ashgate, pp. 1–10.

Barth, F. (2005) *One Discipline, Four Ways British, German, French, and American Anthropology*, University of Chicago Press, Chicago.

Barwick, L. and Thieberger, N. (2006) *Sustainable Data from Digital Fieldwork: Proceedings of the Conference Held at the University of Sydney, 4–6 December 2006*, Sydney University Press, Sydney.

Bauman, Z. (2000) *Liquid Modernity*, Polity Press, Cambridge.

Boellstorff, T. (2008) *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*, Princeton University Press, Princeton, NJ.

Boellstorff, T. (2012) *Ethnography and Virtual Worlds: A Handbook of Method*, Princeton University Press, Princeton, NJ.

Bourdieu, P. (ed.) (1990) *The Logic of Practice*, Stanford University Press, Stanford, CA.

Bowker, G. C., Baker, K., Millerand, F. and Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, in Hunsinger, J., Klastrup, L. and Allen, M. (eds), *International Handbook of Internet Research*, Springer Netherlands, Dordrecht, pp. 97–117.

Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out: Classification and its Consequences: Inside Technology*, MIT Press, Cambridge, MA.

Bowker, G. C. and Star, S. L. (2004) How to infrastructure, in Lievrouw, L.A. and Livingston, S. (eds), *The Handbook of New Media: Updated Student Edition*, Sage, London; Thousand Oaks, CA; New Delhi, pp. 230–245.

Brett, M. R. (2012) Topic modeling: A basic introduction, *Journal of Digital Humanities*, 2(1), pp. 12–16.

Drucker, J. (2013) Performative materiality and theoretical approaches to interface, *Digital Humanities Quarterly*, 7(1).

Ehn, B. (2011) Doing-it-yourself: Autoethnography of manual work, *Ethnologia Europaea*, 41(1), pp. 53–64.

Escobar, A. (1994) Welcome to Cyberia: Notes on the anthropology of cyberculture, *Current Anthropology*, 35(3), pp. 211–232.

Evens, A. (2012) Web 2.0 and the ontology of the digital, *Digital Humanities Quarterly*, 6(2).

Fielding, N. G. (2012) The diverse worlds and research practices of qualitative software, *Forum Qualitative Social Research*, 13(2), Art. 13.

Fiormonte, D. (2012) Towards a cultural critique of the digital humanities, in Thaller, M. (ed.) Controversies around the Digital Humanities, *Historical Social Research*, 37.

Galison, P. (1997) *Image and Logic: A Material Culture of Microphysics*, University of Chicago Press, Chicago.

Gibbs, F. and Owens, T. (2012) Building better digital humanities tools: Toward broader audiences and user-centered designs, *Digital Humanities Quarterly*, 6 (6).

Giddens, A. (1990) *The Consequences of Modernity*, 1. University Press, Stanford, CA.

Gold, M. K. (2012) *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis, MN.

Gorman, M. E. (2010) *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration, Inside Technology*, MIT Press, Cambridge, MA.

Greengrass, M. and Hughes, L. M. (eds) (2008) *The Virtual Representation of the Past: Digital Research in the Arts and Humanities*, Vol. 1, Ashgate, Farnham; Burlington, VT.

Griffith, M. and Mayer, V. (2013) MediaNOLA: A digital humanities project to tell stories of cultural production in New Orleans, *Journal of Digital Humanities*, 2 (2), pp. 27–39.

Hannerz, U. (2010) *Anthropology's World: Life in a Twenty-first Century Discipline, Anthropology, Culture, and Society*, Pluto Press; distributed exclusively in the USA by Palgrave Macmillan, London, New York.

Harpold, T. (2012) The underside of the digital field, *Digital Humanities Quarterly*, 6 (2).

Hayles, K. (1999) *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago Press, Chicago.

Herzfeld, M. (2001) *Anthropology: Theoretical Practice in Culture and Society*, Blackwell, Malden.

Hine, C. (2000) *Virtual Ethnography*, Sage, London.

Hunter, J. and Yu, C.-h. (2011) Assessing the value of semantic annotation services for 3D museum artefacts. Sustainable data from digital research: Humanities perspectives on digital scholarship. Proceedings of the conference held at the University of Melbourne, 12–14 December. Available at: http://hdl.handle.net/2123/7951 (accessed 12 August 2014).

Ihde, D. (1993) *Postphenomenology*. Northwestern University Press, Evanston, IL.

Janlert, L. -E., and Jonsson, K. (2000) Kulturlaboratoriet, *Tvärsnitt*, 1, pp. 54–61.

Kanfer, A. G., Haythornthwaite, C., Bowker, G. C., Bruce, B. C., Burbeles, N., Porac, J., and Wade, J. (2000) Modelling distributed knowledge processes in next generation multidisciplinary alliances, paper presented at Academia/Industry Working Conference on Research Challenges (AIWORC'00).

Kelty, C.M. (2008) *Two Bits: The Cultural Significance of Free Software: Experimental Futures*, Duke University Press, Durham, NC.

Kjaerulff, J. (2010) *Internet and Change: An Anthropology of Knowledge and Flexible work*, Intervention Press, Højbjerg, Denmark; distributed in North America by Left Coast Press, Walnut Creek, CA.

Kuhn, T. (1976) *Die Struktur wissenschaftlicher Revolutionen,* 2. Revidierte Auflage, Frankfurt, A.M.

Lampland, M. and Star, S. L. (2009) *Standards and their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life,* Cornell Paperbacks, Cornell University Press, Ithaca, NY.

Latour, B. (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*, Harvard University Press, Cambridge, MA.

Lemercier, C. (2011) Formal network methods in history: Why and how? *Social Networks, Political Institutions, and Rural Societies*, pp. 281–310.

Lipp, C. (2003) Struktur Interaktion räumliche Muster: Netzwerkanalyse als analytische Methode und Struktur Interaktion räumliche Muster: Netzwerkanalyse als analytische Methode und Darstellungsmittel sozialer Komplexität, in Göttsch, S. and Köhle-Hezinger, C. (eds), *Komplexe Welt – Kulturelle Ordnungssysteme als Orientierung*, Waxmann Verlag, Münster, pp. 49–63.

Löfgren, O. (2014) Routinising research: Academic skills in analogue and digital worlds, *International Journal of Social Research Methodology*, 17(1), pp. 73–86.

Lövgren, J. and Stolterman, E. (2004) *Thoughtful Interaction Design: A Design Perspective on Information Technology*, MIT Press, Cambridge, MA.

Markham, A. N. and Baym, N. K. (2009) *Internet Inquiry: Conversations about Method*, Sage, Los Angeles, CA.

McPherson, T. (2009) Introduction: Media Studies and the Digital Humanities, *Cinema Journal*, 48(2), pp. 119–123.

Nader, L. (1996) *Naked Science Anthropological Inquiry into Boundaries, Power, and Knowledge*, Routledge, New York.

Pickering, A. (2010) *The Cybernetic Brain: Sketches of Another Future*, University of Chicago Press, Chicago.

Rabinow, P. (2003) *Anthropos Today: Reflections on Modern Equipment*, Princeton University Press, Princeton, NJ.

Rabinow, P. (2008) *Marking Time: On the Anthropology of the Contemporary*, Princeton University Press, Princeton, NJ.

Rabinow, P., Caduff, C. and Tees, T. (2004) *Was ist Anthropologie? Suhrkamp Taschenbuch Wissenschaft*, 1687(1). Aufl., Originalausg, Suhrkamp, Frankfurt am Main.

Rabinow, P., Marcus, G. E., Faubion, J. D. and Rees, T. (2008) *Designs for an Anthropology of the Contemporary*, Duke University Press, Durham, NC.

Ratto, M. (2006) Epistemic commitments and archaeological representation. Book of Abstracts of the XV World Congress, International Union for Prehistoric and Protohistoric Sciences, available at: www.uispp.ipt.pt/UISPPprogfin/Livro2.pdf.

Ridge, M. (2014) *Crowdsourcing our Cultural Heritage: Digital Research in the Arts and Humanities*. Farnham, Ashgate.

Rosenbloom, P. S. (2012) Towards a Conceptual Framework for the Digital Humanities, *Digital Humanities Quarterly*, 6(2).

Schiffauer, W. (2011) *Parallelgesellschaften: Wie viel Wertekonsens braucht unsere Gesellschaft? Für eine kluge Politik der Differenz, X-Texte,* 2., unveränd. Aufl, Transcript-Verl, Bielefeld.

Schnapp, J. T. and Shanks, M. (2009) Artereality. (Rethinking Craft in a Knowledge Ectonomy), in Henry, S. (ed.), *Propositions for the 21st Century*, MIT Press, Cambridge, MA, pp. 141–157.

Star, S. L. (1999) The ethnography of infrastructure, *American Behavioral Scientist*, 43 (3), pp. 377–391.

Star, S. L. and Strauss, A. (1999) Layers of silence, arenas of voice: The ecology of visible and invisible work, *Computer Supported Cooperative Work*, 8, pp. 9–30.

Strauss, A. L. and Corbin, J. (1990) *Basics of Qualitative Research Grounded Theory Procedures and Techniques*, Sage, Newbury Park.

Svensson, P. (2010) The landscape of digital humanities, *Digital Humanities Quarterly*, 4(1).

Terras, M. (2013) *Defining Digital Humanities: A Reader*, Ashgate, Farnham.

Thaller, M. (ed.) (2012) Controversies around the digital humanities, *Historical Social Research*, 37.

Traweek, S. (1992) *Beamtimes and Lifetimes: The World of High Energy Physicists,* 1. Harvard University Press, Cambridge, MA.

Welz, G. (2012) Die pragmatik ethnografischer temporalisierung, in Hess, S., Moser, H. and Schwertl, M. (eds), *Europäisch-ethnologisches Forschen: Neue Methoden und Konzepte, Reimer Kulturwissenschaften*, Reimer, Berlin.

6 Building personal research collections in art history

Christina Kamposiori, Claire Warwick and Simon Mahony

Introduction

Over the past decades, the increase in the use of digital resources and the growth of research conducted in digital environments has transformed academic scholarship. Never before has there been such breadth of information and services available for scholars to use; most importantly, though, such developments offer the advantage of not only speeding up the research process, but also for facilitating innovative research enquiry. Yet, especially in the area of the Arts and Humanities, the capabilities developed by new technologies have had a great impact on the scholarly practices of its disciplines.

In fact, these capabilities have affected the way that scholars approach, create and manage information; the personal research collections of information that scholars build in the context of their research and teaching projects constitute great examples that illustrate how these practices evolve over time, the challenges that scholars meet and the particular needs that they have. In this chapter, we focus on the way that art historians build their personal information collections as a means to identify and understand the specific behaviour and requirements they have in terms of tools and services in the digital age. This will involve examining the way they gather, use and manage information in the context of their research and teaching projects.

Most of the previous studies examining the practices of scholars in the field have focused on their information-seeking behaviour (Stam, 1997; Rose, 2002; Larkin, 2010; Beaudoin and Brady, 2011). Actually, this can easily be justified if we consider the importance that the initial stages of research have for the whole research process – for example, the seeking and discovery of accurate information plays a key role in producing reliable and credible results. However, there is not much information on what art historians do after they find the material they need for their research (Palmer *et al.*, 2009). Thus, there are several questions that can be asked regarding information use and management in art history, some of which we will attempt to address in this chapter. Why do scholars choose particular resources over others for gathering the material they need? How do scholars use and manage the information they collect and what tools do they

employ for these purposes? What are the challenges faced and which are the key characteristics of art historians' work when interacting with information? Managing to answer the above questions could foster significantly our understanding of art historical practices which, to a great extent, take place 'behind the scenes', at scholars' personal workspace. This knowledge, in turn, could be applied to the building of customized tools and services that better support research and teaching in the field.

We argue that the personal collections of information that art historians build for research and teaching purposes, being at the core of their workspace (Long and Schonfeld, 2014, pp. 23–25), are an important starting point for understanding behaviour and practices that are otherwise difficult to study, due to their private nature and the various personal criteria in which they develop. Finally, the observations and issues raised in this chapter are based on the results of research conducted in the context of the first author's Ph.D. thesis at the UCL Centre for Digital Humanities, under the supervision of Professor Claire Warwick and Mr Simon Mahony. Before continuing with the presentation of our research, we consider it necessary to provide a brief overview of some of the main studies conducted with the aim of understanding the behaviour and practices of scholars, as they can provide us with a context for examining the issues raised in the following sections.

Scholarly practices in the digital age

The necessity to study the information behaviour and practices of researchers has been addressed by several studies. Ellis (1993) conducted one of the key studies in the area, which influenced later work in the field. Drawing on empirical and qualitative research, he examined the information behaviour of scholars across various disciplines. As a result, he found some common processes to occur which he presented then as features: starting, chaining, browsing, differentiating, monitoring, extracting, verifying and ending (Ellis, 1993, p. 482). A decade later, Meho and Tibbo (2003), after studying a group of social scientists, discovered similar characteristics in the information behaviour of their participants with those that Ellis had found, but they added three more features: accessing, networking and information managing.

Regarding the area of Arts and Humanities, the importance of identifying the common practices and characteristics of the above disciplines has been stressed by many studies. Thus, Unsworth (2000) introduced the concept of the scholarly primitives, arguing that there is some common, basic behaviour that can be identified across disciplines, especially in the Arts and Humanities; in particular, he listed the following primitives: discovering, annotating, comparing, referring, sampling, illustrating and representing, and used as examples various projects of that time.

In 2001, Brockman *et al.* (2001) examined the way that researchers in the Arts and Humanities work in the new information environment. More specifically, they examined the behaviour of scholars during several research

stages, such as searching, reading, writing and networking, and attempted to identify both the needs that characterize the Arts and Humanities scholars as a whole, as well as the distinct needs of the separate disciplines included in the area. Their argument was that, as new technologies bring changes to the behaviour and needs of scholars, it is important for the institutions, in that case the academic libraries, to adapt to the new circumstances.

A year later, Palmer and Neumann (2002) examined the needs of interdisciplinary scholars in the Arts and Humanities. In their paper, they highlighted the complex nature of this type of research and consequently, the additional needs that these researchers had. More specifically, two additional activities were found to occur during the process of this kind of research, exploration and translation, which complicated the research workflow. According to the authors, 'interdisciplinary exploration and translation complicate the already intensive information gathering, reading, and writing processes inherent in the humanities' (Palmer and Neumann, 2002, p. 107).

Palmer *et al.* (2009), based on Unsworth's concept of scholarly primitives, studied the scholarly activities and primitives in a variety of different disciplines, including the Arts and Humanities. In particular, they suggested five core scholarly activities, each of them including two or more primitives, and four additional cross-cutting primitives which can take place during more than one activity. The main activities were searching, collecting, reading, writing and collaborating. The objective was to report on the research process of scholars in different fields, make relevant comparisons and, therefore, provide useful information on the requirements needed for building effective digital infrastructure for scholarship.

In addition, Blanke and Hedges (2013), by bringing together the results of various e-humanities projects conducted at the Centre for e-Research at King's College London, aimed to identify the common practices and needs of traditional and digital humanists, with respect to infrastructure work. Thus, they suggested five scholarly primitives: discovering, collecting, comparing, delivering and collaborating. Finally, Benardou *et al.* (2013) reported on work conducted in the context of the European projects DARIAH and EHRI, and they argued for the importance of understanding scholarly practices in order to provide research communities with appropriate tools and services. For that purpose, they suggested a user-centred qualitative approach – using semi-structured interviews and a conceptual model – which can be used for exploring the practices and methods employed by scholars among the various Arts and Humanities communities.

Building personal collections of information in the digital age

For the purposes of this project, single face-to-face and Skype interviews with twenty art historians were conducted. These took place from June 2013 to October 2013 and, in terms of format, they were semi-structured, based on an interview guide, while each of them lasted approximately

one hour. The interviews were recorded with the written consent of the participants and then transcribed using the Express Scribe software. Moreover, the interviewing phase included, when possible and with the interviewees' consent, observation of the interviewees' personal physical and/or digital collections. Again, when possible, gaining photographs of scholars' collections constituted a part of the observation process.

When the transcription phase was completed, and before the analysis stage, the transcripts were sent to the participants who confirmed the accuracy of their content. The analysis of the interviews was conducted using the NVivo software for qualitative research and the transcripts were coded according to a grounded theory approach. We started with axial coding, through identifying themes, categories and subcategories in the data as well as specifying their properties. Then, selective coding was considered necessary for creating links between the different themes and categories, and refining the codes until the issues raised during their analysis were of satisfying quality and depth, and could be used to answer the study's research questions. The analysis procedure was complemented by the memos (reflective notes) that were kept during the coding process, and the photographs and the notes taken during or after the observation which worked as a method for achieving triangulation, and thus, more accurate results.

Regarding the participants, at the time of the interviews, sixteen were based at UK institutions, while two scholars were based in Europe and another two outside of Europe. Eleven of the participants were female and nine of them male, whereas age was not a prime concern for this study; whenever is considered appropriate in this chapter, though, relevant comments will be made with regard to issues such as this and specific features of art historians' behaviour. Moreover, it is worth stating that their technical skills varied from advanced to basic. Also, the interviewees were at different career stages and, thus, they ranged from established academics to Ph.D. students, early career researchers as well as independent scholars. Based on the recruitment criteria, participants needed to actively conduct research or teach in the broader area of art history.

We were particularly interested in creating a pool of interviewees consisting of two groups: one where scholars worked on commonly studied areas (e.g. various areas of European art, like Renaissance art) or employed traditional art historical methods (e.g. stylistic analysis, historical investigation) and another where the topics examined (e.g. non-Western art, digital art) or the methods employed (e.g. quantitative, digital) were considered less traditional. Yet, it should be mentioned that this categorization was based on the premise that the practices of scholars in the first group (twelve scholars in this study) had been frequently examined by previous studies in the field while the behaviour and needs of those in the latter (eight scholars in this study) had been less studied before (also see Rose, 2002, p. 37). Identifying any similarities and differences between these two different groups of scholars could provide a better insight into the needs that art

historians in different areas of the field have in terms of resources, tools and services.

In the following section, we will present our results related to the information gathering and organization habits of art historians in an attempt to highlight some of the key issues these scholars face and suggest potential lines of action for facilitating the needs that occur. Furthermore, we should mention that our chosen framework for presenting our observations is based on Palmer *et al.*'s (2009, pp. 16–19) term of the activity of *collecting* which has two primitives: *gathering* and *organizing*.

Gathering information

Establishing personal collections is a common and much valued activity in art history that usually follows the seeking and discovery of information. In particular, our results showed that scholars' collections consisted of a wide variety of information objects. Since all of the participants kept both physical and digital personal collections, these information objects were either conventional or digital, and of textual, visual, audio-visual or multimedia format. Regarding textual resources, they ranged from monographs, various types of books, journals, magazines, newspapers and conference proceedings to dictionaries, photocopies and scans of various textual resources, correspondence, genealogical records, legal documents, bibliographic references, notes and mailing lists. On the other hand, the visual resources consisted of images found in printed, illustrated books, and exhibition catalogues of museums and galleries, including photocopies and scans, other two-dimensional kinds of material, such as 35mm slides, and digital images of art objects, manuscripts or other types of objects. Finally, scholars' collections included CD-ROMs, videos, interview recordings and 3D models and visualizations.

Generally, as art historians participating in this project told us, the beginning of research and the need to look for specific resources is usually driven by the topic under examination and the research questions. However, sometimes, it is an art object itself or its image found somewhere on the Web that marked the initiation of research and the path to be followed.

> Personally, I tend to start with objects or images. So, an interest will often be sprung by looking at an image – often online just because it's easy to access – either in an image library or normally a museum website. And then if it's in a museum environment I'm going to see it. Actually, often, I mainly just rely on images. . . . And then, I guess, I'll go to a library and try and learn more about it through books.
>
> (Participant 04)

Moreover, during these initial stages of research, scholars tended to gather as much information as they could find and would consider it relevant to

current and future projects (research or teaching) or to longstanding interests in a particular topic (also see Brockman *et al.*, 2001, p. 8). As one of our participants argued:

> So, in the beginning, I was capturing everything from administration documents to personal notes on important painters. I was capturing so much information that I didn't realize it. It wasn't until I was at trip twenty that I really knew exactly what I needed. And so the first ones were blunt work, were cannon blast, and the next ones were surgical incisions of the impression that I was getting.
>
> (Participant 05)

Yet, as the excerpt above demonstrates, a change was often identified in the intensity with which this practice was conducted when scholars' understanding of the research topic developed, affecting accordingly the amount of information gathered. More specifically, when they reached that point, scholars looked for particular information that would enable them to construct their research argument rather than material that would provide them with an overview of the subject. Consequently, this issue may suggest that art historians may have different information needs at different stages of the research process which, in turn, may affect the information-seeking practices and requirements they have in terms of resources and digital infrastructure to support their needs.

Continuing, an interesting aspect of the findings with regard to the places (digital or physical) that scholars visited to gather information was the list of reasons why they preferred certain resources over others. The most common factors included the degree to which the material that could be found at a certain resource was relevant to the research subject: convenience, which was mostly linked to issues of accessibility and availability; cost due to travel or copyright reasons; language; breadth and quality of material; and reliability of source, which was often determined by the trustworthiness of the provider (institution, publisher) or the author. Participant 10, below, provides an idea of the mixture of factors that influenced their decision when choosing particular resources.

> The primary factor for me was firstly its accessibility, how easy or hard would be to get, how much it would cost, the language it was in and whether I understood that language or some of my colleagues were able to assist me with that. So, primarily, I stick to resources that are in English, but I can also have resources that are in Italian, Spanish and French.
>
> (Participant 10)

Moreover, during the discussion around the main criteria upon which art historians in this study decided to use specific resources, it became apparent that some of the places visited (digital or physical) were more

frequently mentioned. Therefore, the British Library was one of the places most scholars mentioned visiting at some point, either physically or virtually, when looking for information. The fact that sixteen out of the twenty participants were based at UK institutions at the time of the interviews may explain this finding to some extent. Yet, convenience or familiarity were not the only factors for visiting this well-known institution; despite their location (London-based, UK-based, outside of the UK), several of the participants mentioned visiting the British Library physically, when possible, and electronically for the great amount and variety of resources available, such as manuscripts, newspapers, and so on. Other resources commonly used for research and teaching were the online collections of well-known museums such as the V&A (Victoria and Albert Museum, London) and the MET Museum (Metropolitan Museum of Art, New York).

> The Victoria and Albert Museum has a very good site where they actually have uploaded things like interviews with makers and designers. Their site is also an educational resource, so I use that.
>
> (Participant 19)

> I get quite frustrated if the images are low quality. I'll try and find something in as high quality as possible. Not least because I'll usually be using it in a presentation. . . . Now the MET Museum offers very high quality images for free, for academic purposes, which is great.
>
> (Participant 04)

Furthermore, with regard to the digital collections used by scholars in the field of art history, it may be worth noting that only one of the interviewees referred to any of the large, collaborative, European online initiatives that enable the discovery of primary and secondary material. In this one example, Participant 09 referred to Europeana Regia, a resource that provides access to manuscripts from the Middle Ages and the Renaissance; however, the brief commentary that follows suggests that the design of the resource might not fit the needs some of the art historians interested in manuscripts might have.

> The ways you can search are very limited and it's very rare actually to find a project that has been created with the end user in mind.
>
> (Participant 09)

Additionally, Participant 16, while describing their experience with resources containing digitized print material, gave an example of potential problems that can be encountered when using such a collection.

> I find that a lot of the stuff I look at it's pretty accessible. I mean, sometimes you're very dependent upon how a digitization has been done, so a lot of time if I'm looking at reports and engravings in a periodical via an online archive, the digitization might be of very bad quality.

> They are often very dark, so you can't see engravings very well or they miss out the engraving completely and you don't know whether that's because of the book that they're working from, if the plate has been taken out of it or they just missed it. So, you have things like that which are a bit of a pain. You always have to remember that there are holes in the online record.
>
> (Participant 16)

Hence, these observations may indicate that even when material is already available digitally or freely accessible, it might still not be useful to researchers in the field. It can also be argued that a problematic experience with a particular resource may be considered by scholars as an additional obstacle in their journey to find and gather useful information for their projects; as a result, it may influence their decision as to which resources to consult and how to consult them (physically or digitally). Then, if physically is regarded as the preferable way to consult a resource, our interviews and observations revealed that art historians would probably decide to take their own photographs in order to enrich their personal collections with material that meets their own criteria.

Actually, visiting a place in person and taking photographs is considered essential in the cases where access to resources is problematic. More specifically, apart from few disciplinary areas which benefit from greater availability of digital resources (as in the case of Participant 16 who researched on 19th-century British art), access is an issue that still concerns art history researchers. For instance, scholars dealing with primary material often encountered challenges in gathering the required information for their research since much of the material they were interested in was not digitized, in some cases was in private collections or even lost. Participant 05's account illustrates the previous argument.

> So, the constant battle that I'm under is whether or not it exists, whether or not it ever existed, and if it does exist, who has it and will they let me see it?
>
> (Participant 05)

Regarding especially visual material, access problems can be related to copyright and cost as well as to issues concerning the design of online resources. For example:

> But, for example – and I probably shouldn't say this on the recorder [laughs], but I don't care if they know – I had an image that was in a French museum's [name changed] collection and they wanted to charge me 300 euros. And so, I just said no way. And I went and photocopied it by scanning a book. Because it is a thesis, I have no qualms about that at all.
>
> (Participant 04)

> So, every time I went to the actual site, I would take a lot of photos and thus, I have got a lot of photographs of it before it was restored. They've just been restoring it over the last year. . . . So, I have them all in my computer and I do tend to use those. I guess if I was going into the field now I might actually think about trying to get familiar with the website, but I haven't because I know where my pictures are and I know what I've taken. The problem is you often want bits that aren't necessarily attractive pictures. . . . It's a more useful archive for me [the interviewee's collection], because actually the one they've got you on the Web hasn't a great detail on it. Because it's actually very slow, when all you want is to come through the sort of things and, I mean, it's actually difficult to work out where you're going. I don't find it that easy to use.
>
> (Participant 03)

Therefore, it becomes evident that getting access to research material is only one aspect of the problem in art history. Accessing resources that are truly useful to scholars is another. For instance, Participant 03's comment, above, suggested that there are certain needs in the field regarding types of visual material that are not covered yet, a problem that could potentially be resolved through collaboration between information professionals and scholars.

Although digitization and infrastructure projects have come a long way over the last few decades, there are still things to be done not only for enhancing access to research material, but for presenting this material in a way that is useful and easy to find (as Participants 09's and 16's comments also indicate). Especially when it comes to accessing visual resources online, it is important for institutions, such as museums (which constitute the principal information providers in the field), to employ a more user-centred approach in terms of website design and facilities provided. Finally, given the interdisciplinary nature of art historical subjects, there may be cases where useful resources already exist but are not currently known to scholars (e.g. in the case of Participant 03, architectural resources might have been useful). In such circumstances, collaboration between scholars and content providers as well as better communication avenues for reaching researchers might help clarify such matters and improve the discovery and use of existing resources.

Organizing information

The building of personal collections of information in art history not only constitutes an ideal choice for dealing with various issues regarding research and teaching in the field, but also for creating collections that can be indexed and organized upon personal criteria. However, as collections grow, the need for their management emerges (also see Borgman, 2003, p. 3; Meho and Tibbo, 2003, pp. 582–583; Palmer *et al.*, 2009, p. 18;

Beaudoin and Brady, 2011, p. 32). In art history, especially, this can be a very challenging task due to the variety of formats and the personalized practices involved.

Based on our findings, although all of the scholars participating kept both physical and digital collections, there was a growing tendency to store and organize the collected research material digitally; the most popular reasons given were convenience and easy transfer of data. However, in some cases, this tendency was not necessarily in agreement with the scholars' preference over the digital or physical part of their collection. This was especially evident in the responses given regarding the potential sharing of resources – for example:

> I'll send electronic files to anybody who wants them. . . . And other people have done that for me as well and no problem there at all. Books, it's a completely different matter for me. My books are mine, I don't lend them; I don't lend them to students, I don't lend them to colleagues.
>
> (Participant 11)

Generally speaking, personal collections were found to play an important role during research – for example, scholars would go back to the gathered resources when they needed to refresh their memory or rethink and re-evaluate the scholarly argument. However, what seemed to be common practice was revisiting material from their collections during the writing stage – for instance:

> I guess, when I'm writing something, I just get out a pile of books and use that in conjunction with resources on my computer.
>
> (Participant 20)

Furthermore, personal criteria determined the way that personal collections, both physical and digital, were organized and managed; some examples included organizing the files by project, labelling them by author or just keeping the label provided by a digital resource, such as an online library. Also, a common practice among scholars who were teaching at the time was to separate the material used in class from material used for research.

Concerning the 'toolkit' of the art historians in this study, it typically included several means for storing, filing and organizing information. These ranged from more traditional means, such as physical folders and filing cabinets, to digital tools and services, such as digital filing systems, note-taking software (e.g. Evernote), reference and bibliographic tools (e.g. Zotero, Endnote), Microsoft Office applications and other smartphone and tablet applications (e.g. Feedly, DoggCatcher). Dedicated tools for the management of images, like FileMaker Pro, Mac's iPhoto or Google's Picasa, were considered to be of great importance, even by the

participants who did not use one at the time, since researchers had to deal with a great amount of visual information which often constituted a big challenge for them. Moreover, it is worth noting that the tools scholars tended to use mostly were those suggested either by their colleagues or their institutions (also see Warwick *et al.*, 2008), while they were more likely to use a tool when given at the start of their project and when receiving relevant training and support.

Yet, some scholars appeared to face difficulties that could not be tackled easily by using a particular tool, since they were not aware of one that could fit their needs exactly. According to the quotes below, the first scholar researching on non-Western art found it challenging to keep together the related visual and textual information needed for their research while the second, a digital art historian, faced problems due to the ephemeral nature of digital information. As a result, both cases illustrate the need for a better understanding of art historians' practices whose subjects do not belong to the traditionally researched areas, like Western art.

> I need to find a method for saving the picture together with the transcription. It's no use for me if I can't read them both together. And the only way I've been able to do it is by using Power Point. Power Point is quite good for me.
>
> (Participant 08)

> I've migrated across several different platforms in terms of how I stored various things and then there's always a new way of working; there's always a slightly different way to keep art which is difficult. Quite simply, some other things I end up copying and pasting into Word documents and adding as many links as I possibly can, that means I should be able to find that content again if it's migrated to different platform or the link breaks later on. But part of the nature of a lot of things that I've spent my time researching is that I won't be able to find them maybe in a few years' time anyway. So, that's why I'm trying to contextualize them now so there's at least some meta-narrative I guess on what was there even if it's not there later on.
>
> (Participant 17)

An aspect of the discussion surrounding the information practices of art historians in this study which is worth highlighting is the role that visual memory played in the organization of the gathered material (also see Brilliant, 1988). Some of the scholars, in particular, argued that their organization systems reflected the way their mind works and as a result it was easier for them to locate the information they needed. Additionally, organizing information in specific ways could constitute a source of inspiration. For example, based on our observation of scholars' workspaces, one of the interviewees (Participant 12) was accustomed to printing out

images related to the topic under examination and pin them on a board over their desk, while they would rearrange them when needed. Looking at the pictures and creating themes through organizing them in specific ways worked as a source of inspiration during research; yet, our interviewee could only follow this practice by using conventional material, since they were not aware of any related tools that would enable them to work in the same intuitive way.

Actually, there were several cases where scholars preferred to use conventional formats when carrying out specific tasks (e.g. for annotations) or for organizing information (as described above). Yet, it is worth mentioning that, in this study, there was no strong relationship that could be identified between the preference for printed formats and the type of research conducted (traditional or non-traditional), the career level (junior or senior scholar), the background of the art historians (two of the participants with a background in the Sciences expressed their preference for printed formats while one highlighted their difficulty in dealing with technology) or age (younger scholars often had large collections of printed material or noted facing challenges with regard to the use of digital tools and services).

Regarding age, in particular, it may be useful to note that there have been studies arguing that, eventually, the employment of ICT tools will increase as a new generation of Arts and Humanities scholars takes over (e.g. Wiberley and Jones 2000; Barret 2005); this assumption was based on the premise that younger scholars tended to adopt digital technologies more readily. However, more recent studies (e.g. Rowlands *et al.* 2008; de Jong *et al.*, 2011; Warwick, 2012) have highlighted that issues such as lack of technical skills or low employment of digital tools and services can still be encountered among Ph.D. students and younger scholars, making it difficult to predict with certainty the future use of digital technologies in the Arts and Humanities disciplines.

On the other hand, the flexibility that conventional formats offered for interacting with the material in creative and, often, visual ways tended to be one of the main reasons when choosing to work with paper resources. This brings to mind Rose's (2002, p. 38) finding about the preference that art historians in her study showed for paper systems since computer screens did not allow them to work in the way that conventional formats did – an interesting issue considering that, to a great extent, is still true after more than ten years since Rose's study took place. Hence, as visual memory and creativity in terms of interaction with information was found to constitute a significant part of art historians' work conducted 'behind the scenes', at their personal workspace, we believe that it should be taken into account when designing tools and services to facilitate research in the field. Similarly, further research should be conducted in order to identify the extent to which these qualities affect other information practices, such as information seeking, and how these can be integrated to digital infrastructure work.

Finally, scholars in this research wished especially for better ways to combine textual and visual information and for applications that would enable them to manage visual information according to their needs. Some, also, wished for a tool that would allow them to know what they already have in their collection, since it was sometimes difficult to remember all the material they had gathered over the years; such a tool would help them use the collected material effectively in their research and teaching. In fact, it is pertinent to say that these wishes are highly in agreement with the issues stressed so far, emphasizing the importance of information organization and management in conducting efficiently other core scholarly practices. Therefore, it is worth considering how information professionals could assist art historians in carrying out such practices while scholars' requirements and concerns should be taken into account when building digital resources and tools for facilitating research in the field.

Conclusion

Our aim in this chapter was to use art historians' personal information collections as a departure point for looking at the practices that often follow the discovery of information in the field. More specifically, we argued that these practices, which are conducted 'behind the scenes', can be especially challenging to study; however, as they play a key role in the research process, they should be taken into account when building digital infrastructure and tools for research purposes. As a result, we presented our study's findings based on Palmer *et al.*'s (2009) term of the collecting activity with two primitives – gathering and organizing.

In particular, we had the chance to stress some of the issues that scholars participating in our research encountered during the gathering and organization of information and how these practices affected or related to other scholarly activities. In fact, some of these problems seemed to be associated with larger issues, such as information access in the digital age, while others related to the requirements of specific projects. Moreover, we looked at the tools that scholars employed during these practices and presented their wishes with regard to digital services that could fit their needs. Finally, based on our results, we pointed out issues that could be of interest to digital infrastructure stakeholders, such as scholars' criteria when choosing resources to use and key aspects of their information practices, like the role of memory and the creative aspect of art historians' work, that should be considered when creating digital services and research environments for enhancing digital scholarship.

To conclude, we strongly believe that the provision of tools and services that fit art historians' needs with regard to information use and management has the potential of facilitating the research process as a whole. However, we suggest that further study focusing on these practices is needed;

as scholars will increasingly collect or create more information, one of the emergent issues to occur in the immediate future will be the necessity to manage and curate large amounts of digital data. Thus, only by developing a sound understanding of how these practices evolve, along with the challenges scholars face, can we effectively facilitate scholarship in the field with relevant and user-centred digital resources and services.

References

Barrett, A., 2005. The information-seeking habits of graduate student researchers in the humanities. *The Journal of Academic Librarianship*, 31(4), pp. 324–331.

Beaudoin, J. E. and Brady, J. E., 2011. Finding visual information: A study of image resources used by archaeologists, architects, art historians, and artists. *Art Documentation: Journal of the Art Libraries Society of North America*, 30(2), pp. 24–36.

Benardou, A., Constantopoulos, P. and Dallas, C., 2013. An approach to analyzing working practices of research communities in the humanities. *International Journal of Humanities and Arts Computing*, 7(1/2), pp. 105–27.

Blanke, T. and Hedges, M., 2013. Scholarly primitives: Building institutional infrastructure for humanities e-Science. *Future Generation Computer Systems*, 29(2), pp. 654–661.

Borgman, C. L., 2003. Personal digital libraries: Creating individual spaces for innovation. *NSF Workshop on Post-Digital Libraries Initiative Directions*. Chatham, MA, September 2005.

Brilliant, R., 1988. How an art historian connects art objects and information. In: Stam, D. C. and Giral, A., eds. 1988. *Linking Art Objects and Art Information. Library Trends*, 37(2), pp. 120–29.

Brockman, W. S., Neumann, L., Palmer, C. L. and Tidline, T. J., 2001. *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, DC: Digital Library Federation and Council on Library and Information Resources.

de Jong, F. M. G., Ordelman, R. J. F. and Scagliola, S., 2011. Audio-visual collections and the user needs of scholars in the humanities: A case for co-development. In: *Proceedings of the 2nd Conference on Supporting Digital Humanities (SDH 2011)*, 17–18 November. Copenhagen, Denmark. Copenhagen: Centre for Language Technology.

Ellis, D., 1993. Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4), pp. 469–486.

Europeana Regia. (online) Available at: www.europeanaregia.eu/en (accessed 20 August 2015).

Larkin, C., 2010. Looking to the future while learning from the past: Information seeking in the visual arts. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1), pp. 49–60.

Long, M. and Schonfeld, R. C., 2014. *Supporting the Changing Research Practices of Art Historians*. ITHAKA S+R. (online) Available at: www.sr.ithaka.org/research-publications/supporting-changing-research-practices-art-historians (accessed 1 May 2014).

Meho, L. I. and Tibbo, H. R., 2003. Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, 54(6), pp. 580–587.

Palmer, C. L., Teffeau, L. C. and Pirmann, C. M., 2009. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Dublin, OH: OCLC Online Computer Library Center (online). Available at: www.oclc.org/programs/publications/reports/2009-02.pdf (accessed 16 January 2010).

Rose, T., 2002. Technology's impact on the information-seeking behavior of art historians. *Art Documentation: Journal of the Art Libraries Society of North America*, 21(2), pp. 35–42.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H. R., Dobrowolski, T. and Tenopir, C., 2008. The Google generation: The information behaviour of the researcher of the future. *Aslib Proceedings*, 60(4), pp. 290–310.

Stam, D. C., 1997. How art historians look for information. *Art Documentation: Journal of the Art Libraries Society of North America*, 16(2), pp. 27–30.

Unsworth, J., 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? *Humanities Computing: Formal Methods, Experimental Practice*. King's College London, 13 May 2000 (online). Available at: http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html (accessed 10 January 2013).

Warwick, C., 2012. Studying users in digital humanities. In: Warwick, C., Terras, M. and Nyhan, J. (eds.), *Digital Humanities in Practice*. London: Facet Publishing, pp. 1–21.

Warwick, C., Terras, M., Galina, I., Huntington, P. and Pappa, N., 2008. Library and information resources and users of digital resources in the humanities. *Program: Electronic Library and Information Systems*, 42(1), pp. 5–27 (online). Available at: http://dx.doi.org/10.1108/00330330810851555 (accessed 23 May 2012).

Wiberley, S. and Jones, W. G., 2000. Time and technology: A decade-long look at humanists' use of electronic information technology. *College and Research Libraries*, 61(5), pp. 421–431.

7 Making sure the data fit the researchers

Data identification and investigation in European Holocaust Research Infrastructure (EHRI)

Veerle Vanden Daelen

Introduction

The European Holocaust Research Infrastructure (EHRI) is a Horizon 2020 project that started in October 2010 as a FP7 project (7th Framework Programme for Research and Technological Development) of the European Union. EHRI's main objective is to support the Holocaust research community by opening up a portal that gives online access to dispersed sources relating to the Holocaust, and by encouraging collaborative research through the development of tools. 'By integrating currently fragmented archives online and streamlining the access to these information resources, the EHRI project aims to facilitate research into one of the most traumatic aspects of Europe's shared history: the Holocaust' (EHRI, 2013).

Even though many archives have started digitizing the archival materials, the lack of metadata describing the materials and the sharing of all this information beyond the archival institution itself, remains a huge challenge: 'The difficulties lie in the sheer amount of information, the range of cataloguing approaches and technologies, the complexities of interfacing divergent database structures and the disparate nature of the material, which includes papers, photos, films and art' (EHRI, 2013). In July 2015, the EHRI portal contained information on 57 countries, over 1,800 collection-holding institutions and over 150,000 archival descriptions in 465 institutions. While EHRI's first phase under FP7-funding (2010–2015) concluded with the presentation of the EHRI portal in Berlin in March 2015, the project continues in its efforts until at least April 2019 under Horizon 2020. The data and use case described in this article reflect the work of EHRI's first phase under FP7.

How to bring content to the portal was and is a big challenge for EHRI. Because of the nature of the event, Holocaust-related sources are spread worldwide, there have been attempts to destroy the evidence, the Holocaust survivors have migrated and there exist a multiplicity of documentation projects since the events. Given these facts, it was sheer impossible to map everything within the first four years of funding of the project.

Therefore, the identification and investigation work bringing the content to the portal needed a clear focus (as even so, it remained a huge area of research). This contribution presents a geographically focused case study – namely, on Holocaust sources in Greece.

The data integration and investigation work was in this particular case closely linked to user requirements as it built upon the contacts and information from the EHRI user requirements' interviews with Greek Holocaust scholars. It is our premise that attention to user requirements should be reflected in all work of the project, especially in data identification and investigation, the content brought into the portals and infrastructures.

EHRI's identification and investigation activities

A considerable part of the content of the EHRI Virtual Research Environment (VRE) is the identified Holocaust-related material. The ICA-standard compliant EHRI-customized database brings together EHRI terms and authorities, country reports, descriptions of collection-holding archival institutions, and archival collections. Next to this, EHRI also developed research guides and online teaching courses, and publishes newsletters next to information on the project shared via its website and Twitter account. The objectives of the 'Identification & Investigation' work package in EHRI's first phase as an FP7 were outlined as follows:

> 1) To identify Holocaust-related collections and archives, existing services and infrastructures; 2) To draw up a list of all institutions where such collections and archives are located, with some basic information on each of these institutions and their collections; 3) To develop criteria to enable selection of key collections, infrastructures and services and to apply these to all investigated collections, infrastructures and services; 4) To create a functional structure of a database integrating collections and services.
>
> (EHRI, 2011a, p. 44)

The identification work can be carried out directly, via institutions that hold these collections, or indirectly, via aggregators, and other finding aids. However, investigation work needs a framework, both content- and technical-related. As to the content-related decisions, EHRI worked on a working definition of the Holocaust to guide active surveying. These guidelines include a time frame, a geographical focus and a content-related focus. The technical-related guidelines concerned the working database and the metadata standards. All of these aspects were discussed and agreed upon at the early stages of the project (EHRI, 2012).

The starting point for the identification of repositories was the Directory of Holocaust-Related Archives (Conference on Jewish Material Claims Against Germany, 2012). This list was completed with the *Guide des archives*

sur la Shoah (*Mémorial de la Shoah*, 2000), an overview list of institutions with which Yad Vashem worked together and has copied archival material from, the online catalogue of the United States Holocaust Memorial Museum (USHMM, 2013), and specific additions from national archival guides, experts, and published studies on the subject (EHRI, 2011b).

In order to structure the EHRI inventory of Holocaust-related archival institutions and collections, EHRI developed country reports (Bennett *et al.*, 2015). They give an overview of the Second World War and Holocaust history as well as of the archival situation in the covered countries, and include a section on EHRI's research. To make the country reports easily searchable and as standardized as possible, they all follow the same general structure. The first paragraph of the history section deals with the history of the country during the Second World War, covering questions of statehood as well as German rule and influence. The second paragraph focuses on Holocaust history and includes information on the size of the pre-war Jewish community as compared to the total population of the country, as well as an estimate of the number of Jewish victims in the country.

The second section of the reports presents a concise overview of the archival situation. A first paragraph deals with the archival culture of the country: how the archives are organized (centralized system or not; public and private archives; general information about access, etc.). The second paragraph gives more information on which archives are most relevant for Holocaust research. When available and known to the project, the report also gives references to aggregators and research guides on the country.

The country reports are fully integrated and connected to the identified archival institutions and collections in the portal. Their ultimate goal is to offer the researcher a tool within the VRE which explains in broad lines the Holocaust history of a country, which relevant Holocaust-related archives are located in the country and which research EHRI already carried out. It is the combination of these aspects and the concise and structured format that make the reports a unique tool. At the time of its Berlin presentation in March 2015, EHRI has published reports for the Axis and Nazi-occupied countries in Europe (including the North-African colonies),[1] plus Israel, Portugal, Spain, Sweden, Switzerland and the United States (as these countries also hold important Holocaust-related archival collections). Ideally, every country that holds Holocaust-related archival collections should receive a country report, and the reports are intended as a dynamic, growing resource: new reports can be added, and existing ones can be updates and annotated. The reports have been published as work-in-progress on to the EHRI portal, and are already accepting annotations and comments by registered users.

When looking at such a vast geographic area, different languages, different kinds of archival sources and repositories, there is a need to focus in the further identification work. This focus can be, for example, on the type of repositories or collection-holding institutions (e.g. state or national

archives, memorial sites, private archives), type of source (e.g. document, photo, film, audio, object, ego-document), or thematic approaches (e.g. ego-documents, looting of Jewish assets), but also looking into existing aggregators, surveyors, finding aids, or other projects. EHRI has focused on geographic approaches, such as, for example, on Ukraine or Poland, and the Greek case in this chapter also falls under this category. However, this particular case came into being via EHRI's user requirements work.

EHRI's linking user requirements to the content

In its case study on the Holocaust in Greece, EHRI has organized its identification and investigation work for data integration on the work carried out by the 'user requirements' work package. This work package, among other things, held researcher interviews with both junior and senior scholars conducting advanced research in the field of Holocaust studies. The aim of these interviews was to identify and analyse scholarly research practices and use of archival materials in the area of Holocaust studies, in order to analyse how digital methods and approaches could be used to support and enhance these practices (EHRI, 2011a, pp. 47–48). All user requirement interviews were conducted between September 2011 and June 2012 (Angelis *et al.*, 2013, p. 17). Even though these interviews were processed to create a set of data and functional requirements for the VRE, they also brought in names of repositories that were not yet included into the repository database. This made for the decision to further engage with the interviewees, especially with the four researchers on the Holocaust in Greece.

Maintaining and expanding the contacts with the user requirements interviewees seemed a novel and productive way forward, not only for the sake of the data this could bring in, but also for a general engagement with the research community. The importance of involving the researcher in the technical requirements has been stressed very explicitly and rightly so:

> To ensure usability, VREs need to be focused on the needs of researchers and specific research communities, putting them in the driving seat of VRE development . . . The development of a VRE needs to be broached not as a technological project but as a community building project, since without this community buy-in, the VRE cannot fulfil its function. Community outreach beyond the initial community is also essential for the future sustainability of VRE projects, with members feeding new applications, new content, and the social context that allows for effective use of data and other resources, back into the VRE.
>
> (Carusi and Reimer, 2010, pp. 5–6)

> The best projects have approached user engagement as a lifecycle process taking place before, during and after the creation of a digital resource and informed by a strategic approach including a range of

> activities aimed at soliciting users' input in a two-way exchange, where users shape, at various points, and in various degrees, the development and creation of a digital resource.
>
> (Marchionni, 2009)

Both data integration and user requirements are important, and both require contact with the target audience and users. By linking both aspects, EHRI wished to ensure that the EHRI portal would not only attract historians, archivists and the general public for its functionalities, but equally for its content, that the portal contained the (type of) data the researchers hope to find in this VRE, in order to ensure greater interest via the right content in the right format. This led to our premise that linking these more technical requirements to the actual data integration (meaning the data which are being brought into these infrastructures) makes for a better adjustment of the content to the format and vice versa, and henceforth for more successful – because more adapted to the users – digital research infrastructures. The two aspects (content and technical) clearly need to be intertwined: '[A VRE] requires the participation of scholars, who combine methodological and technical skills (interfacing between humanistic research and information technology). The key is research-driven development' (DARIAH-DE, p. 21). Indeed, 'the importance of not building a VRE for, but with the researchers cannot be emphasised enough' (Carusi and Reimer, 2010, p. 24). And since VREs 'always build on presumptions of what the user will be looking for' (DARIAH-DE, 2013, p. 23), we wanted to make sure and check them with the users.

The researchers of the Holocaust in Greece interviewed by EHRI's user requirements Work Package were asked to suggest further researchers specialized in Holocaust history in Greece to all participate in an EHRI workshop on data integration and Holocaust research in Greece which took place in Athens on 4 December 2012. It is important to note that for two user requirement interviewees who participated in the workshop, this was their second in-person contact with EHRI; the other two had close colleagues attending the workshop, so this was also a tool not only to 'set up' a relationship with users, but also to 'nurture it throughout in an interactive fashion' (Marchionni, 2009). Four representatives of repositories in Greece, three graduate students, two junior and two senior researchers joined for the December meeting.[2]

Next to the connection to user requirements, other important reasons to focus on the Greek case study were 1) the historical reasons – the size of the pre-war Jewish community (about 80,000) and the number of victims (over 80 per cent) – and 2) the language and the transcription, making the integration of Greek material a challenge. Further complicating the search for archival sources are the multiple occupational zones in war-time Greece (Italian, German and Bulgarian) and the two armed forces (the British and the Red Army) entering the country in 1944, which make that sources on the Holocaust in Greece are spread over multiple countries; Greece

itself and the occupying and liberating countries' archives (Bennett *et al.*, 2015). Large Holocaust aggregators' archives (Yad Vashem and the United States Holocaust Memorial Museum) have already brought much material together, but there is still much work to do.

Because of changes in boundaries since the Second World War until today, and because the current borders define the different archival systems, EHRI works within the framework of current national borders. Therefore, this Greek test case focused on repositories and their collections in present-day Greece, leaving out the sources on Greece in other countries. Before meeting with the researchers, EHRI was aware of ten repositories (institutions that preserve Holocaust-related archival materials). The purpose of the workshop was to receive expert feedback on this list, to further complete EHRI's overview of repositories and to identify Holocaust-related collections within these institutions.

EHRI also needed further understanding of the format of the metadata and their availability. The project also wished to determine whether and where further surveying was necessary (as well as which type of surveying). Altogether, the goal was to create a state of the art report of identified Holocaust-related sources in Greece, and to bring this information into the EHRI portal, as much as possible via the technical work packages (as opposed to manually entering the information into a database).

The workshop was an eye-opener: instead of discussing digital data, their format and their integration into the EHRI database, it soon became clear that the main issue and challenge for Holocaust research in Greece was access. As the following overview indicates, we were far from discussing merging digital (meta)data from different sources, the more-or-less expected issue when talking about developing a VRE. To bring the right content to the EHRI portal and to be relevant to the user community in Greece, interacting with the workshop, participants learned that EHRI had to invest its time on identification of the sources and bring them together for the first time in a digital, easy-searchable online platform. Only by doing so could EHRI really play a role in meeting the needs of the researchers.

User input on digital data

Only a very limited number of institutions share descriptions of their holdings online. One of them is the Hellenic Literary and Historical Archive, better known as E.L.I.A. (part of the National Bank of Greece Cultural Foundation, M.I.E.T.), which has an extensive and international metadata standards compatible Greek language catalogue online available on their own website (E.L.I.A., 2013). This archive is one of the only ones that could be used quickly and effectively to help put a VRE together.

A second, but different example is the Jewish Museum of Greece (2013), in which a Holocaust collection consisting of more than 2,000 artefacts, comprising books and pamphlets, periodicals, maps, documents, manuscripts

(personal notes, correspondence, memoires, diaries, etc.), false IDs, personal objects, ritual objects, jewellery, textiles and costumes, clothing, badges, armbands, numismatics (coins, banknotes, stock certificates) and an art collection. The Holocaust Archive consists of the Y.D.I.P. archive (Yperesia Diatheseos Israelitikon Periousion, Central Agency for the Custody of Jewish Property), the O.P.A.I.E. archive (The Heirless Property and Jewish Rehabilitation Fund), private and personal archives, which include material pertinent to the Shoah, the JMG audiovisual archive of oral histories, and personal testimonies and the photographic archive. The Holocaust Photographic Archive is a rich and expanding collection of photographic images spanning from the pre-war period to the early 1950s. The broad subject areas, each including several subcategories, are Jewish life before, during and after the Second World War and the Holocaust. An extensive amount of document scans and photographs of the museum's collection have been uploaded in Judaica Europeana (Judaica Europeana, 2013).

The workshop participants also drew EHRI's attention to the Database of Greek–Jewish Holocaust Survivors' Testimonies (2013), a project led by Professor Rika Benveniste at the University of Thessaly. This database assembles all the audio as well as the visual testimonies from Greek Jews who survived the Holocaust. The database systematically brings together testimonies from repositories in Greece and abroad, public and private collections, collected in different languages and places, at different times, forming thus a virtual archive. Next to the archives with online finding aids, there are a couple of institutions with a high degree of digitization (such as the Parliament Library or the Karamanlis Foundation).

Understanding how to integrate Holocaust-related sources from the General State Archives in Greece

The largest institution for EHRI's purposes in Greece is the General State Archives (national archives of Greece, GSA). They are the national institution responsible for the preservation and promotion of the Greek archival materials. The institution was created in 1914 and their operation defined by law. EHRI had one central entry for the GSA in its database before the workshop, but it immediately became clear that this was not sufficient as the Greek State Archives consist of a central service, 47 regional services and 16 local archives. Based on the location of the wartime Jewish communities in Greece (in at least 27 cities according to the Central Board of Jewish Communities of Greece, 2009), selections had to be made into which branches of the GSA had to be added to the portal.

This was not a straightforward procedure as the selected archives often required on-site surveys, as was also outlined during the workshop by EHRI's target audience, the researchers. There is a digital finding aid, Archeiomnimon (2013), 'Memory of the archives' (containing about seven million pages of documents), which allows navigation through the archival

collections originating from 37 agencies of the General State Archives (the central service and 36 branches). The metadata fields in Archeiomnimon are provided in English, the actual descriptions are in Greek. However, not all branches have descriptions on this digital finding aid yet, and also for the branches that do have data in this online system not all materials have been digitally described so far. Also, not all the digitized materials have already been included in Archeiomnimon. This is, among other things, due to lack of resources (for scanning, uploading and adding adequate metadata). For the agencies and materials that are missing, the holdings either remain uncatalogued, or there are published Greek-language finding aids or index cards in filing cabinets in the reading rooms of the respective branches. The latter, however, is less standardized than the online database. The use of standardized software for encoding (meta-)data clearly has proven its benefits:

> There is a crucial difference between the time before and after in the digital collections. Before, we had real difficulties to find the same language in cataloguing the archives, between the central office and the branches. But, after the digital collections, I think there is a new era for the archives. In our software, we follow clearly the international system, and I think it's only the state archives, E.L.I.A., and the archives of our National Bank that follow this system in Greece, as far as I know. And now, thanks to Archeiomnimon, we have a new database and we gathered all the information from the level of the collection to the level of the file and that will be accessible online, in the beginning in Greek.

Hence, we can expect a further standardization as more and more (meta-) data are encoded in standardized software. How much material in the GSA remains non-digital and non-described is hard to estimate. Workshop attendees explained that the 19th-century materials have been completely catalogued and that the degree of cataloguing of 20th-century materials – which are the core focus of EHRI – remains limited. This is somewhat linked to the history of the institution. As was said during the workshop, the GSA 'was founded in 1914 and then the emphasis was on cataloguing and providing access to the material on the big war of independence, the first years of Greek statehood. Then there was a long period of war etc.' Hence, the cataloguing department still is catching up with this backlog.

So, how can EHRI at this time identify the right materials for its project? During the workshop, we learned that in March 2008, all branches of the GSA had been asked to provide an overview of collections and documents related to the history of the Jews in Greece and the Holocaust in order to prepare the survey work of the United States Holocaust Memorial Museum (USHMM). The responses were rather limited: 15–20 local offices submitted such a report, which varied from a short note confirming there is such

material available to an overview of records. Nevertheless, this list served as a first orientation tool for EHRI's surveying. Having a representative of the USHMM survey team in Greece present at the workshop furthermore ensured EHRI of a smooth, personal communication with a key player in this field.

The most important challenges for Holocaust research in Greece

The workshop participants raised EHRI's awareness about several challenges for data identification of sources on the Holocaust in Greece. A first challenge is that finding aids are often only consultable at the repository itself. The limited number of online information and locally available cataloguing systems make access to the actual materials time-consuming and often costly and inefficient (a research trip is hard to prepare). A major challenge in Greece is that there is an unknown number of materials that remain totally invisible because there is no indication about their existence online and because there is the huge challenge of the (digital) description backlog which requires a thorough identifying and inventorying campaign. Uncatalogued materials remain hidden. Whether one receives information and what type of information often has a 'serendipity factor'. While some researchers signalled 'precious and unlimited help', others struggled to receive information.

Receiving permission for consultation, for example, can be a long process. Access to public archives in Greece often requires an application procedure before any access is granted. Two laws regulate the protection of individuals with regard to the processing of personal and sensitive archives (1991 and 1997). The time limit for keeping records restricted was reduced from 50 to 30 years (although there is room for flexibility in certain cases). In some cases, it is not possible or there are limits to take photographs, scans or photocopies. Sometimes, lack of capacity of the reading room is an issue.

In one archive, the Archives of the Ministry of Foreign Affairs, where not all archives can be made publicly accessible (for reasons of national security among others), the documents concerning the Holocaust have been removed from the original collections and have been put in one separate collection. This causes problems to the researchers as they lose the context of the original files. Researchers need to see the other materials to fully 'place' the sources on their research subject in the right context:

> You can find separate pages from a single file, like page 4–5–6 and nothing more than that, pages about Jews and the other [pages] were about other minorities [and therefore not included]. So you don't have the dates . . . when you think like a historian, you want to have the context, you want to have the whole collection so you can understand how to integrate these cases into the whole story.

Another issue, signalled by the workshop participants, pertains to the destruction of archives. The destruction of archives does not only apply to Holocaust archives, but did include sources on the topic. For example, the archives of the Greek Agency for War Crimes, which was founded in the autumn of 1945, were totally destroyed in 1975. In some cases, materials have been scanned before destruction. The Jewish Community of Thessaloniki, which engages in such digitization projects, has preserved certain sources.

Yet another challenge is that much material is kept with private organizations who don't necessarily consider themselves archival institutions. an archival institution. These archives are hard to locate and often remain 'hidden treasures'. During the workshop, the researchers signalled that local Jewish communities 'don't know what treasures they have hidden' or that they 'don't even know they have an archive'. The researchers passing by and asking, opening up boxes in little rooms, discovering sources of historical importance, are a first step to raising awareness. It often leads to the opening up of the sources, cataloguing and improving of conservation conditions. Having some critical knowledge on the content of the archives, as well as having some general procedures for consultation, was suggested to help improve access to the sources. Being able to share the knowledge and sheer existence of these hidden treasures by entering or annotating them to the EHRI portal was put forward as a first major step forwards for Greek Holocaust research, especially for the local level of community, the Jewish community archives.

Meeting with the researchers made us aware of many other archives and the search for still others. There is, for example, the search for police and army archives, of which the location or whether they have been kept to begin with, remains unknown until now. The same goes for sources on two concentration camps in Greece where Jews were imprisoned, but where archival documentation on their presence has not been not yet been indentified. Archives of the Chamber of Commerce, the Court House and other juridical archives, municipal archives, church archives (for example, the archives of the Holy Synod of the Orthodox Church), Communist Party Archives, Red Cross Archives, public and private company archives (electricity, water, port companies), and private collections kept in museums, at home or in private institutions, are yet another mixture of private and public archives that remain hidden or closed. A participant explained:

> My experience is that you apply and they tell you they don't have [an archive], and then you have to go and prove them that they have [one]. Usually, they publish a book about 50 years of the association . . . so when you . . . say: 'I spoke with this professor who wrote your book', they say: 'Oh, it's in the second basement to the left'.

Even though it may sound easy to simply include all municipal archives of cities with a Jewish community at the time of the Second World War,

the expertise of the workshop participants pointed EHRI to the fact that in the Bulgarian-occupied zone demographic records of some towns have been transferred to the local branch of the GSA. In addition to identifying repositories that need to be included, the workshop participants also informed EHRI about other archives that had been checked because they were thought to contain Holocaust-related materials, but in fact do not, as was established by the experts.

Implications and conclusions

The one-day workshop was a genuine success. It allowed the project to collect a huge amount of expert information that would have been very hard to collect if EHRI had not directly contacted the Greek Holocaust research community. This input allowed EHRI to grasp the situation and understand the implications on its work to support (Holocaust) research in Greece efficiently and effectively. A first conclusion is that the content-provider, i.e. the collection-holding institutions and their administrations, clearly influences the research. Bureaucratic procedures can cause delays and there can be a problem of access which is more personal-related (personal relationships influence what you can see and what you can't access). The limited number of online information and locally available cataloguing systems makes access to the actual materials time-consuming and often costly and inefficient. And then there remains the – impossible to estimate – amount of materials that remains totally invisible because there is no indication about their existence, not in the digital or in the analogue (metadata) world.

Based on this input on challenges of access to information on Holocaust-collections in Greece, EHRI had to revise its strategy about data integration. Indeed, most of the discussions about VREs are about how to bring the digital material into the VRE and how to match the different digital/technical standards. This, however, would only cover a limited part of the Holocaust-related sources in Greece considering that the formats of the data are clearly not living up to this expectation (PDF, Word docs, paper-printed, index cards, or no inventory at all). However, this did not mean that a VRE and a project like EHRI could not step in and support Holocaust-research in Greece into a high degree. Access to material was identified as one of the key needs in the user requirement interviews (Angelis *et al.*, 2013, pp. 58–60). This access to material, be it awareness and location of repositories or identifying relevant collections within them, has been strongly supported by EHRI, not only by bringing in as much information as possible during the project, but also by engaging with the research community and the collection-holding institutions.

Considering the positive and constructive atmosphere during the workshop, a kind of material-sharing 'collaboratory' could work (see Anderson *et al.*, 2013). Researchers may not share their work notes or their personal databases made out of the material, but when talking about identifying

repositories and collections within them, researchers and representatives of repositories are indeed 'highly motivated to access and share data' (Carusi and Reimer, 2010, pp. 19; see also Angelis *et al.*, 2013, pp. 62–64). On the EHRI portal, a researcher can, for example, add a repository and include information about its collections or a publication based on these archival materials in an annotation. As such, 'the results of one search may, taken together with other information available to a researcher, influence the questions that are asked of others. Importantly, this may be a collaborative activity, requiring the ability to record interpretation and opinion as annotations, and thus start a dialogue within the relevant community' (Blanke *et al.*, 2010, p. 3821). Merging this kind of information will make further research less dependent on individual archivists and 'who is asking whom'. A second important role for EHRI became to reach out to the institutions and offer them tools and a platform to share the information about their holdings to the research community. It was especially rewarding to hear the moving and personal speech of Aliki Arouh, who works for the Jewish Community of Thessaloniki Archives in Greece, at the presentation of the EHRI portal in March 2015 in Berlin. She had been invited to the 2012 EHRI workshop, and 'has ever since enjoyed a mutually beneficial collaboration with the project' (EHRI, 2015).

Creating a 'collaboratory' and reaching this type of involvement does not happen by itself: the importance of the workshop cannot be underestimated. During this December-day in 2012, EHRI was able to bring the 'virtual' to the 'in-person', creating a first group of active users who have 'met' EHRI and feel they have a connection to the project (for those having been interviewed by the user requirement Work Package, this was a second in-person meeting with EHRI). Just as in-person contacts help significantly to become a successful virtual team (one of the paradoxes of virtual teams, see Dubé and Robey, 2008 and Segers *et al.*, 2012); likewise, having a first in-person contact with a core group of potential VRE users can help a VRE like EHRI reach the 'critical mass of users' that it needs for its sustainability. As has been pointed out in the literature, the sustainability of VREs is one of their biggest challenges, and the best guarantee for sustainability is community support:

> Ultimately the sustainability of VREs is dependent on their acceptance and use by the communities that they are intended for. The more VREs prove themselves to be indispensable for research activities, the more likely they will be seen as priorities for continued funding and support at the institutional level and ideally, also beyond. This reinforces the need to develop VREs in close collaboration with research communities, since they need in some ways to be self-built by research communities in order to ensure that different kinds of research can actually be supported by them . . . A key issue for any virtual community, not just those engaging in research, is to encourage enough contribution to

> make it seem interesting enough for others to join and/or engage with that community. . . . Getting the critical mass of users together is also important for the sustainability of a VRE, as further funding will be very difficult to secure without visible uptake in the research community.
>
> (Carusi and Reimer, 2010, pp. 33 and 36)

> VRE development is a resource-intensive undertaking that is almost certainly doomed to failure if the needs and desires of the intended user community are not taken into account. . . . 'bottom-up' approach, ensuring that researchers were asked what it is they do, how they do it and what might be useful to them.
>
> (DARIAH-DE, 2013, p. 10)

> 'If you build it, they will come' has been perhaps too common a mantra, underpinned by the belief that it will be enough to publish rare or unique content on the Web in order for users to flock in and for a Website to thrive . . . 'the value of a . . . project is quantified by the benefits it creates for users', which has an impact on how much a resource will be used and, in turn, its chances of survival and sustainability.
>
> (Marchionni, 2009)

VREs depend on their users: without users, no VRE. Therefore, it is logical that user requirements are given high priority in the development of VREs. This contribution wanted to shed light on how the content-providing work – bringing the data into the VRE – also should be very much in dialogue with the envisaged users. It gives a tool to engage with the research community, to learn about their work, their challenges (which in the Greek case study almost all fall under the label 'access to the sources'), and their needs, not only from a technical point of view (tools and functionalities of the VRE), but also – and very importantly to attract the researchers' attention – on the content. 'Putting a face' to the VRE creates engagement and trust and the invitees to the Greek workshop may, for example, be more willing to annotate and contribute, become active users of the VRE, than if they did not meet anyone from the EHRI project in person.

Moreover, we triggered the attention of the researcher in their 'core business', in their research activities, 'without emphasising the technology as such' (Marchionni, 2009). It has often been stressed that 'the learning curve must not be too steep' (DARIAH-De, 2013, p. 20), as the technology side of the VRE most often is not the main attraction of the researcher to the VRE. The benefits lay in how the VRE accommodates the researcher's needs in his core business: 'scholars use the features, they don't need to understand the technology behind them!' (DARIAH-De, 2013, p. 20). The Greek workshop has had multiple benefits for EHRI as it ensured that the content fits the user, it is a time-efficient identification method, avoiding duplication of work and at the same time investing in a smooth communication with the future VRE users.

The presentation of the EHRI infrastructure took place on Thursday, 26 March 2015. Until then, and also during the second phase of the project, EHRI continues to be in touch with the users from its user requirement interviews, and the project team is implementing the top priorities as outlined by the user requirements reports. User feedback continues to be gathered via the data identification team (including subcontractor Holocaust source-experts in different countries), which is using the database and all features in development, and user feedback is equally explicitly included in EHRI's second funding period.

By engaging in this way with the VRE's core audience and future users of the infrastructure, EHRI hopes to address the needs of researchers as concerns both content and tools. The success of these endeavours seems to bear fruit: by the time of the kick-off meeting for the second funding period of the project, the portal counted over 500 registered users and over 200 sessions a day (and this does not indicate how many researchers visit the infrastructure without registering). For the Greek case, EHRI is now able to assist the research community with an overview of 102 collection descriptions, scattered over 34 collection-holding institutions (plus information on five more institutions for which the collections still have to be added). Merging technical with content user requirements was a fruitful experiment and provided the project with invaluable information that continues to feed into its development.

Notes

1 This includes the following countries: Albania, Algeria, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Kosovo, Latvia, Libya, Lithuania, Luxembourg, the former Yugoslav Republic of Macedonia, Moldova, Monaco, Montenegro, Morocco, the Netherlands, Norway, Poland, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Tunisia, Ukraine, United Kingdom and the Vatican City State.

2 As this was a closed workshop, the identities of the participants will not be revealed in this article for privacy and confidentiality reasons. Quotes from the workshop are clearly indicated by quotation marks and identities of the respondents can be requested by contacting the author.

References

Anderson, S., Speck, R., Links, P., Benardou, A., Constantopoulos, P. and Dallas, C. (2013). *Deliverable D.16.6 Functional specifications (EHRI internal documents)*, November.

Angelis, S., Benardou, A., Constantopoulos, P., Dallas, C., Fotopoulou, A., Gavrilis, D., Manola, N., Anderson, S., Speck, R. and Links, P. (2013). *Researcher Practices and User Requirements, EHRI Deliverable DL 16.4 (EHRI internal documents)*, November.

Archeiomnimon (2013). http://arxeiomnimon.gak.gr/, retrieved 13 June 2013.

Bennett, G., Pohl, D. and Vanden Daelen, V. (eds) (2015). *EHRI Country Reports*, www.ehri-project.eu/country-reports, retrieved 21 September 2015.

Blanke, T., Candela, L., Hedges, M., Priddy, M. and Simeont, F. (2010). Deploying general-purpose virtual research environments for humanities research, *Philosophical Transactions of the Royal Society,* 368, pp. 3813–3828.

Carusi, A. and Reimer, T. (2010). Virtual research environment collaborative landscape study. A JISC funded project, http://jweblv01.jisc.ulcc.ac.uk/media/documents/publications/vrelandscapereport.pdf, retrieved 11 June 2013.

Central Board of Jewish Communities of Greece (2009). *The Holocaust of the Jews of Greece*, www.kis.gr/en/index.php?option=com_content&view=article&id=401:holocaust&catid=99:2009-06-04-07-06-01&Itemid=76, retrieved 25 March 2014.

Conference on Jewish Material Claims Against Germany (2012). Directory of Holocaust-Related Archives, www.claimscon.org/archivist_forum/archive_search.asp, retrieved 20 December 2012.

DARIAH-DE (2013), *VRE Blueprint* (M 2.5.1), DARIAH-DE (GCDH, WP2.5, version 21 February 2013).

Database of Greek-Jewish Holocaust Survivors' Testimonies (2013). http://gjst.ha.uth.gr/en/theproject.php, retrieved 18 June 2013.

Dubé, L. and Robey, D. (2008). Surviving the paradoxes of virtual teamwork. *Information Systems Journal*, 19, pp. 3–30.

EHRI (2011a). *Internal documents, Description of Work (DoW)*, 2011-01-12.

EHRI (2011b). *Internal documents, Deliverable 15.2*, August.

EHRI (2012). *Internal documents, Deliverable 15.5, Selection Criteria and Guidelines for Collections*, August.

EHRI (2013). *Project Factsheet European Commission*, www.ehri-project.eu/drupal/webfm_send/123, retrieved 2 May 2013.

EHRI (2015). *Presentation of the European Holocaust Research Infrastructure*, www.ehri-project.eu/presentation-ehri, retrieved 14 September 2015.

E.L.I.A. (2013), www.elia.org.gr/pages.fds?pagecode=01&langid=2, retrieved 13 June 2013.

Jewish Museum in Greece (2013). www.jewishmuseumgr/, retrieved 13 June 2013.

Judaica Europeana (2013). www.judaica-europeana.eu, retrieved 13 June 2013.

Marchionni, P. (2009). Why Are Users so Useful?: User Engagement and the Experience of the JISC Digitisation Programme. *Ariadne*, 61, www.ariadne.ac.uk/issue61/marchionni, retrieved 18 June 2013.

Mémorial de la Shoah (2013). *Guide des Archives de la Shoah* (www.memorialdelashoah.org/b_content/getContentFromNumLinkAction.do?itemId=182&type=0, retrieved 24 June 2013. The website is an updated version of the following book: *Mémorial de la Shoah* (2000). *Guide européen des sources d'archives sur la Shoah (European guide of archival sources on the Shoah).* Paris: Centre de Documentation Juive Contemporaine.

Segers, J., Meulepas, A. and Marichal K. (2012). *Visie: Actuele tendensen in leiderschap: vijf paradoxen voor e-leiders. HR square: gids voor arbeidsrelaties en personeelsbeleid, 120*, pp. 54–56.

USHMM (2013). www.ushmm.org/research/collections/archives/, retrieved 12 June 2013.

8 Mubil

A library-based immersive virtual environment for situated historical learning

Alexandra Angeletaki and Marcello Carrozzino

Libraries in the 21st century face a great challenge: to improve the integration of digital technologies with their archive material and to find new ways of disseminating their collections to the public. Research and evaluation on the use of digital tools and new software technology for educational purposes in current library practices is still under development. There is a constant demand among current practitioners for creating a working model to explain how these technologies can be adapted. Heritage institutions in general need a conceptual frame in order to discuss the diversity of the strategies applied to the digitalization processes, as well as analyse and disseminate the experience this creates for users and for society in general.

In this chapter, we examine the idea of a paradigm shift on the concept of what a book is, by creating a new 3D space for reading and studying. Knowledge from the past can be delivered in many ways and 3D technology has been an inspiring tool to work with. Our aim is to investigate new ways in which archive knowledge can be acquired by the general public and how to leverage a dynamic and exciting exchange between users and information within a brand new platform.

Introduction: educating the researcher of the future

The special collections of the Norwegian University of Science and Technology (NTNU) University Library were established in 1768, and our library was the first scientific library of Norway. During this era of scientific enlightenment, scientists started to use positivism and experimentation, and thus created powerful new ways of thinking and developing knowledge. Technology today might provide a similar paradigm shift, as long as it contributes to create dynamic engagement.

The ongoing goal of NTNU University Library is to continuously promote its services, improve access to its available scientific resources and allow the public to use the library facilities as needed. Traditionally, most users would come to the library desk to ask a librarian for help regarding a book. The development of the World Wide Web has changed this altogether and it challenges our preconceptions and expectations about our

future roles as librarians (Tonta, 2008). Increasing the amount and speed of digitization is a demanding task, forcing changes in attitudes and teaching practices (Dilevko and Gottlieb, 2004; Bailey, 2010).

Digital archives are all too often treated like electronic shelves of dusty books. We consider the increased digitization of books and archives as a unique opportunity to challenge and change the current relationship between user and institution. We argue here that today's libraries and archives have the chance to explore the role of emerging technology in disseminating practices, in a way that reaches a wider public – especially at a young age. Workshops, gaming and 3D environments: these are all tools in the new and diverse archivist kit for the 21st century.

3D technology is used currently by various projects (CYARK500, 3D-Icons) for documentation, preservation and dissemination of heritage monuments and museum or archive collections (Cosmas *et al.*, 2001; Grosman *et al.*, 2008; Forte *et al.*, 2012). Its use has influenced the user experience and created new possibilities for fragile objects like manuscripts or cultural remains handled as digital representations rather than as physical objects (Almeida *et al.*, 2006). Research in the field has borrowed ethnographic tools and usability analysis methods from social sciences (Nelson *et al.*, 2005) to evaluate the learning outcome of the visitor experience in such immersive environments.

A series of projects suggest that most computer applications afford digital involvement through active participation and they typically enhance learning through a cognitive dialogue with their users (Lindgren *et al.*, 2016). According to knowledge creation theory, an individual seems to learn more effectively when tacit knowledge meets explicit knowledge in a social context (Zaqout and Abbas, 2012). An individual contributes his or her knowledge to a group task and learning occurs through participation.

The goal of an immersive simulation is to create an illusion of the real world. When an individual enters into a computer-mediated 3D environment, the experience of presence becomes immediate through the feeling of being there (Pavlidis *et al.*, 2007). There are few empirical studies on collaborative learning in computer-mediated environments, where empirical data are collected (Arbaugh, 2000; Bailey and Kurland, 2002). Some researchers argue that learning itself should be redefined and that educational goals should be questioned regarding their aim and methods (Biesta, 2005; Bayne, 2014).

Our approach builds on all the above ideas, and the idea that the metaphor of a real world to a virtual game creates increased engagement, motivation and integration of knowledge that is modified by the users themselves (Bailey and Kurland, 2002). Our hypothesis is that the use of 3D technology coupled with the potential of user interaction with virtual objects leads to a better conceptual understanding of the context.

There is no doubt that digital literacy will be a skill useful to acquire for all of us as Internet resources and their content grows (Davis, 1989). The number of digital assets produced is enormous and the possibilities for

developing strategies in evaluating the knowledge available is compelling. How will the scientist of the future research a subject if they have all the materials and knowledge in a mobile device that they carry in their pocket? What if virtually performed tasks could replace activities in physical labs of chemistry, botany, physics, etc.? Immersive technology could offer great training opportunities during one's study period and improve contextualization of learning compared to 2D alternatives.

Mubil: an interactive environment of archive exploration

Mubil-Lab was established in 2012 by NTNU University Library in Trondheim in order to develop open access digital educational workshops and applications using innovative technology. Our research concentrates on educational tools enriched with game applications for students and teachers with unique source-material drawn from our archives.

The idea of creating a 3D lab (Figure 8.1) for the Norwegian University Library of Trondheim came as a response to the challenge of modernization that libraries face today. We wanted to investigate different possibilities of delivering knowledge from the past in order to meet the user's level of knowledge and skills and help their quest of knowledge.

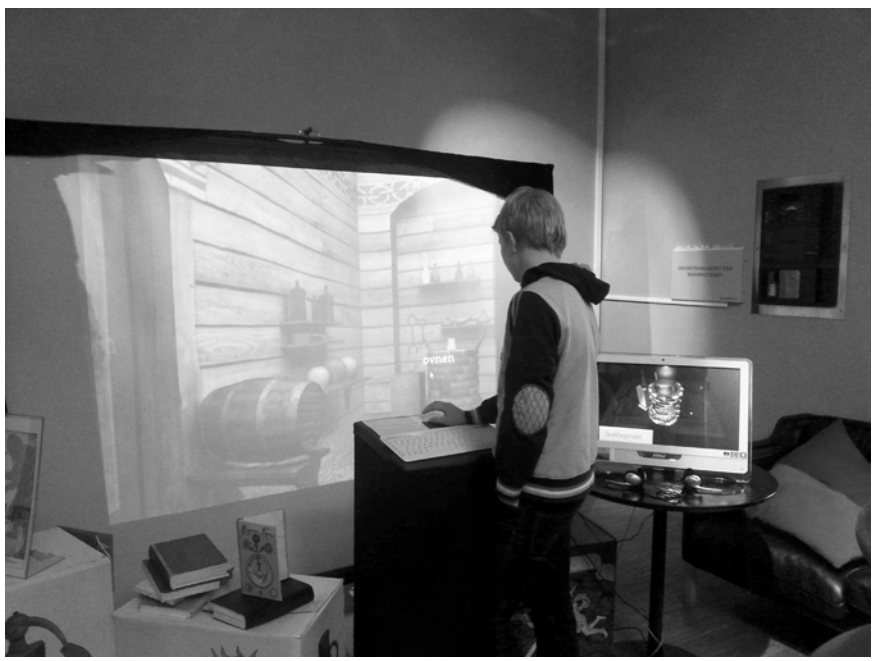

Figure 8.1 An exhibition with the augmented book and game on medicinal distillation at the NTNU Museum in Trondheim, 2014.

Mubil started as an experiment where old books would be enriched with stories, narratives and knowledge from the past, with the use of 3D models and multimedia content, to trigger curiosity and create an immersive learning environment (Carozzino *et al.*, 2013). We were trying to figure out whether technologies like 3D imaging and virtual environments could actually boost our intentions to seek knowledge from the past but in a format packaged for the needs of the 21st century.

We decided to put together a cross-disciplinary team, bringing people with different backgrounds and expertise together to create the book of the 21st century. It would not just be a flat digitized version of a physical book. We tried to think outside the box, with a vision that books or manuscripts too fragile or too valuable to be handled by many hands could instead become accessible to the general public in a new format more suitable for our technological era. We then created two 'immersive' books. We believe that this became a paradigm shift in the experience of 'reading a book', combining the simulated experience of handling an ancient manuscript with the excitement of gaming.

Our collaborators were:

1 NTNU Department of Computer and Information Science Trondheim, Norway.
2 PERCRO LAB Scuola Superiore, Santa Anna, Pisa, Italy.
3 Namtaru Creations Design and Art Direction studio, Trondheim, Norway.

Our collaborators brought expertise from different disciplines such as IT, Museum studies, pedagogics, software engineering and educational design. The research team worked in a cross-disciplinary way to explore 3D technology as a tool for educational workshops, with the aim to entertain but also support learning. Two teams were established: one worked with the content of the archive objects to be presented at the Gunnerus library and one developed the technical applications at the Percro lab in Pisa. Both teams have had regular contact throughout the whole period.

In order to investigate our hypothesis and collect data on the user experience, we created an educational workshop concept around the Mubil lab for schoolchildren between 12 and 19 years old at the NTNU University Library. This concept was developed in collaboration with teachers, beginning in 2013 on different subjects like chemistry, botany and archaeology, where the applications developed in the project were a central part of the workshop experience offered.

Two books were presented to our users. One was a treatise of medicinal distillation from 1569 by Adam Lonicer, the second one was a handwritten travel-journal started in 1670 by a young noble, Hans Lilienskiold. Both were presented in an immersive environment, where the users wore 3D glasses, but also as flat scans on touch-screens for them to browse. The first book was also complemented by a game that re-created an experiment in an alchemist's

laboratory, reconstructed from drawings found in the original book. The physical book thus became an 'augmented digital book' with new qualities and character, enriched with information around the period, the writer and the issues on which it reflected. The concept of augmented book (Billinghurst *et al.*, 2001) is in this case exploited in a completely virtual dimension with the added value of a wider range of multimedia elements contributing not only to the user engagement, but also to a better understanding of the text.

In that process we engaged researchers working on the history of Chemistry and one of our collaborators wrote his Master's on the subject (Kirkemo and Kvittingen, 2014). As they note:

> Libraries with collections of old and rare books seem to have two inherently irreconcilable purposes: to curate the historical artefacts that the books are and at the same time make them available to the public. The traditional solution to that is to let the public carefully read through them . . . Mubil project was instituted to develop a way of using digital methods and modern technology to amend the irreconcilability of these purposes.
>
> (Kirkemo and Kvittingen, 2014: 2)

As an example, the book of Lonicer emphasizes how in 1569 it was valuable for a doctor or chemistry student to understand, for example, the pharmaceutical properties of a plant. There is a plethora of ways to procure such information. To make it available in 1569, one had to write a book. Today, we have many ways to disseminate knowledge around the subject. It is still valuable, however, to understand what we thought we knew about the pharmaceutical properties of a plant that existed five hundred years ago.

Today's students seek information, not only through physical books, but many other digital tools and ways. It is important to notice that connecting with digital embedded information can be presented within a historical framework, so that today's students also connect with the past. Thus, they become curious about it and, most important of all, understand that 'facts' can be inconstant, and that knowledge is fluid and forever developing. This is something that an immersive library or archive can uniquely bring to a wide range of educational tools. The laboratory concept has been designed with the overall aim to reinforce public interest and to create dynamic links between old collections, through innovative communication platforms and via exhibition layouts. We have also collected user behaviour data in order to analyse it, and to see changing patterns and the demands that are placed on the design of such tools.

Learning by doing: an experiential learning environment

Our study considers learning a process of sociocultural character, defined as a system of participatory competencies and activities (Leinhardt and

Knutson, 2004). This is an area of growing interest for museum studies and is exploited also by library professionals to market their collections and attract new users. Understanding how learning occurs within a virtual interaction and trying to measure the effect that virtual application and ICT technologies have on learning is a demanding task for researchers (Roussou, 2004; Arnab *et al.*, 2015). The quest for identifying 'the weaknesses and strengths of a virtual heritage environment's ability to provide a cultural learning experience' needs to be further investigated (Champion, 2011: 177).

Several studies seem to introduce various usability tests to evaluate the learning outcome, based on visitor/user experience. The result is often not of universal value as it is closely connected to the particular context (Champion, 2011; Carbonell, 2012). A few studies have looked into the role of social conduct in mixed reality environments and the role of collaborative production of meaning in a space of technological innovation in museums (Galani and Chalmers, 2002; vom Lehn *et al.*, 2007). Others have tried to establish evaluation methodologies of the design architecture, looking at collaborative skills in educational virtual environments as second life, where 3D technology is applied as well (Bouras and Tsiatsos, 2006; Dawley, 2014).

Mubil-lab has considered all of the above approaches in the research design phase and organized several workshops for school students and university students in the field of learning. The study has used a constructivist approach that suggests that every visitor in a given context assigns his or her own understanding to it, through the process of constructing meaning (Greenhill, 1992). The educational workshop concept was established as a visit to our library, where the participants were informed of the scope and context of the project, and then participated in a demonstration of the applications developed by the project.

The new element of the visitors' experience in our study is the practical engagement of free choice in tasks of interactive character, in an immersive environment. The conceptual model used in the study is borrowed from the so-called 'Enactive Knowledge' and refers to 'knowledge gained through a perception-action interaction' (Angeletaki *et al.*, 2013) with the object itself. This kind of knowledge seems to be direct and personal, based on the experience one gains through multi-modal tasks connected to exploit the particular object.

In contrast to traditional library visits, Mubil lab offers new possibilities of cultural experience that adapt and follow the user's interest. When Mubil's first applications were launched, we tried to reach out to schools and other user groups by email and creating websites (Mubil.no), blogs (http://gunarchives.blogspot.no/), Facebook pages and participated in exhibitions and shows. None of the above actually brought users to the library. The Gunnerus library has been visited by researchers and other adult users, mostly interested in local and national history.

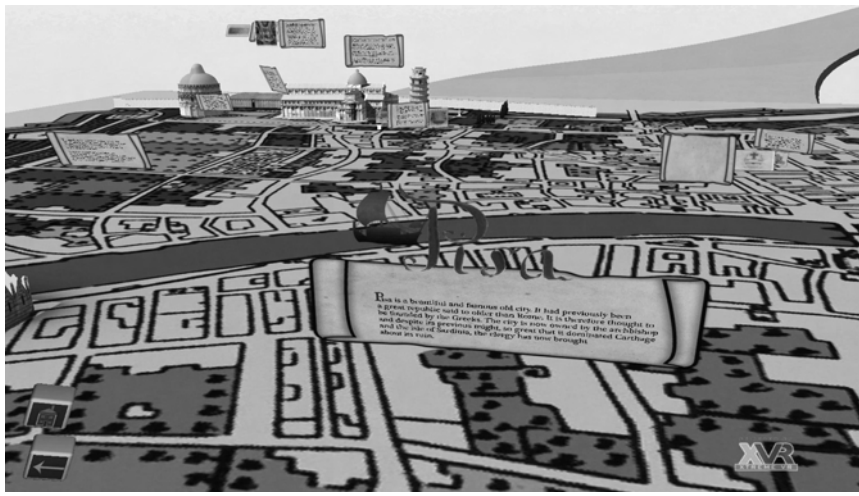

Figure 8.2 A manuscript transposed in an Information Landscape travel with texts and drawings from Hans Hansen Lilienskiold and his trip to Pisa, 1670.

Our special collections are known mostly to researchers, historians, journalists and local institutions, such as museums and other libraries. The library has never had school classes visiting, except on special occasions. The Mubil lab workshops were thus brand new, as a public space, into which school groups were invited to experiment and learn. The visits were planned in collaboration with their subject teachers as part of their school curriculum.

The books presented are classified as 'rare' and are not accessible to a wide public because of their fragility and value. The tool developed and proposed by project Mubil transcribes digital representations of objects or texts in a 3D environment and allows the interaction of the user within it.

Technological challenges and choices

The 3D tool developed by our collaborator Percro-Laboratory in Pisa supports different modules of interaction with the user. The books were presented on a touch screen, as digital books transcribed to a 3D format, enriched with text, video and audio on the subject. They were developed using the XVR technology (Tecchia, 2010) and Information Landscape (IL) technology integrating 3D models created from the book's images that the user can activate and with which she can interact. All resources are loaded from a custom XML file that allows the book's designer to quickly author, modify, or even create a completely different 3D book on the same container XVR application.

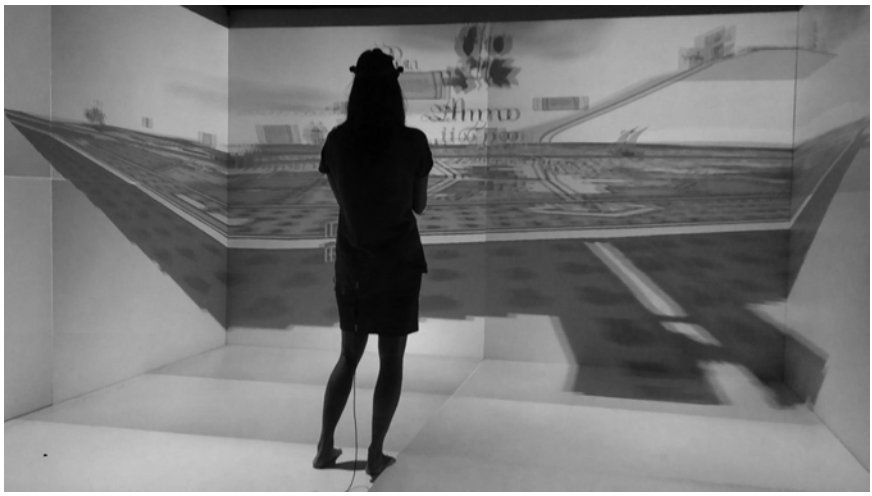

Figure 8.3 The Lilienskiold's Information Landscape experienced immersively in the X-CAVE at PERCRO Laboratory, Pisa.

It is also supplemented by a game in 3D format, where an experiment from the book was available to the participants, either in a 3D screen with 3D glasses on, or on a computer with a regular flat screen. The 3D game was also developed using the XVR technology, using the same container/content paradigm, with the XML content file describing resources and the game mechanics (Lorenzini *et al.*, 2013).

From a technological point of view, as new stereoscopic devices like low-cost Head Mounted Display (i.e. Oculus Rift HMD) became increasingly popular, we added support of these appliances to the Mubil applications in 2014. We used different interaction paradigms to navigate the virtual environment in order to cope with the absence of a standard interaction interface for such appliances. (Oculus Touch controllers, which will be available in the first semester of 2016, are expected to overcome this, however.) This could lead to an interactive development approach, which gradually enhanced the immersive character of the game experience.

The choice of the particular XVR platform for the development of the applications was influenced by two main factors: its ability to rapidly change to address any need (as it is developed internally by one of the Mubil partners) and its flexibility in terms of rendering devices. The latter feature has allowed the rapid deployment of Mubil applications for a range of visualization alternatives, such as 3D stereo wall, CAVE-like systems (Figure 8.3), HMD and Touch-Screen devices. Concurrently with the development of the Mubil project activities, new 3D platforms have been spreading in the market, such as Unity and Unreal 4, at affordable rates or even free. Although XVR adaptability remains superior, these platforms, by supporting the

visual authoring of scenes and storyboards, commonly enable a quicker prototyping of applications and a better visual quality. As these tools are developed mainly for computer gaming, they would probably be more suitable for applications such as the Virtual Lab (but not entirely appropriate for more customized applications, such as the ones specifically developed for the 3D visualization of the books).

How we did it: methodology, data analysis and results

In order to collect data on how people interact with 3D technology tools, we invited school-classes and university students to visit our 3D lab together with their teacher, and they were asked to interact with both the books and the game. They worked in groups of five and moved from one application to the next, as they wished. The overall architecture of the workshops allowed our users to seek knowledge by free choice in an informal learning environment. Social interaction and collaboration was also encouraged during the experiments. The main research focus was to investigate and analyse the learning strategies of the participants involved when interacting with 3D technology.

In order to be able to evaluate the user experience data we had to develop a mixed evaluation methodology following a learner-centred approach, with the participation of both pupils and teachers. We referred to the users as our experts and throughout the project period we have used their feedback to decide on changes and improvements as the project was developed. We tried to collect the opinion of the participants in discussion rounds, and to get feedback, not only on the applications presented, but also on the idea and goals of 'Mubil project' to use immersive technology and support learning.

In order to capture the experience of the participants, we implemented two different data-collection procedures: first, we used a quantitative approach to identify how students perceive the activity. To develop this, we used a survey, based on questions identified in the literature as important for IT activities. Second, we applied a qualitative approach through group discussions. Questions were designed to probe different aspects of a) IT usage and b) creativity. We carried out quantitative and qualitative analyses on the data obtained from the survey. We used a questionnaire for the quantitative analysis of the user experience where we tried to measure the computer literacy the users had, which allowed them to have or not to have control over the activity. In addition, we measured enjoyment during the activity, which seemed to be positively related to perceived ease of use and to user satisfaction. Satisfaction in terms of 'I enjoyed the activity' was also high among the participants in their survey answers.

During the design of the qualitative analysis, we arranged the results from the interviews into categories in order to analyse them more effectively. When coding the interview transcripts for concepts, we reduced the

text to manageable categories in order to focus on content that was relevant to the research area (such as 'technology', 'performance', 'intuition' and 'reflection on learning'). The participants did not give high scores to the first three categories, but felt that they had actually learned something. That was also confirmed by the knowledge performance test where they scored highly.

In their interviews, the students said that a visit to such a digital archive environment was a rich learning experience and they wished to learn more about the subject. It was apparent in both the survey answers and the interviews that users appreciated experiential learning and task collaboration. An overall positive appreciation of the 3D environment comprised 70 per cent of the participants, but 90 per cent of the participants considered the sequence of the experiment tasks difficult to grasp and the duration of the game too long. The fact that this kind of 3D technology was employed to augment the book experience increased their motivation for learning, and this was explicitly commented on by many of them.

Our survey showed that 90 per cent of the male users that participated in our experimental workshop felt that they could move around the lab freely and enjoyed the game, while only 40 per cent of the females participated willingly in the workshop, and they reported 90 per cent satisfaction. Female users remained more cautious and less decisive in their behaviour throughout the project. The interviews showed that the female participants play fewer games that demand collaboration in contrast to the male participants.

More precisely, the participants described the activity as interesting, but the sequence of game tasks seemed difficult to follow. They were not willing to spend time listening to instructions, since most of them were used to intuitive gaming structures. They all liked the idea of the augmented book with the 3D drawings, but in order to be able to use it for school activities they would have chosen more multimedia content that would inform potential users about the book. We observed the participants while they were performing and conducting the visit, and noticed that all the groups were using social collaboration in their performance.

We also saw that young students used their innate aptitude (as members of a digital generation) and gradually felt comfortable in a 3D environment. Their active involvement and social conduct in solving the given tasks also grew as more time was spent in the immersive environment. The most interesting field observation by the Mubil team was that while the groups were solving tasks in the immersive environment, social presence and emotional connection were very important. Expressions like 'do that', 'move this' or 'oh no, not that', were often used, showing that even though only one person was holding the mouse, all the others were engaging actively with the game. Collaboration made them more aware of what the other players were doing. Competition between groups was also a positive factor in fostering higher engagement in the user tasks.

Both the school students and the university students of the teacher's educational programme were very positive about the digitally simulated environment of a 3D book and the experiment game. They all liked the idea of using 3D to revive and reconstruct a book experience and an experiment as such, but they could see the limitations of such 3D software especially when it came to the graphic quality. They also thought that the software was a bit slow to use. 'Schoolchildren, accustomed to playing games with better graphics and faster software, may find it boring,' they said.

Young users, the target group of this study, seem to be technology confident and genuinely interested in interactive tools. As observed in our workshops, they seemed to engage actively in performing knowledge quest tasks and support their learning experience towards their own interest (Costabile *et al.*, 2008). Nonetheless, the so-called digital native generation is not altogether as confident and competent in their digital skills as it is usually assumed, and their digital behaviour is dependent on several factors: age, gender, socioeconomic status, mastery levels and prior use (Angeletaki *et al.*, 2014), and social interaction and participation.

Where to from here? Discussion and conclusions

Mubil 3D lab constitutes a paradigm shift in the experience of 'reading a book'. The difference between Mubil books and digitized collections of rare and old books is that Mubil seeks to enrich the experience of the book with 3D models and multimedia content, thus creating a hybrid rather than a digital facsimile. Mubil provides evidence that virtual environments can support experiential learning because it involves active participation and the engagement of innate knowledge.

This study casts such technology in a different light, promoting heritage-related virtual environments in informal learning situations. Our hypothesis is that when interaction remains a passive character, it does not support learning because it does not allow active involvement. Mubil is based on active involvement and encouraging social interaction via problem solving. It has created an active new public space and a digital exhibition environment for the presentation of archive and content to young users at a library that had been used mostly by grown-ups and researchers up to now. Thus, it has constituted a novel experiment in the social context of the learning community with the help of new technology.

The novelty in our experiment is the use of 3D technology and the fact that we allowed the groups to work through a framework of social involvement with a kind of mixed reality installation and we observed them during this process. The observations and the interviews have provided us with new narratives that are visitor-centred around our archives. We used the interaction between the participants and the objects as an observation field so that we could examine what was interesting for the users to actually do and what was not. This way we could adjust the applications to their behaviours and interests.

As established by many in this field, (Roussou, 2004; Bronack *et al.*, 2006; Abrahamson 2014) participatory activities create engagement with the objects presented in the virtual world in a frame of collaborative learning processes. Users become increasingly involved in the digital landscape of their immediate interest field. In our case, the interest field was defined by their teacher and they were expected to answer several knowledge tests that would form part of their overall grade. This was a motivating factor that was actually outside the user's personal interests. Even though they had free choice to participate or not in the workshop, they did not really have a choice when it came to the subject. It was more a constructed interest, acting as an educational framework encompassing the workshop by the Mubil team and the class teacher. This factor may have acted to distort the results and scores of satisfaction tests.

The relationship between instructions and learning, though, through digital gaming is very complex and we need to investigate further on the effect of instructions on students in a digital educational game, since this is a fast-evolving research area (Hawlitschek and Joeckel, 2017). One of the main drawbacks, however, is that the ongoing development of such tools contains technological and performativity problems. 3D technology is developing apace and the film and gaming industries are investing a lot in its exploration. As a public institution with limited funding, we cannot compete with the 3D gaming industry that has recently launched the new Oculus head mounted display, Google Glass, or Samsung Gear. Some technical issues such as bugs cannot be avoided and they seemed to affect the user-experience in the workshop environment. We have actually utilized user feedback to develop our applications and find new solutions to the performance errors registered during the lab experiments.

Another challenge was our involvement in the actual design of the workshops, which did not allow us to have a neutral view of the participants. There is a pressing need for further study of the different external factors involved in learning due to the framing of our study. It might be influencing the performance of the users through the influence of the teacher on the students, and his or her role in the planning and executing of the workshop.

We strongly believe that Mubil is exploring a new way for historical archives to create a more engaging learning space by deploying novel technological tools. The project has had a flexible and dynamic design in its overall architecture, which allowed us to work in an interdisciplinary research frame and learn about each other's disciplines. It has forced us to understand terminology and methodological discussions in all the disciplines involved, and construct a shared language for the project.

Technological solutions to learning tend to involve the regurgitation of current knowledge on a subject – even a 'dumbing down' in some cases. The trend is to produce an almanac of collected but often superficial knowledge on a subject, akin to a Wikipedia approach. But we already have Wikipedia

and, technology being what it is, there is little doubt that it – or something like it – will develop into an immersive environment as technology begins to facilitate such a direction more widely and cheaply.

The difference with the Mubil project is that it has sought to deliver 'old' knowledge, by digitizing texts and books that would be too fragile or precious to be made freely accessible, and make them available and attractive within an immersive environment. By doing so, we submit that an archive's true purpose is brought forward into the 21st century, because an archive's true purpose is not only to function as a repository of current understanding, but also to provide a series of historical snapshots of knowledge on a subject over time.

As Alberto Manguel wrote in his 2015 article in the *New York Times*:

> If we change the role of libraries and librarians without preserving the centrality of the book, we risk losing something irretrievable.
>
> By creating augmented books we preserve them as heritage artefacts that can enrich the knowledge process but also help them to blossom into digital tools that will change and augment the learning strategy of today's students.

References

Abrahamson, D. (2014). Building educational activities for understanding: An elaboration on the embodied-design framework and its epistemic grounds. *International Journal of Child Computer Interaction,* 2.1: 1–16.

Almeida, R., Cubaud, P., Dupire, J., Natkin, S. and Topol, A. (2006). Experiments towards 3D immersive interaction for digital libraries, in *Technologies for E-Learning and Digital Entertainment*, Springer, 1348–1357.

Angeletaki, A., Carrozzino, M. and Johansen, S. (2013). Implementation of 3d Tools and Immersive Experience Interaction for Supporting Learning in a Library-Archive Environment: Visions and Challenges, *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(2), 37–41.

Angeletaki, A., Carrozzino, M. and Johansen, S. O. (2014). MUBIL: Creating a 3D experience of 'reading books' in a virtual library laboratory, *International Journal of Heritage in the Digital Era*, 3(2), 271–286.

Arbaugh, J. (2000). How classroom environment and student engagement affect learning in Internet-based MBA courses, *Business Communication Quarterly*, 63(4), 9–26.

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., Suttie, N., Berta, R. and De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46: 391–411. doi:10.1111/bjet.12113.

Bailey, C. W. (2010). *Digital Scholarship 2009*, Charles W. Bailey Jr.

Bailey, D. E. and Kurland, N. B. (2002). A review of telework research: Findings, new directions, and lessons for the study of modern work, *Journal of Organizational Behavior*, 23(4), 383–400.

Bayne, S. (2014). What's wrong with 'technology enhanced learning', in *Proceedings of the Networked Learning 2014 Conference*, 7–9.

Biesta, G. (2005). Against learning: Reclaiming a language for education in an age of learning. Reprint for 25th Anniversary Issue of Nordisk Pedagogik. Selected as one of the three most significant articles in the 25-year history of the journal, *Nordic Studies in Education*, *Nordisk Pedagogik*, 25(1), 54–66.

Billinghurst, M., Kato, H. and Poupyrev, I. (2001). The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications*, 21(3), 6–8.

Bouras, C. and Tsiatsos, T. (2006). Educational virtual environments: Design rationale and architecture, *Multimedia Tools and Applications*, 29(2), 153–173.

Bronack, S., Riedl, R. and Tashner, J. (2006). Learning in the zone: A social constructivist framework for distance education in a 3-dimensional virtual world, *Interactive Learning Environments*, 14(3), 219–232.

Carbonell, B. M. (2012). *Museum Studies: An Anthology of Contexts*, John Wiley & Sons.

Carozzino, M., Angeletaki, A., Evangelista, C., Lorenzini, C. and Tecchia, F. (2013) Virtual technologies to enable novel methods of access to library archives, *SCIRES-IT: SCIentific RESearch and Information Technology*, 3(1), 25–34.

Champion, E. (2011). Augmenting the present with the past, in *Playing with the Past,* Springer, 157–176.

Cosmas, J., Itegaki, T., Green, D., Grabczewski, E., Weimer, F., Van Gool, L., Zalesny, A., Vanrintel, D., Leberl, F. and Grabner, M. (2001). 3D MURALE: A multimedia system for archaeology, in *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage*, ACM, 297–306.

Costabile, M. F., De Angeli, A., Lanzilotti, R., Ardito, C., Buono, P. and Pederson, T. (2008). Explore! Possibilities and challenges of mobile learning, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 145–154.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 319–340.

Dawley, L. and Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. *Handbook of Research on Educational Communications and Technology.* Springer, New York, 723–734.

Dilevko, J. and Gottlieb, L. (2004). *The Evolution of Library and Museum Partnerships: Historical Antecedents, Contemporary Manifestations, and Future Directions*, Libraries Unlimited.

Forte, M., Dell'Unto, N., Issavi, J., Onsurez, L. and Lercari, N. (2012). 3D archaeology at Çatalhöyük, *International Journal of Heritage in the Digital Era*, 1(3), 351–378.

Galani, A. and Chalmers, M. (2002). Can you see me? Exploring co-visiting between physical and virtual visitors, Department of Computing Science, University of Glasgow.

Greenhill, E. H. (1992). *Museums and the Shaping of Knowledge*, Routledge.

Grosman, L., Smikt, O. and Smilansky, U. (2008). On the application of 3-D scanning technology for the documentation and typology of lithic artifacts, *Journal of Archaeological Science*, 35(12), 3101–3110.

Kirkemo, F. M. and Kvittingen, L. (2014). Adam Lonicer's Kreuterbuch and 16th century distillation: An experimental approach to the study of Adam Lonicer and some of the technology applied by him and his contemporaries in the production of medicines, unpublished Master's thesis, NTNU.

Hawlitschek, A. and S. Joeckel (2017). Increasing the effectiveness of digital educational games: The effect of a learning instruction on students' learning, motivation and cognitive load. *Computers in Human Behavior.*

Leinhardt, G. and Knutson, K. (2004). *Listening in on Museum Conversations*, Walnut Creek, CA, Rowman Altamira.

Lindgren, R. *et al.* (2016). Enhancing learning and engagement through embodied interaction with a mixed reality simulation. *Computers and Education*, 95: 174–187.

Lorenzini, C., Carozzino, M., Evangelista, C., Tecchia, F., Bergamasco, M. and Angeletaki, A. (2013). Serious games for disseminating the knowledge of ancient manuscripts: A case study, *SCIRES-IT*, 3(2), 135–142.

Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C. and Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment, *Educational Technology*, 45(1), 21–27.

Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V. and Chamzas, C. (2007). Methods for 3D digitization of cultural heritage, *Journal of cultural Heritage*, 8(1), 93–98.

Roussou, M. (2004). Learning by doing and learning through play: An exploration of interactivity in virtual environments for childre, *Computers in Entertainment (CIE)*, 2(1), 10–10.

Tecchia, F. (2010). A flexible framework for wide-spectrum VR development, *Presence*, 19(4), 302–312.

Tonta, Y. (2008). Libraries and museums in the flat world: Are they becoming virtual destinations?, *Library Collections, Acquisitions, and Technical Services*, 32(1), 1–9.

vom Lehn, D., Hindmarsh, J., Luff, P. and Heath, C. (2007) Engaging constable: Revealing art with new technology, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1485–1494.

Zaqout, F. and Abbas, M. (2012). Towards a model for understanding the influence of the factors that stimulate university students' engagement and performance in knowledge sharing, *Library Review*, 61(5), 345–361.

9 Digital heritage tools in Ireland
A review

Sharon Webb and Aileen O'Carroll

Introduction

The Digital Repository of Ireland is an interactive national Trusted Digital Repository for contemporary and historical, social and cultural data held by Irish institutions, providing a central internet access point and interactive multimedia tools, for use by the public, students and scholars. DRI is an exchequer-funded project, currently in its fifth year of operation, comprising six Irish academic partners. Stakeholders include the National Library of Ireland (NLI), the National Archives of Ireland (NAI) and the Irish national broadcaster RTÉ. This review of digital heritage tools is based on the results and findings of requirements and policy elicitation work undertaken as part of the initial development phase of the project. It draws on qualitative interviews carried out by DRI between October 2011 and August 2012, and additional supplementary research on digital tools and repository infrastructures carried out by the authors.

Our stakeholder interviews, which include interviews with the National Library of Ireland, the National Archives of Ireland, the National Gallery of Ireland, the Irish Museum of Modern Art, and a number of university libraries as well as academic research institutions and independent archives, highlight practices and trends in the provision and development of digital heritage tools. The interviews provide a unique insight into the challenges, as well as the opportunities, associated with online digital cultural heritage material and the ways by which organizations promote and support user interaction and engagement of this content. They demonstrate the commitment of cultural heritage institutions in providing users with tools that support functionality beyond access to content, but also highlight the challenge of maintaining digital tools in the long term. We highlight that key challenges for those providing user tools are associated with issues of preservation and sustainability of digital tools, and argue that for cultural heritage organizations the provision of digital tools is as important as providing access to digital content.

Stakeholder engagement and DRI's requirements

DRI's stakeholder interviews served two purposes: requirements elicitation and policy development. The interviews were semi-structured; we used a topic guide to ensure that essential information and points of interest were captured but left room for free and open discussion with participants (see O'Carroll and Webb, 2012). We asked our interviewees about their current practices for archiving their digital, as well as analogue, collections in order to capture the activities and policies that drive this important area. From these interviews we captured and specified a number of core functional requirements from the content provider or depositor's perspective. These include (but are not limited to):

- A REST-based API to support content reuse and repurposing.
- A single ingest web form and command line bulk ingest.
- Support of multiple metadata standards and data formats.
- The implementation of access rights.
- A full set of DRI's requirements are now available at: http://dx.doi.org/10.3318/DRI.2015.6[1]

DRI's number one, high-level, business requirement, the creation of a Trusted Digital Repository, was fulfilled in June 2015 when the Repository publicly launched and was subsequently awarded the Data Seal of Approval.

DRI's core business requirements state that the Repository must be robust, scalable and provide long-term digital preservation services to content contained therein. However, since DRI does not own any of the content in the Repository, the technical and social definition, and earning, of 'trust' was a concept that featured heavily in each of our stakeholder interviews. Our interviewees sought assurances in terms of DRI's sustainability model and its ability to provide long-term, sustained access to ingested content. While sustainability is an ongoing concern, the Repository's requirements and operational and legal policies have been developed specifically with this user cohort in mind – all our policies are publically available and have undergone significant peer and legal review where necessary, all our processes are transparent, and we have published our infrastructure choices as well as business sustainability and funding model proposals and approaches, among other best practice reports and guides. These concerns, policies and requirements are important from a particular perspective, that of the content provider or depositor. However, equally, if not more important to consider is the end user's perspective – the audience for whom we are building the infrastructure in order to provide sustained access to the digital content in the Repository.

Beyond preservation

Long-term digital preservation, concerned with providing sustained access to digital objects, is an activity that is rarely finished – when can we say that

something is preserved? As a core functional and business requirement of DRI, long-term digital preservation is, of course, priority. However, supporting user engagement and interaction with digital content, in ways that can help users overcome digital content saturation or fatigue, is equally important given the exponential growth of digital content over the last decade. As such, all content providers must look and think beyond just preserving digital object (the content of our national archives, libraries, museums, etc.) and consider how to support the presentation of, and interaction with, digital objects and the data (or knowledge) contained therein. While simply making content available and accessible is a task in itself (and one that requires resource allocation and commitments), the provision of additional functionality in the form of end-user tools is essential, given significant increases in user expectations as well as the sheer volume of digital content available.

We may define the success, or failure, of websites, online archives, digital libraries, digital repository, etc. by their ability, or inability, to capture the attention of their users or their audience. The 'if you build it [they] will come' mantra is long silenced as online content providers struggle to capture the attention of digital audiences in new and innovative ways. The data deluge and significant changes in online reading and consumption patterns (driven by and a result of hypertextuality) mean that digital audiences are spending less time reviewing, looking at and reading online content, and are instead opting to speed read and click; we are only ever one click, one tangential distraction, away from losing a user's attention (Carr, 2010). Our understanding of User eXperience (UX) and user engagement stems from significant studies in human–computer interaction (HCI), an area of computer science that foregrounds fundamental design concepts, principles and methodologies in software, or more specifically UI design and development. It informs us, at a basic level, that products – i.e. software – must be useful, usable and used (Dix *et al.*, 2004, p. 5). Throughout our project planning, DRI discussed reducing barriers to sharing data but, as Alison Prince discusses in 'Engaging digital audiences', we must also consider 'reducing barriers to engagement' (Prince, 2013, p. 253). It is in this sense that cultural heritage organizations must not only provide access to the digital cultural heritage content, which is stored, harvested and aggregated, but must also engage users with digital tools that enhance the user's experience and engagement with that content.

Digital cultural heritage tools

The provision and development of digital heritage tools serve to promote an institution's holdings, and to encourage user engagement and interaction with content contained in both digital as well as physical collections. Digital tools, as Prince observes, are used to 'reach, converse with, enthuse and promote specific actions among an audience', and can include 'websites, social

media, email and mobile technologies' (Prince, 2013, p. 350). The tools listed by Prince support a particular type of engagement, that of content dissemination and communication with digital audiences. These tools, websites, social media, etc., in many ways support the means by which users gain access to digital holdings and as such it is unsurprising to find that the majority of our stakeholder interviewees cited their finding aids as their most important digital tool to support user interaction with content – i.e. access.

However, while supporting access to content is a fundamental concern, we were also interested in the types of digital, or more specifically user, tools and/or methodologies, developed or employed by institutions to help users engage more thoroughly with the cultural heritage content they accessed. Significantly, 60 per cent of those interviewed provided additional tools to support user interaction with digital collections and/or objects; additional tools included 'user-generated annotations, crowd-sourced transcription correcting, networked mapping and graphing of relationships between material, text comparisons, data and geospatial visualisations, interactive maps, online exhibitions, interactive guides, interactive tables . . . educational tools', mobile apps, as well as finding aids that rendered results in various visualizations – e.g. temporal, geographical, etc. (O'Carroll and Webb, 2012).

The range of tools is reflective of the diverse data types held by the various institutions, and while image, text, audio and moving image files are the predominant data types being catered for, a number of our interviewees dealt with complex data types and sets such as geo-spatial data, CAD, 3-D images, BIM (Building Information Modelling) as well as various statistical data sets.

The rendering and presentation of image, text and audio files was often supplemented by the creation and curation of online exhibitions. The RTÉ Archives developed a number of online exhibitions that utilize the varied media content in their holdings. The exhibitions collate content on various subjects ranging from 'National Holidays, Festivals, Anniversaries' and 'Politics and Economics' to 'Arts and Culture' and 'Society and Social Issues' and provide additional contextual information on the exhibitions, in both English and Irish (RTÉ Archives).

RTÉ's position as the national broadcaster is reflected in the important historical and cultural exhibitions they have curated. Bringing their collections together in this manner contributes to and creates part of the national historical and cultural narrative while providing an opportunity for users to engage with the contents of the archive in a manner that brings interconnected images, audio and moving image together – e.g. 'RTE Coverage of General Elections' (from 1965 to the most recent in 2011). A number of cultural institutions created online exhibitions that mirrored the physical exhibitions on display, while a number used this feature as a provenance tool or a way to retrospectively display exhibitions. For example, the Irish Museum of Modern Art (IMMA) offers virtual tours of some of their past exhibits, while the National Library of Ireland (NLI) offers similar online tours, most notably 'The Life and Works of William Butler Yeats' and

'The 1916 Rising: Personalities and Perspectives'. Similarly, the National Gallery of Ireland allows users to create their own collections and curate according to their specific needs and interests. Yet, rather than just a space to curate or manage artworks of personal interest, the ability to share collections with other users, as well as enabling other users to edit their collection, means that this feature is not only a personal research tool, but also a collaborative tool, reflecting online research habits and patterns.

All interviewees were cognizant of the need to manipulate, and use, multiple digital channels to optimize engagement with digital audiences. The NLI, for example, use multiple social media sites not only to publicize events, but their collections and holdings as well. More interesting, however, is their use of Flickr Commons (see www.flickr.com/people/nlireland/). The NLI populate their Flickr Commons account on a regular basis with images that range from landscapes and buildings, to people and events, and ask the public to help identify the people, places, monuments, etc., in the photos. The simple use of this medium engages the audience in an activity that allows individuals to directly interact with the Library's holdings. The task of identifying key features and the opportunity to comment on images of historical and cultural importance not only adds value to content through the enrichment of descriptive metadata, creating information rich digital resources, but what is more important is that it enriches the user's experience.

Similarly, the Military Archives launched an 'outreach initiative', the 'Military Archives Imaged Identification Project'. The project, which also uses Flickr, seeks to identify personnel, locations, machinery as well as background information such as date and location. Other forms of crowdsourcing were utilized by a number of institutions and libraries.

Specifically, Clare County Library was able to generate a database of graveyard inscriptions through transcription projects carried out by individual and group volunteers. Built over time, and with direct input from the local community, the initiative has created a useful, crowd-sourced and locally developed tool that supports genealogy services as well as other historical research. The Chester Beatty Library's 'Islamic Seals Database' also employs crowd-sourced means to gather information on 'seal impressions found in Islamic manuscripts'. The final results of this process form 'an online, interactive database'.

This type of user engagement is a tangible connection to cultural heritage. It provides an opportunity for individuals to contribute to the national, historical narrative, and moves the user from a passive observer to an active participant in a national memory institution. The use of social media to promote and enrich collections, creates online communities that are connected to national treasures and, through simple engagement powered by social media, can open collections to the physical (and online) audience (citizens/netizens, diaspora, etc.) for whom it exists. Promoting this type of user engagement with content solidifies and generates relationships between the cultural institution and the audience, and enhances the user's

experience as they feel connected to collections or content on display through these various online interactions.

Developing resources and promoting collections through multiple digital channels was also complemented by developing resources for multiple devices, for our interviewees were cognizant of the need to develop mobile-friendly services. A number of the university library catalogues, including the Maynooth University, can now be accessed through mobile apps. The University of Limerick, along with the National University of Ireland Galway and the Royal Irish Academy, also developed a mobile app, 'Ireland Under Siege', which, as an augmented reality application, interacts with the landscape of important Irish battle sites.

The development and provision of mobile-friendly services, such as the examples provided, are a response to, and indicative of, changing Internet usage patterns. They reflect specific requirements and use cases derived from developments in new media content delivery and, as Prince (2013) outlines, reduce barriers to user engagement as mobile-friendly services (e.g. websites optimized for mobiles) focus on 'providing content into the right channels for the audience' (Prince, 2013, p. 354).

The digital tools mentioned so far are focused on the reuse and repurposing of image, text, audio and moving image. However, our interviewees also discussed digital tools and resources for the rendering, presentation and manipulation of numerical, geospatial and statistical data sets. As Catriona Crowe, senior archivist at the National Archives of Ireland, points out, 'today's social science data always eventually becomes humanities data', indicating that institutions such as AIRO (All Ireland Research Observatory), which deal with spatial data, are capturing and organizing data that becomes part of the historical narrative and of interest to memory institution and heritage organizations.

Rendering numerical, geospatial and statistical data through mapping, graphing and other visualizations captures the audience's attention and provides users with an opportunity to visually engage with content and the information contained therein. Mapping tools such as AIRO's National Census Mapping Viewer tools also provide an opportunity to manipulate, interrogate and analyse these datasets, and is but one example of a number of in-house tools developed by various institutions and archives.

Challenges: preservation and sustainability

The tools described above are content- and data-type driven, and in many cases represent traditional archival material, albeit in digital form. Given the fact that national heritage institutions, and other independent archives, are invariably concerned with historical records and artefacts, it is not surprising that the content delivered in the majority of these cases is digitized material, consisting of images, audio and moving image files organized and managed in, and by, various content management and database systems. The tools developed support the presentation, interaction and engagement

with digital objects that are derived from real world objects and are a result of digitization projects.

These projects create digital surrogates and serve as a means to share the content of an archive or library, and help preserve the integrity of primary sources as they reduce the physical handling of, often fragile, material. However, once an object, artefact or source has been digitized, we are faced with new preservation issues, specifically the long-term preservation of the digital surrogate or object. As Daniel Cohen and Roy Rosenzweig observe, 'digitization is not yet the preservation silver bullet' (2006, p. 244). The imaging of material, to archival standards, can create storage, as well as content delivery challenges. A number of our interviewees cited this as a major challenge, the solution to which was often the creation of lower quality images – a compromise that may not fit all archival or cultural content, or indeed institutions, if images of last resort are being created. Audio and moving files create even greater storage challenges, highlighted by the fact that 'a two hour film of [digital] preservation quality would be approximately 6 terabytes' (O'Carroll and Webb, 2012, p. 25).

However, we are now entering a new phase for digital cultural heritage material, born digital (and digital only) material. Cultural institutions and archives must be able to handle new types of archival records that include text documents, spreadsheets, databases in various proprietary, open source and bespoke formats, not to mention emails, online ephemera (blogs, posts), websites (both static and dynamic), software applications and programs, and as well as unstructured and structured big data sets.

A number of institutions highlight this as a challenge they face – one archive spoke of the difficulty archiving the collection of a now defunct political party because it consisted of desktop computers – i.e. the CPU – as well as documents, and had ethical as well as technical challenges. As stated, managing and preserving digital objects is resource intensive. Yet, while we may concern ourselves with long-term preservation of digital objects, we must also consider the preservation of the tools that provide access and functionality to these digital objects and data sets. Long-term preservation is 'sustainable access' – it must consider the form as well as the functionality of the digital object.

The tools themselves become our digital cultural heritage; they are part of a national historical narrative that requires public interaction and engagement. The preservation of user-generated content is also of concern as in many ways they become primary sources themselves – they document and are an expression of contemporary political and social concerns and interests. It is in this sense that we must also seek to provide long-term preservation or sustained access to digital resources and tools, but also to the digital objects that drive these.

A key step in the process of obtaining a TDR certification is identification of the repository's 'designated community' (Trusted Digital Repositories: attributes and responsibilities, 2002). It was reported at the Digital Preservation Coalition, Workshop on Trust, (Dublin, May 2013) that this first step is one that commonly causes difficulty. While many institutions have a

remit to serve the public, in reality, it rarely services everyone. The National Gallery, for example, identified their public in layers: academic researchers, school students, lay researchers, and only then, the general public.

When an archive creates user tools, it should similarly consider for whom they are being developed – at times the answers may be unexpected. Clare County Library discovered that their public/users also included diaspora, Irish emigrants abroad who engaged with their online resources in order to maintain a connection to their home county. A challenge therefore is defining the various user communities so that tools are developed to meet an appropriate need, at an appropriate scale.

Conclusion

Providing access to our digital cultural heritage is an important task. As a result of this interview process, we found that access alone does not suffice given the growth of online cultural heritage, as well as user expectations in terms of data use, reuse and interaction. As such, the development and provision of user tools that support innovative analysis, visualization and curation of digital content is an important feature of digital archives, libraries, repositories and other such systems.

Providing users – that is, researchers, scholars, the general public, etc. – with the means to interact with content, whether through a visualization, curated exhibitions, user light-boxes, social media sharing or the ability to add comments, through multiple channels, on multiple devices, enhances the user's experience and takes into account significant changes in user expectations and online habits.

Apart from enhancing the user's experience, a useful, usable, used digital tool will also promote and secure future engagement and attract visitors to both the online and offline (physical) sites. While this chapter describes current tools, many of the cultural heritage institutions, archives and libraries we interviewed spoke of future developments in this area, and were acutely aware of the need to develop tools to present and render born digital content. The interviews revealed a number of challenges that our interviewees faced in this area. Many users spoke of the technical limitations they faced, particularly in an age of budget cuts. While short-term funding may allow for the development of tools initially, in the long term there is the very real danger that they may become obsolete as there are not technical resources to maintain them as software systems develop. Yet tools that support resource, content and knowledge discovery are an important aspect of current developments within the Irish context, and this was highlighted during our stakeholder interviews and is reflected in DRI's requirements and policy decisions and implementation.

Interviewees were overtly aware of the need to preserve digital objects. However, the need to preserve the digital tools that support engagement and interaction with those objects was an area that many had not yet considered. Digital tools, which support dissemination and access – e.g. websites,

finding aids, etc. – as well as tools that support additional functionality – e.g. visualization, mapping, analysis, etc. – are, however, more difficult to 'preserve' than, for example, digital objects like text, image or audio files. Yet, regardless of the complexity of software preservation and/or emulation, we must consider how digital tools and resources can be sustained or maintained into the future – not necessarily for their functional value per se, but for the value they bring to our digital cultural heritage. National Trusted Digital Repositories, such as DRI, must consider how they might support this type of preservation activity.

Note

1 Accessed 8 June 2016.

References

Carr, N. (2010) *The Shallows: How the Internet is Changing the Way We Think, Read and Remember*. New York: W.W. Norton & Company.

The Chester Beatty Library, Islamic seals database. Available from: www.cbl.ie/islamicseals/Home.aspx (accessed 8 June 2016).

Cohen, D. and Rosenzweig. R. (2006) *Digital History: A Guide to Gathering, Preserving and Presenting the Past on the Web*. Pennsylvania, PA: University of Pennsylvania Press. Available from: http://chnm.gmu.edu/digitalhistory/preserving/7.php (accessed 8 June 2016).

Dix, A., Finlay, J., Adown, G. D. and Beale, R. (2004) *Human–Computer Interaction* (3rd edn). Essex: Pearson Education.

Ireland Under Siege. Available from: www.irelandundersiege.com/ (accessed 8 June 2016).

The Military Archives, Military Archives Image Identification Project. Available from: www.militaryarchives.ie/?id=888 (accessed 8 June 2016).

National Archives of Ireland, *The Life and Works of William Butler Yeats*. Available from: www.nli.ie/yeats/main.html (accessed 8 June 2016).

National Gallery of Ireland. Available from: www.nationalgallery.ie/ (accessed 8 June 2016).

National Library of Ireland. Available from: www.nli.ie/ (accessed 8 June 2016).

National Library of Ireland on the Commons. Available from: www.flickr.com/photos/nlireland/ (accessed 8 June 2016).

O'Carroll, A. and Webb, S. (2012) *Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences*. Maynooth: National University of Ireland Maynooth, Available from: http://dri.ie/digital-archiving-in-ireland-2012.pdf (accessed 8 June 2016).

Prince, A. (2013) Engaging digital audiences. *Journal of Digital Media Management*, 1(4), 349–58 (accessed 8 June 2016).

RTÉ Archives. Available from: www.rte.ie/archives/exhibitions/ (accessed 8 June 2016).

Trusted Digital Repositories: attributes and responsibilities, an RLG-OCLC report. (2002) Available from: www.oclc.org/resources/research/activities/trustedrep/repositories.pdf (accessed 8 June 2016).

10 From Europeana Cloud to Europeana Research

Tools, users and methods

Agiatis Benardou and Alastair Dunning [1]

Introduction

At a time when digital research infrastructures across Europe attempt to support digital research in the Humanities through meeting documented user needs of the so-called related research communities, we, the Europeana Cloud project team working on user requirements and community engagement, considered that it is not particularly agreeable to view the Humanities and the Social Sciences as two separate disciplines (or to consider any particular disciplines within them as self-contained), when almost all empirical studies on user requirements in the context of digital research infrastructures have revealed numerous underlying variations in the ways in which researchers carry out research.

Work conducted in the context of other EU infrastructures (e.g. EHRI, ARIADNE *et al.*) has already convincingly showed that, similarities notwithstanding, research methods, activities, resources, research goals, as well as interaction with data, both digital and analogue, vary between subdisciplines in the Humanities and the Social Sciences. Such an approach, we believe, must be considered in the design of effective user-centred digital research infrastructures. Needless to mention, research subcommunities are neither autonomous nor self-sufficient. Only those digital research infrastructures that address this granularity in subcommunities of practice and their respective user needs are to be meaningful to researchers. As Ian Hodder put it recently:

> One of the things I thought was very important during the 1980s was the idea that culture is meaningfully constituted. I still think that is right. But now, I put the emphasis on the 'meaningful constitution' rather than the 'cultural' bit. . . . prefer to break it down and talk about the various processes that constitute it.
>
> (*To The Trowel's Edge: An Interview with Ian Hodder*, http://ucexchange.uchicago.edu/interviews/hodder.html (accessed 7 October 2015))

To this end, and within the course of our work on Europeana Cloud (the project that will be the main topic in this chapter) we soon realized that a

focus on particular research communities and, in some cases, in selective aspects and areas of their research lifecycle, would provide a much better insight into both the user requirements, tools use and content strategy recommendations than a generic approach to Humanities and Social Sciences researchers per se.

Europeana Cloud (2013–2016) was a Best Practice Network coordinated by the Europeana Foundation, designed to establish a cloud-based system for Europeana and its aggregators. Europeana Cloud provided new content, new metadata, a new linked storage system, new tools and services for researchers and a new platform – Europeana Research.

Through a user-centric approach based on work conducted within DARIAH (preparing DARIAH and affiliated projects such as EHRI and CENDARI), CESSDA (through Data without Boundaries and DASISH) and NeDiMAH, part of the work in Europeana Cloud focused on the identification and assessment of researcher needs in the Humanities and the Social Sciences within the Cloud environment, as well as their interaction with Europeana content and metadata, while ensuring the active engagement of the wider research community. Within Europeana Cloud we investigated researchers' use of and interaction with Europeana and The European Library (TEL) material, as well as identified, documented and assessed existing tools and services from the domain of discovery, curation and use of The European Library (TEL), and Europeana material. Moreover, we explored and analysed the usefulness and relevance of the existing content within Europeana, alongside new content that could be ingested as part of this project, as well as other types of content. Not least, within our work for Europeana Cloud, we sought to actively engage the Humanities and Social Sciences research communities in establishing user requirements for the development of Europeana Research.

The starting point of this exhaustive piece of work was a thorough literature review of user information behaviour studies, followed by a web survey carried out at the beginning of the project. The Europeana Cloud Web Survey was built on, and ran in parallel to, related surveys by EHRI and DARIAH-EU, and the Survey greatly benefited from both of them. In addition to this quantitative part of the work, three other in-depth methods were employed to make sense of user requirements and content needs for Europeana Research: case studies, a study of researcher behaviour in using data at the API level, as well as a close look at specific research themes for aggregating digital content.

The information collected provided evidence revealing how Europeana Cloud could provide a valuable platform for the Humanities and Social Sciences research community – Europeana Research – and how it could explore and answer some key questions pertaining to research infrastructures:

- What advantages might the aggregation of digitized content from Europeana cultural heritage institutions bring for researchers?
- What benefits might different levels of authentication over such an aggregation of content permit?

- What kind of tools could be built on top of that infrastructure and which of them are the most appropriate to support this kind of communities – i.e. tools for transcription, tools for data analysis, tools for visualizing or enriching metadata, tools for sharing?
- And, importantly, how might these findings and information relate to other infrastructures, both existing and future ones?

Within this broader picture, the purpose of the case studies (that will be described later in this chapter) was to investigate what researchers use innovative digital tools for when they work with the types of material that a future Europeana could contain. We achieved this by analysing actual cases of how researchers within three disciplines work with three innovative digital tools. The three cases were selected through what can be referred to as purposive sampling. (For a more thorough discussion of this, see Teddlie and Yu, 2007, pp. 80–83) – that is, the cases are selected as illustrative specimen to provide a useful basis for the discussion. It is important to note here that the qualitative approach employed aimed to offer general insights into researchers' activities through detailed investigations of particular instances, as well as the implications of these insights on the future development of Europeana and Europeana Research. It should be noted here that these user scenarios do not constitute a review of digital tools or research practices in total, nor should they be taken to be representative or comprehensive of Humanities and Social Sciences research as a whole.

The case study as a methodological approach involved multiple sources and methods of gathering data which provided enough material for in-depth analysis. Thus, through a case study approach, some research practices and activities were illuminated from a new perspective and in more depth than had been previously undertaken in earlier tasks.

Case studies employed a mixed-method approach to a varied material and although it was not possible to generalize to populations, they were generalizable to theoretical and analytical propositions (Yin, 2013, pp. 40–44; see also Flyvbjerg, 2006. pp. 219–245). For an investigation of 'actual use', case studies offered a highly effective way of approaching the material.

In this study, an innovative digital tool is taken to refer to a tool that either performs functions that were previously unavailable, or that performs already available functions in a qualitatively different way. Examination of how researchers used the tools required them to be well acquainted with the tools, which, along with the inclusion of research publications as material, meant that there was a limit as to how recent the tools could be in order to be included in the cases.

Our case studies focused on the Europeana 'content types' (image, text, sound, video and 3D – the final type is somewhat vaguely defined in Europeana, however, and represents only a small proportion of the content and thus has been largely excluded from this study). The actual and practical use of the tools should thus involve material of types corresponding to the

Europeana content types. Rather than trying to find researchers and tools that work with current Europeana material for this task, the cases selected aim to investigate tools of kinds that might be used in future research on Europeana content.

The case study research approach involves a variety of data collection and analysis techniques that allowed for in-depth analysis and representation of complex phenomena (Yin, 2013, pp. 9–19). Even though the approach lacks a strictly defined set of rules and procedures, it is a suitable methodology for combining different data sources and research procedures for in-depth analysis of interrelated aspects of social phenomena. This makes the case study useful for examining innovative tools and services in Social Sciences and Humanities research.

One of the most comprehensive and widely cited works on case-study research methodology is Robert Yin's *Case Study Research* (2013, pp. 9–19), which discusses a number of central components in the research design of case studies. For the purpose of the present study, two of these were of particular relevance. The first of these was the study's questions, which are of great importance since they will have implications on every other aspect of the study – for example, the cases selected in this task will be presented following the structure of the research questions. The second component regarded the unit of analysis. At this stage, the researcher had to define what the case is – e.g. an individual, an event or some kind of process. These two components are highly interrelated as the nature of the question(s) must depend on the unit of analysis and vice versa. The following sections define the cases and describe the questions for the case studies of this task.

Three disciplines, History, Sociology and Education, were selected as they were likely to work with material from, or similar to material available through Europeana – that is, cultural heritage material of the four major Europeana content types. Moreover, History and Sociology were selected because they are fairly typical for the fields of Humanities and Social Sciences, respectively, whereas Education is a research field that straddles the border between the two fields.

The cases themselves centre on one particular tool and the way in which a researcher or group of researchers make use of that tool. The case study thus focuses both on the tool and its documentation as well as researchers' experiences of using the tool with one or more types of research material in several stages of research.

The application of an innovative tool in research can be characterized by a variety of interrelated dimensions involving the research traditions in a particular scientific discipline and the role of digital tools in it; new research trends and the digital tools; functionality of the tools, and the skills and experiences of the users; compatibility of the innovative tools and other tools typically used by the researchers in the discipline, etc. The overarching research question is therefore how researchers in particular disciplines use digital tools, and it can be broken down in five related subquestions:

- **How does the tool work?** Data types that the tool works with; functionality of the tool in different stages of the research process; the level of competences needed to work with the tool etc.
- **In what way has it been used?** Disciplinary background of the users, types of data they have worked with; description of the research project(s) the tool has been used in; the scholarly activities the tool has been used for, etc.
- **What are its strengths and weaknesses?** The motivation for choice of the particular tool; the problems encountered while using the tool, improvements needed for the tool and the gaps in the discipline that could possibly be addressed by new tools.
- **What is its further potential**? How could it be used with Europeana-type material? The potential of the tool in applying it to Europeana material; the changes needed for the tool and for the material.
- **What kind of new research (methods) does it facilitate/enable?** The issues and questions in the discipline that can be addressed better by using the tool.

The satellite (complementary) cases aimed to answer the following questions:

- **How has this alternative tool been used in the discipline in relation to the main tool?**
- **What are the strengths and weaknesses of the alternative tool in the discipline, working with different data types, etc.?**
- **How can the alternative tool be potentially used with Europeana material?**

These questions have been elaborated in data-collection instruments (as interview questions) and explored in the analytical process in data material.

Selection of cases

From each discipline, one digital tool was selected: *NodeXL* for Sociology, *Transana* for Education and *HyperImage* for History. Thus, we are in compliance with the suggested sample size of three to five case studies in order to be able to perform analytic generalizations based on the results. For further discussion on sampling schemes and sample sizes, see Onwuegbuzie and Collins, 2007, pp. 281–316. Suggestions for appropriate tools were solicited largely by querying researchers in our network, not least those who had in some way assisted in previous tasks (in particular the Expert Forums).

The researchers participating in the Forums were selected by the relevance of their research to potential use of Europeana material. Other aspects taken into account in a selection of cases were related to the available documentation of the tool and its application in research; coverage of Europeana content types; and relation of the case to new trends, new research environments and/or new research questions arising in the disciplines.

The application of the tool should be documented in publications. By its nature, a 'cutting-edge' innovative tool is used by few researchers, and there would be little, if any, documentation available while the well-documented tools may be considered mainstream rather than innovative. The tool selected for the case study needed to be new to researchers in the discipline, but at the same time fairly well documented in publications.

It should be noted that no attempt was made to identify cases that engaged with current Europeana material, for two reasons. Europeana as it stands today does not offer content and metadata of sufficient quality to make it an attractive tool for research or source of research material (Europeana Cloud Deliverable 1.5 (1 of 4) Expert Forum Case Studies Report; Deliverable 1.5 (2 of 4) Expert Forum Tools & Content for Humanities Research Report; Deliverable 1.5 (3 of 4) Expert Forum Tools & Content for Social Sciences Research Report). Moreover, identifying researchers, in particular social scientists – who worked with Europeana content as well as put digital tools to innovative use in their work – quickly proved more than challenging (it would be hard, if not impossible, to identify such use, especially in disciplines that are not currently involved in cultural heritage material).

More importantly, the tools that were selected are such that they – or tools with similar functionality – can quite conceivably be used for future research on Europeana content. The forum reports that an important feature of a future Europeana could be the ability to export content to third-party tools for analysis, and the three tools are the kind of tools to which Europeana could be expected to export content. It is, of course, impossible to predict exactly which particular programs will be used with Europeana even in a few years' time, but *Transana*, *NodeXL* and *HyperImage* all belong to categories of tools that would likely be employed with exported content.

The three tools finally selected all included functionality that enabled or facilitated what was broadly considered innovative research methods and covered the major Europeana content types as well as metadata. *Transana* facilitates collective work with audio and video transcription and qualitative analysis; *NodeXL* enables mixed-method network analysis; and *HyperImage* is a tool that makes it possible to annotate images and parts of images. For each of the tools, an appropriate researcher or group of researchers was chosen (mainly on the basis of publications or previous contacts) for closer study.

For Education, two alternative tools that support or facilitate similar scholarly activities were selected in the course of the data-collection process: *NVivo*, a commercial and well-established tool in the research community, and *Voyant*, an innovative web-based text reading and analysis environment. These tools form the basis of the two satellite cases – that is, complementary cases of the use of digital tools, which, however, are not innovative per se.

Data collection

In order to capture the multifaceted nature of how digital tools and services are employed in research, the uses of innovative tools were studied by integrating data from different sources. Yin (2013) lists six sources of evidence (documentation, archival records, interviews, direct observations, participant-observations and physical artefact), of which three can be applied to the subject under examination here:

- analysis of published documentation (of tool and research applying the tool);
- semi-structured interviews with researcher;
- observations (experiences) of how the tool works.

For each case, information was gathered via semi-structured interviews with the researchers, desk study of the available documentation of the tool in research as well as observations of the demo versions or similar examples of how the tool worked. Six semi-structured interviews were made between November 2013 and January 2014. They were both face-to-face and web-based interviews, and lasted about one hour each.

Analytical process

The semi-structured interviews were recorded. The audio files were transcribed and coded using different methods and tools:

- transcribed and coded using *Transana* (for the *Transana* and *NodeXL* cases);
- transcribed and coded in MS *Word* and *NVivo* (for the satellite cases);
- transcribed and coded in Windows media Player and MS Word (*HyperImage*).

Published research was reviewed in order to present the novelty and value of the tool in the discipline (what types of research questions are researchable because of the tool, what new research questions can the use of the tool facilitate), as well as the stages of the research in which the tool has been used, and the data import and export procedures.

Observations of the tool in function were made by using the documentation of the tool, personal experiences with the tool (*HyperImage*, *NodeXL*, *Transana*) using sample data sets (*HyperImage*, *NodeXL*) and actual research material (*Transana*). These were of particular importance in discussing the research activities supported by the tools and their possible applicability to the Europeana content.

Data was analysed and presented to provide answers to the main case study questions and to highlight the various applications of the digital tool

and its relevance to facilitating more use of Europeana content. Finally, analytic generalizations were made, where appropriate.

For the purposes of this chapter, we will present the case study developed primarily by Trinity College Dublin on *HyperImage and History Research*. Research into History, particularly research including images, requires a tool or suite of tools that can recognize and also annotate images in a way that allows for both qualitative and quantitative analysis. Traditional Art History research involving images has used print images. At best, these may have been photocopied or reprinted for annotation purposes. Digital tools are not typically involved in research practice within the field of Art History, although they are increasing in usage. Most traditional work within Art History would be focused instead on the interpretation of the artworks rather than doing something digital with them. It is an open source resource and sits on an Omeka platform – see http://omeka.or – as an exhibition and archive. There are one or two exceptions. For example, *Gothic Past*, a project based in Trinity College Dublin, is an online image archive of artefacts and architectural items of significant historic value found around Ireland.

One tool that enables digital annotation of images is *HyperImage*. This tool is the result of collaborative research between the Humboldt-Universität in Berlin, Leuphana University Lüneburg and Zweitwerk, a content software company. The tool was built in response to a significant dearth of tools for image archiving and annotation in comparison to what is available for texts. In 2007, just after the team began development on *HyperImage*, image annotation was limited to metadata text. The resulting software was to be a 'single integrated system, whereby each stage will be based on open standards and therefore compatible with other systems (Warnke *et al.*, 2007. It is this claim to direct compatibility with other systems that prompted us to investigate its suitability for use with Europeana. The *HyperImage* software is available to download for free from *SourceForge* and is open source.

In investigating this particular tool and its uses in History and Art History, we spoke to Karolina Badzmierowska, a Ph.D. student in Art History who uses images as the main focus of her research. She uses *HyperImage* for annotating those images. We also looked at some pilot projects connected to the development of *HyperImage*.

Functionality and features

According to the *HyperImage* website, '[the] HyperImage platform supports the linking of (audio)-visual objects, texts and mixed-media documents'. The tool is web-based, which allows the user to access it from multiple computers, as information is stored online. There is also an option to store the information on your local server, if required. Users work with the *HyperImage Editor* and *HyperImage Server* to annotate and store images and associated data.

HyperImage allows the user to 'drag-and-drop' images into the *HyperImage* interface ready for annotation, either from their image management software

(such as Windows Explorer or Adobe Bridge) or from a website. *HyperImage* works by allowing the user to select items, or 'elements', within an image and annotate them. It is then possible to link to similar elements within other pictures. For example, in the *HyperMedia* project, we can see highlighted elements within the image, the annotations on that image, and the links it makes to related images.

As an image annotation tool, *HyperImage* includes space for metadata. It comes with three metadata standards out of the box: basic, Dublin Core and CDWA (Categories for the Description of Works of Art). Dublin Core is an international standard and is encouraged for richer metadata purposes. Metadata fields can be freely tailored and extended to the user's needs on a custom basis. Custom fields and further metadata standards can be incorporated using the built-in template editor. Different metadata templates can be used in conjunction to form a complex set of metadata fields.

The drag-and-drop function for populating the *HyperImage* catalogue allows for single and multiple images to be uploaded to the tool. These images can then be filed into groups, depending on their category type within the context of the research. External URLs can also be created as objects, which can later be linked to one or more images or layers within the project.

HyperImage has its own terminology. Image files are referred to as 'objects'. Items within an object that are annotated, or tagged, are called 'elements'. These annotations are made via 'layers' into which 'polygons' are drawn around the shape of the element you want to annotate.

Layers within images are not a new concept. Image manipulation tools such as *Adobe Photoshop* and *Corel Draw* have been using layers within images for some time. These, too, allow for changes to be made without altering the original digital image. The difference with *HyperImage*, however, is that these layers enable hyperlinks to be created on them. The hyperlinks can link to an explanation of the object in the image, another similar object in another image, or to an external website, thus allowing for interaction between the images and their wider context. For example, within the tool, annotations are made on the layer as polygons, which can be colour-coded to identify individual annotations or groups of annotations around a topic within a single image. Layers can have a transparency level set to prevent annotations from interfering too much with the visibility of the image.

Multiple polygons can be drawn into one layer. This is particularly useful where there are several repetitions of the same element within an image (e.g. trees) or where a single element is in separate parts within the image. Links can be made on elements of images based on themes.

Outputs

The outputs from the Reader and Editor can be downloaded in the XML mark-up language as a back-up. This can then be used to re-create the information on a website or user's server. This download takes the form of a

directory or zip-file, which contains the full structure, metadata, and links and relationships in XML format. The images are supported as image files within the directory and are rendered in different sizes for use online. By exporting in this fashion, it allows the user to import the directory to their server, or even to a WordPress site (or similar) for publication.

> The HyperImage Reader essentially shows the fruits of one's labour as the published data. A preview mode for the HyperImage Reader is available which allows you to see how this online version develops as you annotate more and more images. HyperImage Reader is currently a separate platform, which can be used alongside the HyperImage Editor.

Using *HyperImage* in Art History

Our informant is a Ph.D. student working within Art History. She applies Digital Humanities techniques to her research (images within the Fagel Collection). Her research requires annotation of these images, alongside the use of spreadsheets in Excel. She also uses Zotero to manage citations within her research. By her own admission, Badzmierowska had not done a great deal of research into alternative tools for digital image annotation, but had heard of *HyperImage* through Twitter.

Our informant used *HyperImage* to annotate digital images taken from the Fagel Collection, which is held in Trinity College Dublin's Long Room Library. The Fagel Collection is a single collection that takes up 'a mile of shelf space'. The Fagel family in the Netherlands compiled it during the 18th century, comprising scientific drawings, maps and atlases, cartoons, broadsheet newspapers and political writings. The entire collection contains over 20,000 items. She hopes to work on as many of the scientific images as possible to analyse them for their artistic merit, as well as thematic, iconic and scientific links. Once she has completed her annotation, she will export this as XML in the format available from *HyperImage*. Her overall ambition is to create an online archive that will allow the public to interact with the images, discovering links between them. Her ideal set up will be via touch screen or gesture recognition, but she realizes that this may be beyond the scope of her Ph.D.

Pilot projects

The majority of publications that involve the use of *HyperImage* are only available as the online output of the project. Any journal articles or conference papers on the user-experience for *HyperImage* have, for the most part, been written by the *HyperImage* team. For that reason, we looked to pilot projects for examples of uses of the tool.

To test *HyperImage* once it had gone through the initial stages of development, the team established a set of pilot projects. The complete list of pilot

projects is too numerous to go into detail here (but see Websites, below). However, to see the basic principles of how *HyperImage* was used, we will look at *HyperTaxon*, *HyperGiotto* and *HyperMedia*.

HyperTaxon was a project dedicated to studying a small group of insects known as *Peloridiidae*. The aim of the project was to compile 'existing information' about the insects in one virtual space, thereby making it accessible to members of the public as well as academic researchers. The published version showcased images of the insects, the regions in New Zealand in which they can be found and anatomical information about the creatures. The HyperImage Editor allowed the *HyperTaxon* team to highlight different parts of the insect on the photographs.

HyperGiotto was an Art History project, which looked at the works of Giotto and his contemporaries from the Trecento period (14th century). It investigated thematic relationships among image data and metadata, using a combination of three tools: *HyperImage*, *Mneme* (https://mneme.huberlin.de). Mneme was the name that the Humboldt-Universität in Berlin used for a commercial asset management software developed by Zweitwerk. Humboldt-Universität is discontinuing this system as they now intend to use an open source alternative and the Census of Antique Works and Arts Known in the Renaissance. The project looked for similarities in elements within images of antiquities and picture what were then translated into sculpture in sarcophagi or friezes, or vice versa. An image of a carving on a sarcophagus from AD 190 is repeated again in a sketch in the 14th century, as well as a painting with a different subject from the same era. The metadata from these images was drawn from the Census, as well as from databases such as Prometheus and Mnemic.

HyperMedia was a project run by the University of Munich, which looked at cartoons and paintings created during and immediately after the French Revolution. The project managed to gather over a thousand items for annotation, and looked at how both commercial paintings and cartoon graphics were used in propaganda during this time, and explored the themes that were employed in those images to make their argument. This included the parody of nations through maps, symbolism incorporated into more serious artworks and motifs as a running theme throughout a series of images.

Prometheus and *HyperImage*

In the second phase of the project, the *i* team worked alongside Prometheus, Germany's largest art history database through the Meta-Image project (Dieckmann *et al.* (2010). *HyperImage* was employed as a means for users of Prometheus to annotate images within the Prometheus repository. It did not, however, allow users to upload their own images into the Meta-Image project. Instead, the project aimed to design a function whereby the user could upload images to their own Prometheus account.

MIRROR OF KĀŚĪ

A team in Heidelberg University used *HyperImage* to reproduce and update a previous project on A New Mirror of Kāśī (2011). The project had originally been published in 2001, but they used *HyperImage* to update the outputs of the project to allow for newer features that *HyperImage* can provide. This included grouping elements and better searchability between the images.

HyperImage's strengths and limitations

Strengths

Compatibility

According to the *HyperImage* website, it is compatible with most operating systems – Mac, Windows and Linux.

Interface

The interface is very easy to use, although the display is more functional than graphically pleasing. Indeed, once set up in the program, the interface is reasonably intuitive, although broken up into several windows, depending on whether you are working on a group of images, or an individual image. Images can be put into groups, either thematically, chronologically, by existing collection or by any other categorization means, which makes it easy to find items for annotation. Labels for groups are free text, so this makes creating categories very straightforward.

HYPERIMAGE READER

Once items have been annotated, they can then be previewed in the HyperImage Reader. This makes it very easy to see your project as the public would see it, thereby allowing the researcher to see any problems immediately and update or amend them accordingly.

LAYERS AND THE POLYGON TOOL

Once you are familiar with the interface set-up, creating layers and polygons is fairly straightforward. The Polygon tool allows for quite a high level of accuracy in outlining the element you wish to annotate, and also allows the user to modify the colour and transparency of the annotation so that it is either invisible to the viewer until they hover over it, or is fully visible.

DOCUMENTATION AND SUPPORT

The *HyperImage* website is very comprehensive. Currently, most of the information about the tool is available in both German and English,

although there are still some documents that are only available in German. For background information on the tool, there is little printed information. Members of the *HyperImage* team have produced journal and conference papers on its development. These can be found in Appendix II: Documentation of HyperImage. Researchers found the development team to be the best resource for support, however. The team will provide a training session via Skype if requested, which is very useful for giving basic information about how the tool works, creating links between objects and links to external websites.

Limitations

Compatibility

The reader and editor are not compatible with Google Chrome. Furthermore, *HyperImage* requires Java 7 to run. This is difficult for Mac users working on an OS older than 10.6 as Java 7 is not supported by previous operating system version.

Interface

The *HyperImage* team recognizes that the current interface on *HyperImage 3.0* is not an attractive one. They are currently working on improving this for *HyperImage 4.0*. The multiple windows within the tool make it difficult to keep track of work.

Pop-up windows that open when saving or creating a URL for linking purposes remain on screen after the link has been saved. The save function does not confirm that a change has been saved, making it difficult to know if it is safe to close the pop-up or not.

Future direction for HyperImage

HyperImage aims to expand its use across major EU projects. They have been in discussion with *TextGrid* with a view to establishing the tool for use on *TextGrid Laboratory* and *TextGrid Repository*. We have already seen how *HyperImage* teamed up with Prometheus to enable annotation of repository items. It is the hope of the *HyperImage* team to build on this experience and make further reaches into working with trusted repositories.

HyperImage also aims to expand its scope to include annotation functions for text, video, audio files and 3D objects. They also plan to investigate how to integrate semantic web technologies and to increase the collaborative aspect to the tool to allow greater functionality within group projects, particularly in the classroom. One of the key goals for *HyperImage* is to increase its use in teaching environments.

In *HyperImage* 3.0 and 4.0, the developers hope to integrate the Reader into the Editor platform in order to reduce the number of programs needed

to run the application fully. This should make *HyperImage* more streamlined and allow for instant publication of a project.

The current versions, *HyperImage 2.0 and 3.0*, do not allow for statistical data to be taken from the annotations. However, the team is working on *HyperImage 4.0* in parallel, and hopes to introduce a means to draw statistical data from the annotations that can be used in further quantitative research (personal correspondence with Jens-Martin Loebel, Humboldt-Universität in Berlin, on 14 December 2013). It is envisaged that this statistical data will be in the form of .xls or .csv, although they are open to suggestions for what might be most useful to researchers. A spreadsheet-ready format might make most sense, though, as these are most easily fed into common statistical analysis tools such as SPSS (personal correspondence with Jens-Martin Loebel, Humboldt-Universität in Berlin, on 27 December 2013).

How could image annotation software be used with Europeana content?

Europeana is currently rich with metadata for images, be they photographs, paintings, maps or diagrams. Images such as these could be used in any discipline: a linguist may wish to use an older map of a specific region they are studying; a sociologist may wish to look at photographs of groups of people, or street images that show everyday life. The same can be said for historians, art historians, geographers and botanists to name but a few. In nearly all cases, the research requires cohesive links between themes within the images. Therefore, the match between Europeana content and image annotation software such as *HyperImage* is an obvious and natural one.

Exporting images from Europeana

In the first instance, exporting images may be a simple case of adding an option to the Europeana interface that allows the user to export either individual or multiple images to an image-annotation tool. If multiple images are required for annotation, this export function could also be built into the My Europeana portal where links and Europeana items are stored. Multiple Europeana content items, along with their associated metadata, could then be directly exported from My Europeana into a user's image-annotation software. To save the user from having to specify which items within their My Europeana file they want to export through clicking and opening each individual item, the option could also be available whereby they could right-click on an item and 'add to export list'. This export list could then be available from My Europeana, which they could export to tools such as *HyperImage*, or indeed any of the other tools we have discussed within this chapter.

As Europeana aims to deal with text, video, audio and 3D as well as images, these content types would match perfectly with *HyperImage*'s planned expansion. In the case of *HyperImage*, it has shown that it is capable of working with

repositories such as Prometheus, although these may be on a smaller scale to Europeana. Caution might therefore be exercised in making any large-scale plans initially.

This case study report complements other research within Europeana Cloud, providing insights into how the Europeana collection could be utilized by researchers within the Humanities and Social Sciences. Tools similar to those explored in this context could potentially be effectively applied to the content types of Europeana. For this to be achieved, however, several considerations need to be made with regard to the development of Europeana Research.

In order to be a useful source of research material, Europeana needs to provide a broad selection of data and metadata formats to ensure applicability in a wide variety of analytic tools and long-term stability. Unlike other aspects of our work, enriched metadata was not stressed as important for a future Europeana by the interviewees in this study. Developing Europeana for researchers was not among the subjects of the interviews, however, so the lack of discussions about metadata enrichment should not be taken to imply that such functionality would not, in fact, be appreciated by the research community. What metadata are present (whether added by users or the contributing institutions) should be easily accessible in terms of format as well as comprehensibility. A researcher about to write a publication based on analyses of Europeana content should be able to export analytic information – e.g. transcripts of clips created by another researcher or illustrations of the analysis of a certain network, into a text or graphics editing application, either third-party software or within My Europeana. Those pieces of analytic data could then be the points of departure for scholarly analysis.

Basic visualization and statistical functions within Europeana Research would provide the researcher with a means of evaluating the relevance of a certain selection of data for a specific purpose. Thus, the researcher can avoid downloading and ingesting large data quantities only to discover that they hold no interest to the project at hand. Another useful tool within Europeana Research would be a note-taking tool in which the researcher might document all steps of the research process. This feature would allow researchers to go back and trace analytical decisions made in their own previous research as well as reconstructing other researchers' methods and decisions, and would be invaluable in the documentation process of research data management (RDM).

My Europeana should feature functions that enable users to organize Europeana content into collections. Differentiation of collections should be possible, allowing personal collections to be kept separate from group collections (accessible to members of a user-defined group – e.g. a research team) and from publicly available collections. Europeana Research should consider incorporating a set of collaborative functions connected to the collections function. These might include the possibility

to share analytic data and decisions within Europeana Research, or collaboratively annotate data and metadata.

Being a pan-European project, Europeana Research needs to provide multilingual access to the data. More specifically, a function for users to translate metadata and some types of research data into other languages would be highly useful to facilitate access to a wider array of researchers.

Thus, this particular case study provides the following key recommendations for Europeana Research:

- Provide import and export functionality for a wide variety of formats (awareness of different formats and the problems they bring with them is important).
- Functionality that enables addition, import, export and collaboration metadata for Europeana records should be flexible and easy to use.
- Functions for mark-up, organization and collaboration in My Europeana would be highly useful.
- Provide basic analytic functionality and the possibility of documenting research decisions.
- Support for (the creation of) multilingual data and metadata would greatly increase the material's usefulness.

Note

1 Our work on Europeana Cloud was the outcome of a great collaboration: Eliza Papaki, Nephelie Chatzidiakou and Costis Dallas from DCU/ATHENA RC, Vicky Garnett from Trinity College Dublin, Stefan Ekman, Ilze Lace and Dimitar Popovski from the University of Gothenburg, Owain Roberts from the National Library of Wales, Norman Rodger from the University of Edinburgh, Marian Lefferts and Ingeborg Versprille from CERL and Kees Waterman from DANS all worked with enthusiasm and efficiency. Our thanks should also go to our Research Communities Advisory Board: Lorna M. Hughes (University of Glasgow), Leif Isaksen (University of Leicester) and Karina van Dalen-Oskam (Huygens ING) provided us with useful insights and comments throughout the project. Not least, special thanks should go to The Europeana Foundation and The European Library for their trust and encouragement, as well as to the continuous support and wisdom of our very own President Snow.

References

Benardou, A., Constantopoulos, P. and Dallas, C. (2013). An Approach to Analysing Working Practices of Research Communities in the Humanities. *International Journal of Humanities and Arts Computing*, 7(1–2), Edinburgh University Press, pp. 105–127.

Dieckmann, L., Kliemann, A., Warnke, M. (2010). 'Meta-Image – a collaborative environment for image discourse'. In *Proceedings of the 2010 International Conference on Electronic Visualisation and the Arts*, British Computer Society, pp. 190–198.

Flyvbjerg, B. (2006) Five Misunderstandings about Case-study Research, *Qualitative Inquiry*, 12(2), pp. 219–245.

Hughes, L.M. (2014). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines, chapter in *The New Companion to Digital Humanities*, Schreibmann, S. and Siemens, R. (eds) Oxford: Blackwells.

Onwuegbuzie, A. J. and Collins, K. M. T. (2007). A Typology of Mixed Methods Sampling Designs in Social Science Research. *The Qualitative Report*, 12(2), pp. 281–316. Retrieved from: www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf (accessed 18 February 2014).

Teddlie, C. and F. Yu (2007). Mixed Methods Sampling: A Typology With Examples. *Journal of Mixed Methods Research*, 1(1), pp. 77–100.

Warnke, M., Kuper, H. and Helmers, S. (2007) *HyperImage – Image-Oriented e-Science Networks*. In: Max Planck Digital Library/German e-Science Conference, ID: 315522.0.

Yin, R.K. (2013). *Case Study Research: Design and Methods* (5th edn). Sage: Los Angeles, CA.

Websites

All Europeana Cloud reports can be accessed here: http://pro.europeana.eu/europeana-cloud/europeana-cloud-project-documents

A full description of the other pilot projects associated with *HyperImage* can be found at: www.uni-lueneburg.de/hyperimage/hyperimage/beispiele_E.htm (accessed 6 January 2014).

A New Mirror of Kāśī: Gengnagel and HyperEvaluation Team (M. Arnold, P. Gietz, E. Decker, G. Niebler, W-T Wu) 'Asia and Europe', Cluster of Excellence, Karl Jaspers Center, Heidelberg University 2011: http://kjc-fs2.kjc.uni-heidelberg.de/evaluation/Gengnagel04/PeTAL/index.html#T47593/ (accessed 7 January 2014).

Census of Antique Works and Arts Known in the Renaissance: www.census.de/census/home/view?set_language=en (accessed 6 January 2014).

Fagel Collection: http://kalden.home.xs4all.nl/verm/fagel-dublin-lecturesENG.html (accessed 16 December 2016).

HyperGiotto: www.uni-lueneburg.de/hyperimage/hyperimage/hyperGiotto_E.htm (accessed 6 January 2014).

HyperImage: http://manual.hyperimage.ws/ (accessed 10 December 2016).

Hypermedia: www.uni-lueneburg.de/hyperimage/hyperimage/hyperMedia_E.htm (accessed 7 January 2014).

HyperTaxon: www.uni-lueneburg.de/hyperimage/hyperimage/hyperTaxon_E.htm (accessed 6 January 2014).

Mneme: https://mneme.hu-berlin.de (accessed 6 January 2014).

Omeka: http://omeka.or.

11 Digital humanities research needs from cultural heritage looking forward to 2025?

Seamus Ross

Introduction

The cultural heritage sector and the humanities have a rich and long symbiotic relationship. This engagement has been transformed by the increasingly ubiquitous use of digital technologies to create, make available and study the records of human activity held by heritage institutions. Many changes are taking place in the cultural heritage sector just as there are advances in digital humanities. It is timely to assess what researchers need from the cultural heritage as we look forward to the middle of the next decade.

One cannot turn up in a university in North America or Europe without hearing that the humanities are in crisis. Many pundits argue that by embracing the digital the humanities can avoid their decline and even perhaps benefit from revitalization, if not renaissance. The humanities, like libraries, have never been more significant to society than they are now in the digital age – of course, it appears to take a humanist or a librarian to see that this is so. That said, the digital is more and more radically transforming the way in which we carry out humanities scholarship and the ways in which we represent, record, store and disseminate information. Digitalization changes what we can do with information and what we expect to do with it. At its heart, from ancient times, the humanities have been rich in analytic thinking and story-telling. Cultural heritage institutions play a fundamental role in collecting, curating and making accessible the resources needed for study and reflection. In this way, the interrelationship between the two domains is synergistic, if not symbiotic. In contemporary society, both the humanities and cultural heritage sectors are at a crossroads of relevance and public engagement, and bedevilled by the persistent attacks from managerial, audit and commercialization cultures, egged-on, perhaps, by narrow-minded bureaucrats.

As is now widely acknowledged, 'the cultural and scientific heritage held in or destined to be accessioned by our memory institutions provides a source of raw materials for intellectual, social, cultural, and economic prosperity in the 21st century' (Ross, 2004). Since the digital makes it

possible to perform new kinds of actions on cultural heritage objects and to interconnect them in new kinds of ways, the value of the heritage increases when it is made digital. And it is further enhanced when materials from diverse sources, such as film, audio, documents, artefacts, performances, data, works of art and much more can be explored in tandem. Humanities scholars armed with digital methods, algorithms, and software are, and will, explore this burgeoning content in new and multifarious ways (Hockey and Ross, 2008; Ross and Feeney, 1993; Schreibman *et al.*, 2016). They will make new kinds of discoveries. They will tell new kinds of stories about the past, the present, and our future. Cultural heritage resources in digital form can, as is widely recognized, serve as 'sustainable and renewable resources' that can 'be exploited in an ever-increasing diversity of ways' (Ross, 2006, 115). When represented in digital form, heritage content provides the natural resources to enable new kinds of research by scholars, students, and the general public, to make possible the construction of rich resources that can make learning and educational discovery more engaging, and to facilitate the production of new content and materials by the creative industries. Digital access to the cultural content of heritage institutions contributes to the broad growth of our emergent 'remix culture', cultural heritage 'bricolage' and 'lego-constructor making', all human activities that bring emotional enjoyment through varied intellectual activity, and harness natural curiosity and innovative spirit. There are many places where the virtual and the physical, and one very potent one is in the world of tourism – for several decades now professionals in museums, galleries and special collections have noted that finding materials online helps users to imagine places they would wish to be, and objects and cultural creations they would wish to see (MiBACT, 2014). In this, the improved access to digital content inspires tourism. Essentially, we might draw the conclusion from these arguments that the more varied and the more comprehensive the content the cultural heritage sector provides the digital humanities scholar, the more creatively we will be able to study it and the more exciting the discoveries we will be able to make. We must, though, recognize that the digital humanities do not drive bulk demand for cultural resources, and certainly do not provide the primary case to encourage digital cultural heritage professionals to make content freely available online.

Digital humanities and cultural heritage institutions

Let us begin by reflecting on the broadness of the humanities and the diverse nature of the cultural heritage sector. Then we can consider the complexity of each. The cultural heritage, from a simplistic perspective, is characterized by tangible and intangible outputs of individuals, co-operative entities, and societal institutions and processes associated with them. These outputs reflect, or indeed are, the essence of who and what

we were and in many cases are. The cultural record survives in many forms, including landscapes, architecture, archaeological remains, documentary records, artefacts, paintings, sculptures, stories, food, dress, and so forth. It is preserved in archives, libraries, museums, archaeological research centres, universities, and many similar institutions, including more recently in data repositories housed at these institutions or accessible to them. In some cases, legislation governing conservation of historic landscapes (and other places) or archaeological sites contribute to its preservation, and in other cases trusts or foundations ensure its curation over time. Much of it is dispersed, as, say, the Dunhuang manuscripts are, across multiple institutions (Shenton, 2012). And in some cases it survives only as documentation of performances, stories, or even only as oblique references such as those used by humanists to reconstruct lost philosophical, literary and historical works.

Given my engagement as a researcher with digital libraries, I supposed I could give a very simple answer to the question as to what digital humanities researchers need the cultural heritage to provide: 'give us lots of well-documented stuff and put it into highly functional open access digital libraries', and continue to increase the quantity of content, the diversity of types, and the way it is represented. We digital humanists will take care of the rest. Let me close this line of argument by stating the obvious: we do not merely need the cultural heritage sector to give us large amounts of diverse and multi-sensorial content, we also need the sector to partner with us to enable us to understand how the content they provide was created, how it relates to other cultural heritage materials of the same class or of different classes, and how it relates to other materials that have not yet been digitized. This is not to suggest that we do not need lots of managed content either. There are many weaknesses in this line of argument, such as that more content does not necessarily enable scholars to produce better scholarship, the production of digital representations of analogue resources consumes financial and human resources, the curation over time of these new digital representations requires continued investment by cultural heritage institutions, and this results in a diversion of resources, and users of the digital require new and different kinds of support from those working with analogue resources. All these shifts in the classes of content being managed and how it is managed and supported result in changes in organizational structures, practices and policies governing use, and access to both the analogue materials that have been digitized and their digital representations. For instance, some institutions drastically restrict access to the original analogue materials once they have been digitally represented, which can be unfortunate as there are many questions that can only be asked of an analogue object in the presence of the object itself.

There are evolving challenges in the curatorial responsibilities of cultural heritage institutions. Much of the holdings of heritage institutions in 2016 are primarily analogue and when humanists look to the content being made available in digital form, they are anticipating the digitization of the

resources. But society is now fundamentally digital –we conduct our business transactions in digital form and the records of those transactions are digital, we collect data about the world around us digitally, we use digital technologies to manage processes and activities, and we track them digitally, and we create inspirational outputs from art to videos to animations to music in digital form and enjoy them digitally. Heritage institutions need mechanism to select, acquire, curate and disseminate these new forms of output just as they are struggling to increase access to analogue through digitization.

As humanists, we need to know that the cultural heritage is not just working on building the capacity to handle digital materials if they are deposited, but that they are actively working to ensure that the creators of digital records are securing the material or, if they are not, that they are preparing the case to do so. After nearly 30 years of debate about digital curation and preservation during a time when the trend towards 'being digital' has accelerated exponentially, we find that little progress has been made in curating and capturing the digital. To illustrate this point, we turn to the use of electronic systems to manage government records and to an example from the United Kingdom. In 2016, BBC Radio 4 reporter Phil Tinline, under the banner, 'Too Good To Be Forgotten – Why Institutional Memory Matters', drew attention to a report by Alex Allan published in late 2015 which found what many of us who took part in the 1993 British Library British Academy Workshop on 'Electronic Information Resources and Historian' predicted would happen, that 'almost all [UK government] departments have a mass of digital data stored on shared drives that is poorly organised and indexed' (2015, 1). As a result, the transfer of records to the UK National Archives is not happening at the scale and in the way it should be. Moreover, Sir Alex found that '[m]ost of the focus is on text documents and emails', which he appeared to find concerning, as he noted that in contemporary information systems 'records can encompass a wide range of formats, some of which are complex and not easily stored even in newer systems' (ibid., 5)

Seen from the point of view of Paul Dacre, David Cannidine and Joseph Pilling, in their independent review of the '30 Year Rule' commissioned in 2007 by the Office of the Prime Minister of the United Kingdom:

> [t]he evidence we have heard is that ten years is probably the latest point by which digital records should be reviewed by departments and agencies and, if there are reasons to preserve them, transferred to a stable storage environment. If this is not done, there is a real danger that the official record may not survive, and it is difficult to overstate just how disastrous this would be.

What is true for the United Kingdom, given that the technological challenges of digital records vary little across geographical and political jurisdictions, is

most probably true for records created in other countries. So, in the context of the born digital, what humanists need, actually require, from cultural heritage institutions is more intensive and effective activism to ensure that government records are secured and transferred to institutions for long-term use. Similar challenges appear when we consider other forms of digital content creation, such as digital art or as the documentation of events. In the case of the 2012 Olympic and Para-Olympic Games in London, Andrew Rackley demonstrated that the preservation of 'born-digital content generated by impermanent creators such as LOCOG [London Organising Committee of the Olympic and Paralympic Games] necessitated early intervention and, more significantly, that memory institutions take custody of London 2012 content' (2016: 247). This need for early intervention was recognized as a key factor in the late 1980s by David Bearman and others, but still as Rackley, Allen, and the Dacre, Cannidine and Pilling Committee show, it is overlooked to this day. It is fair to say that there is broad consensus on our inability to manage born digital materials and of our need to do so. In this context, humanists need to know that the cultural heritage sector is developing the capacity in terms of services, human resources, user support, and technological facilities and policies to manage the born digital as well as the products of their investment in digitization.

Another key issue confronting humanists in the domain of archives is what exactly archives and records managers consider a record and a transaction to be. This has implications for the documentation that these institutions will select and acquire, and eventually make accessible to such users as humanities scholars. We need new research into the nature of cultural heritage evidence and the concept of recordness and what we are trying to achieve in construction of holdings of cultural heritage materials. These challenges are acute in the domain of selection and appraisal. Digital humanists can benefit from a wider array of digital materials, and it would be useful to conduct a gap analysis to identify the kinds of digital materials being created in contemporary society which are not subject to selection and acquisition by heritage institutions which should be.

Cultural institutions need to consider how they negotiate relationships with information businesses as these agreements affect the possible research that humanists can conduct. Cultural heritage institutions engaging in the supply of digital information to aggregators should encourage them to provide end users with documentation as to where that information originated. The case of Ancestry.com is, perhaps, alarming in this regard. Ancestry is basically an aggregator either through digitization or through acquiring copies of digital records and packaging this content to facilitate its use by subscribers. Ancestry users often report, as is evident in several Ancestry TV advertisements currently (2016) accessible on YouTube, that the records of their family they found 'they found on Ancestry'. As I have noted elsewhere, this puts traditional cultural heritage institutions at risk because users of cultural heritage content cease to recognize the role of publically funded

cultural heritage institutions in the care and provision of cultural heritage records (Ross, 2012).

In reality, however, this is a minor issue in comparison to the very significant problem of the incremental privatization of the cultural heritage record. Let me share a recent personal experience to illustrate this point. Studies of ideas, theories and methods benefit from an understanding of the context of their derivation and a knowledge of the individuals involved in making them happen or originating them. Recently, in a preparation for a paper on the origins and scope of the concept of literary warrant in the context of information, the challenges of the emerging information landscape became abundantly clear. From an information perspective, the application of the concept of evident literary warrant to the information domain originated with Edward Wyndham Hulme's (1859–1951) innovative work at the British Patent Office (e.g. Rodriguez, 1984). In investigating literary warrant, understanding who Hulme was appeared to have merit. In the process of attempting to locate his papers and information about him, I sought Public Records about Hulme, including birth, marriage and death records as well as census records. Birth, marriage and death records are available, and he appears in UK census records from 1861 until 1911. This is wonderful news – or so you might think. However, in reality available records means obtainable in analogue form at the Public Record Office in the UK. But you might ask: 'Surely these essential historical records have been digitized and can be found online?' Yes, you would be absolutely correct. They are online, but they are online behind 'paywalls' – that is, cultural heritage institutions have permitted private organizations such as FindMyPast.co.uk and Ancestry.com to digitize (or partnered with them to digitize) and to make the digital instances of the records available for a fee. So although there is much public information about Hulme accessible online, it is locked up in the context of new digital representations created by the very process of digitization, which was intended to improve access and use.

Hulme is just an example of the dangers of the separation of the digital from the physical. The digital object is a rendering of the analogue, but it is different, and it is essential that cultural heritage institutions recognize this. This means that it can be subject to different terms of access and use, and can be 'performed' outside of the environment requiring the application of public access principles. So, yes, as a researcher I can discover that there are records that I could see in analogue form if I were to visit The National Archives in the UK, but that are more conveniently accessible in digital form. They are closed to me unless I subscribe to one of the information aggregators. What is true for these records is also true for audio, moving image, data and a rich array of other materials. The digitized versions are new 'legal entities' and have a raft of new rights associated with them, some copyrights, some moral rights, and many more.

As scholars, we need our cultural heritage institutions to resist the paradigm of using the digital rendition of records as a mechanism to generate income. The digital form should be a manifestation of the record, which enables the record office or archive to reach beyond its physical place to make its content transparent to the diaspora of users rather than being perceived as an income-generating fountain. The paywall culture is nothing new, and researchers outside academic institutions without the capacity to purchase or lease digital content struggle for access to resources. Even as a community, we fight the battle for open access to the raw materials of scholarship and its outputs.

In spite of these institutional and sectoral cultural changes, humanists need the cultural heritage to digitize more of their holdings. However, we need more than images: we need metadata-rich and tagged content that can be mined using natural language processing (NLP) methods and machine learning. More than fifteen years ago, representatives of the EU Member States worked to develop strategies and policies to promote the development of digital content to enable Europe to benefit from the eContent age (Ross, 2001). Their work was in an effort to address issues of fragmentation in digitization activities and enable synergistic digitization, to improve interoperability, discoverability and visibility of the outputs of digitization, to enhance online access to content, and to bring the cultural heritage resources more centrally into the activities of European Creative Industries. This resulted in the development of the Lund Principles (European Commission, 2001). These were intended to produce more co-ordinated approaches to digitization among the Member States, to generate a change in cultural behaviour that would support the digitization of 'Europe's unique and significant wealth locked in its cultural and scientific heritage', to foster consistency in digitization standards and processes from selection to capture to documentation to presentation of content, to ensure that the richness of Europe's cultural diversity was reflected in the digital content created, and to enable researchers, creative industries and the general public to both gain enjoyment and economic reward from improved access (ibid.) The European Commission has continued to convene meetings of representatives of the Member States to reflect on digitization.

Much progress has been made over the past fifteen years in the digital representation of the cultural heritage, and in the creation, collection and management of digital data. When we look at our cultural heritage institutions, it is clear that there is much to be done, as only a small percentage of holdings are accessible online. The regional and institutional nature of most digitization projects remains evident, with few projects bringing together institutions across different sectors and across national borders—Europeana being a notable and laudable exception. Cherry picking of key elements of collections for digitization remains common practice, and humanists are not actively involved in the process of making strategic

decisions as to which content to represent digitally and in what time scale. While overall digitization practice has advanced significantly, there are still challenges related to standards, variations in technologies used and differences in approaches to digital imaging and other forms of representation across institutions in the same country as well as across national boundaries. One lacunae that could be addressed is the lack of an inventory, of which materials are amenable to digitization across EU institutions and strategic planning on regional and cross-sectoral levels. To reflect on this, the items in the inventory might then be digitized in a coherent and structured manner. Even in active digitization programs, poor understandings of workflows and inadequate assessment of user needs and expectations detrimentally influence the outcome of digitization activities (Ross, 2004). The value of digital materials to humanities is greatly influenced by the ways in which cultural heritage institutions manage its conversion and how they document the processes of digitization and digital entity documentation (e.g. syntactical, semantic and pragmatic metadata). Quality assurance, which we recognized as a key issue in the *NINCH Guide* (Ross *et al.*, 2002), is central to all digitization initiatives and fundamental if scholars are to depend upon digital content as a resource.

Aligned with this issue is the question of authenticity and its relationship to digitization. What makes a digital representation of an analogue object authentic? Most users adopt the stance that if the material was digitized by a cultural heritage institution they can assume that it is authentic. This notion of the 'presumption of authenticity' has permeated our perception of digital content. Many humanists adopt this even when they know little about the nature of the analogue collection that was digitized or the standards or workflows that were used to carry out the digitization. For instance, some cultural heritage institutions 'touch up' damaged photographs during pre-digitization conservation or during digitization, which can, as the JISC JIDI programme of work showed in the late 1990s, influence the usefulness and evidential value of digital materials to researchers and even the general public. The cultural heritage institutions that are providing access to digitized content need to provide rich evidence as to the processes and practices involved in digitization, as this influences the ways in which humanists and others can use the products of the process. A level of transparency of practice and policy is essential if cultural heritage institutions are to support humanities research adequately.

Of course, one of the great conundrums facing the cultural heritage is whether, if you digitize materials, researchers will be drawn to use them. It depends. As I found in a 2003 study for the National Library of New Zealand,

> those projects that have successfully carried out digitisation to enable research using particular materials tend to have done so with a particular research agenda in mind. It is difficult to point to digitisation activities where material digitised to enable research by one scholar has

> also facilitated research by others. This anomaly may well arise because the little scholarship that has been done using digital resources has been done using those that were created to enable particular researchers to answer specific research questions and there has been little opportunity for the resources to be reused by other researchers. The Cultural Heritage may wish to digitise material in anticipation of scholarship that might be conducted using the resources created, but the effectiveness of this strategy is difficult to predict as the research agendas of disciplines are notoriously difficult to anticipate far enough in advance to ensure that investment can be appropriately targeted for digitally reformatting collections.
>
> (Ross, 2003b)

However, once we reach a critical mass this might change. There is evidence that projects that aggregate information and provide the framework for scholarship are worthy of creating: lexicons, image catalogues for art historical research and large datasets derived from the mining of historical documents. These, though, are more often created by humanists and used by the cultural sector than vice versa. This is all part of the symbiotic relationship, where access to heritage materials enables scholars to produce resources that assist heritage professionals to better understand and curate the materials in their care.

One idea that has intrigued me since the 1980s has been whether or not the use of new ways of representing information and reasoning with it in the humanities would change the publishing chain and transform our relationship with the cultural heritage sector itself (Ross, 1990). Indeed, at the very beginning of my career I was looking at the possibilities that a proliferation of micro knowledge-bases, similar to the one I had made to instantiate the knowledge I had produced from research into Anglo-Saxon pins (Ross, 1992) would proliferate and change the way that archaeological knowledge, a major cultural resource, was packaged, disseminated, used and evaluated. This naive optimism that knowledge could be represented as functional processes existing in active code – what the pioneer of the application of Artificial Intelligence in archaeology, Jean-Claude Gardin, had earlier perceived as a mechanism to bring rigour, consistency and functionality to archaeological knowledge – never came to fruition. Bad timing? Perhaps. However, as I explained in a discussion about the challenges impeding the widespread adoption of the Semantic Web (2003a), the technology needed improving and it required a critical mass of researchers or enablers to become engaged in the process to build the knowledge-bases and semantic interconnections. This just did not happen in the case of knowledge-based systems.

Advances in machine learning, facilitated by new methods and the remarkable developments that have taken place in computational capacity and nature, is changing all that. Here, digital humanities need multiple

cultural heritage institutions to allow us to work with vast swaths of their digital collections to experiment with the application of these methods. In many sectors, the pattern recognition capacity of machine learning, and in particular deep learning technologies, is radically altering our ways of working. Certainly, the new focus on linked-data and open access will change the publishing chain in ways that knowledge-based systems might have but never did. As humanists, we need to know that cultural heritage institutions are harnessing new information representation and processing methods.

There are numerous other challenges. We need cultural heritage institutions to document their holdings which are not yet described or indexed in ways that make them discoverable even if we can only examine them as analogue entities (see above). Although in many European countries there has been a push to document the holdings of cultural heritage institutions, there are many areas, such as in the case of music manuscripts or photographic collections, where even basic documentation has not yet been created.

We need the cultural heritage sector to work with us to create evolving documentation standards and information systems. For example, major institutions with material collections from diverse cultural contexts which often reflect different ways of knowing and even reasoning about the world, use predominantly Western architected information systems to encode data about the objects, thereby imposing colonial knowledge structures on cultures and their records in a post-colonial world.

Although the European Commission and its Member States continue to host initiatives to push forward digitization, public funding for digitization seems scarcer every year, and it has certainly dried up for research into new approaches to digitization and digital curation. We need the cultural heritage sector to work with us to promote greater awareness of governments in the value of funding these activities.

Often in the humanities community we hear it said that the difference between humanists and scientists is that scientists work with structured data and humanists work with unstructured data, commonly known as documents. This is just not true. Documents are highly structured data. One of the pre-eminent strengths of humanities scholars, which is acutely present in the work of digital humanists, is our ability to identify these structures, to describe them, to interpret them, to interrelate them and to create meaning from them whether the document in question is a text, a painting, a broach or a building. Even in the pre-digital age, humanists reshaped the way we handled these different classes of documents and the stories we could tell using them. We need only think of Foucault (*The Archaeology of Knowledge*, 1972), Gombrich (*Art and Illusion*, 1960), Speake (*Anglo-Saxon Animal Art and its Germanic Background*, 1980), or Panofsky (e.g. *Gothic Architecture and Scholasticism*, 1951).

We need cultural heritage institutions to embrace digital humanists as their research partners and to encourage us to develop new ways to apply digital methods and tools to study analogue objects, whether this

occurs by pioneering new imaging techniques, including multispectral imaging, adopting methods for analysing audio from other sectors and studies of the applications of 3D imaging and printing to enable researchers to study such aspects of objects as weight, volume, texture and colour. Current approaches to digital representations separate the scholar from the physicality of the analogue. It is difficult from a digital image to gain a sense of the size, weight or texture of the analogue original of which it is a digital representation. This loss of physicality results in a loss of meaning. Continued research by humanists and the cultural sector into methods of digitization and representation is essential if we are to enable scholars and students to explore the multimodality and multivocality of the cultural heritage. This research needs to be accompanied by more investigations of what it is to be authentic or an authentic rendition in the digital age. As recognized earlier in this chapter, many users presume that a digital rendition of an analogue entity is an authentic representation even when in its production aspects of its physicality have not been adequately captured (e.g. texture and weight). Moreover, researchers and the public are often unaware of the decisions that were taken during the process of digitization that have altered the integrity, and even the reliability of the digital representation as evidence.

Conclusion

In conclusion, let's take a stab at what will be *Digital humanities research needs from cultural heritage looking forward to 2025*. In 1993, the report of the British Library Research and Development and British Library Review (Ross and Feeney, 1993) of Humanities Informatics charted the need for infrastructure, including networks and even computers, technological advances in the form of new tools, support with skills development and data repositories. We recognized the slow incremental, project-by- project basis, a basis on which content was being represented in digital form. This is still the case, although infrastructure, throughput and capacity have become greatly enhanced, and continued investment in education, facilities, content and support services is essential. In many ways, the services that have become available in the last nearly two and a half decades has had a tremendous influence on scholarly possibilities, but scholarship remains a slow process.

We would hope that cultural heritage institutions do not move to license content for financial gain, but put greater focus on public access and openness. There are major issues with intellectual property rights of all kinds in digital materials, and it is essential that the cultural heritage sector and humanities scholars work together to tackle these issues. This collaboration is fundamental if we wish to harness the rich creativity the digital rendering of our analogue heritage can offer to the remix, bricolage and lego-constructor cultures in which we live.

One of the single greatest challenges we will face is that of being confident in the authenticity of the digital materials, whether these digital objects are documents, data or both. In this, the tendency has been for memory institutions to build institutional repositories to manage these entities and, as noted above, to act as the guarantor of the authenticity of the materials in their care and maintenance – reminiscent of the fox guarding the hen house. As our world becomes increasingly interconnected, it is clear that there may be new and better ways of managing digital objects, and this includes post-custodial approaches. For many years now, I have considered that the future of digital curation lay in putting intelligence at the level of the digital object and creating 'safe harbour seeking digital objects which managed their own integrity and indeed existence' (Ross, 2007) but I could not conceive of how such a concept of autonomous digital heritage might be realized. In the last three years, it is increasingly evident that there are emerging technologies that can be used to create 'webs of trust' that may make it feasible to look beyond the custodial view of digital cultural heritage artefacts. One such technology is 'blockchain', which underpins bitcoin (Ross, forthcoming). This will dramatically change the way in which the cultural heritage sector can contribute to facilitating access to digital materials.

We have long seen that collaboration brings numerous benefits to all participants. For me, the question cannot be about what digital humanities researchers need from the cultural heritage sector, but it must rather be about how the humanities and the cultural heritage sectors can forge a shared vision of the ways in which together we can foster mutually beneficial developments. These initiatives will not just be about the scholarship of discovery, but they will also be about the scholarship of digitization, documentation, datafication and curation, whether it is to construct a better understanding of our intangible heritage to ensure that more of it is documented and retained for the future, of our archaeological heritage to enable richer documentation of sites themselves, or of our natural heritage to record how it has been shaped to help us understand how these interventions have shaped us. So, not only do we need collaboration from the cultural heritage sector in the area of the selection of materials for digitization and curation, but we need to forge research networks designed to integrate research by digital humanists and the cultural heritage. Collaboration on advocacy is essential if we are to ensure that such integration takes place, at times when the cultural heritage is increasingly in competition with other societal concerns that appear to have greater social urgencies such as the dire state of healthcare and education. So, to convince cultural heritage institutions, they need to give us content and help to understand it, and we need to give them a value proposition. What that value proposition is, is an open question. However, what I do know is that it will be about how in collaboration with cultural heritage institutions, digital humanists

can create social value and viable social outcomes, and not merely outputs, from content preserved in these institutions.

References

Allan, A., 2015, Government Digital Records and Archives Review, www.gov.uk/government/publications/government-digital-records-and-archives-review-by-sir-alex-allan (completed 2014).

Dacre, P., Pilling, J. and Cannadine, D., 2009, Review of the 30 Year Rule, Independent Report Prepared for the Office of the Prime Minister of the United Kingdom.

European Commission, 2001, The Lund Principles: Conclusions of Experts Meeting, European Content in Global Networks Coordination Mechanisms for Digitisation Programmes, Lund, Sweden, 4 April, https://cordis.europa.eu/pub/ist/docs/digicult/lund_principles-en.pdf.

Foucault, M., 1972, *The Archaeology of Knowledge* (trans. A.M. Sheridan Smith), London: Tavistock Publications, 1972.

Gombrich, E. H., 1960, *Art and Illusion: A Study in the Psychology of Pictorial Representation*, London: Phaidon Press.

Hockey, S., and Ross, S., 2008, Conclusion, in *The AHRC ICT Methods Network – Final Report*, pp. 59–74, www.methodsnetwork.ac.uk/publications/final-report.html.

MiBACT, 2014, Digital Cultural Heritage and Tourism: Recommendations for Cultural Institutions, Italian Presidency of the Council of Europe (Ministero dei Beni e delle Attività Culturali e del Turismo – MiBACT).

Panofsky, E. 1951, *Gothic Architecture and Scholasticism*. Latrobe, PA: Archabbey Press.

Rackley, A., 2016, *Archiving the Games: Collecting, Storing and Disseminating the London 2012 Knowledge Legacy*. Doctoral thesis, University of Central Lancashire, http://clok.uclan.ac.uk/16539.

Rodriguez, R. D., 1984, Hulme's Concept of Literary Warrant, *Cataloging & Classification Quarterly*, 5(1): 17–26, DOI: 10.1300/J104v05n01_02.

Ross, S., 1990, Expert Systems for Databases in the Historical Sciences: A Case Study from Archaeology, in Cacaly, S. and Losfeld, G. (eds), *Sciences Historiques, Sciences du Passe et Nouvelles Technologies d'Information: bilan et évaluation*. actes du Congres international de Lille (March 16–18, 1989) (Villeneuve d'Ascq : CREDO, Université Charles de Gaulle Lille III, 1990), pp. 179–186.

Ross, S., 1992, *Dress Pins from Anglo-Saxon England: Their Production and Typo-chronological Development*, University of Oxford at: http://ora.ox.ac.uk/objects/uuid:3976b772-fccd-41fe-b8c7-f4ae08ac0295.

Ross, S. (rapporteur), 2001, Report of an Expert Meeting on European Content in Global Networks, Coordination mechanisms for digitisation programmes, Lund, SW, 4 April 2001, https://cordis.europa.eu/pub/ist/docs/digicult/lund-report_en.pdf.

Ross, S., 2003a, Position Paper: Towards a Semantic Web for Heritage Resources. In: Geser, G. (ed.) *DigiCULT: Towards a Semantic Web for Heritage Resources*. Series: Thematic issue 3, DigiCULT: Salzburg, AT & European Commission, pp. 7–11.

Ross, S., 2003b, *Digital Library Development Review*, National Library of New Zealand, Wellington, NZ, http://ndhadeliver.natlib.govt.nz/content-aggregator/getIEs?system=ilsdb&id=1218472.

Ross, S., 2004, Reflections on the Impact of the Lund Principles on European Approaches to Digitisation. In: van der Laan, M.J. (ed.) *Strategies for a European Area of Digital Cultural Resources: Towards a Continuum of Digital Heritage*. Ministry of Education, Culture and Science: The Netherlands, pp. 88–98, September 15–16, Den Haag, the Netherlands.

Ross, S., 2006, Approaching Digital Preservation Holistically, *Information Management and Preservation*, Oxford: Chandos Press, pp.115–153.

Ross, S., 2007, *Exploring and Charting the Digital Preservation Research Landscape*, iPRES – 2007 International Conference on Preservation of Digital Objects, hosted at the National Science Library, Chinese Academy of Sciences, 11 October, Beijing, http://ipres.las.ac.cn/pdf/seanus%20ross-ipres2007_keynote_ross_slides.pdf.

Ross, S., 2012, *The Economics of Digital Creation, Curation, and Preservation,* at UNESCO, The Memory of the World in the Digital Age: Digitization and Preservation: An International Conference on Permanent Access to Digital Documentary Heritage, September 26–28 (26 September), Vancouver, BC, www.unesco.org/archives/multimedia/?pg=33&s=films_details&id=2777.

Ross, S. and Feeney, M., 1993, *Humanities Information Review Panel: Information Technology in Humanities Scholarship: British Achievements, Prospects, and Barriers*, London: The British Library Research & Development Department and The British Academy.

Ross, S., Anderson, I., Duffy, C., Economou, M., Gow, A., McKinney, P., Sharp, R. and The NINCH Working Group on Best Practices, 2002, *Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, Washington, DC: NINCH.

Schreibman, S., Siemens, R. and Unsworth, J. 2016, *A New Companion to Digital Humanities* (2nd edn), Oxford: Wiley-Blackwell.

Shenton, H., 2012, Virtual Reunification, Virtual Preservation and Enhanced Conservation, Alexandria, 21.2, 33–45, DOI: 10.7227/ALX.21.2.4.

Speake, G., 1980, *Anglo-Saxon Animal Art and its Germanic Background*. Oxford: Clarendon Press.

Tinline, P., 2016, Too Good to be Forgotten – Why Institutional Memory Matters, BBC, Radio 4 (21 March), www.bbc.com/news/business-35821782.

Index

Page numbers in *italics* indicate figures or images.