

42

Science Policy Briefing • September 2011

# Research Infrastructures in the Digital Humanities

---

## Contents

- 2 • Foreword
- 3 • Introduction
- 4 • Definitions, Taxonomies and Typologies of RIs
- 8 • Bridging Physical RIs in the Humanities with Digital RIs
- 15 • Researchers' Input and Engagement in Producing RIs
- 19 • Digital Research in the Humanities: who is Responsible for RIs?
- 21 • Preservation and Sustainability
- 24 • Evaluation of Digital Research and its Outputs
- 26 • Communities of Practice
- 32 • Cultural and Linguistic Variety – Transnational RIs
- 35 • Education and Training
- 39 • Conclusions: Priorities for Policy and Research
- 42 • References
- 42 • Index of Case Studies
- 42 • List of Authors and Contributors

## Foreword

This peer reviewed document reflects on the centrality of Research Infrastructures (RIs) to the Humanities. It argues that without RIs such as archives, libraries, academies, museums and galleries (and the sources that they identify, order, preserve and make accessible) significant strands of Humanities research would not be possible. After proposing a wide-ranging definition of digital RIs – with the aim of reflecting on the meaning of infrastructure in the Humanities rather than on those parts common to other domains of science – it attempts to relate physical RIs to digital ones. By drawing on a number of case studies – chosen to showcase the variety of research around existing or emerging infrastructures – it demonstrates that digital RIs offer Humanities scholars new and productive ways to explore old questions and develop new ones. Indeed, it is argued that making our cultural heritage accessible in digital form plus its sensitive interlinking with other resources opens a new frontier for Humanities research for addressing ‘grand challenges’ in the Humanities themselves and at the interface with other research domains. These include not only research-based challenges, such as data-driven Humanities, but also institutional and social issues, such as strengthening higher education programmes as well as the recognition of the inherently process character of digital research and the implications of this for evaluation and promotion.

While numerous sophisticated RIs that can inform and further Humanities RIs already exist in other domains of science, ultimately it is also necessary for Humanities scholars to build and have access to ‘fit for purpose’ Humanities RIs, given the nature of their data sets, research methods and working practices. As van Peursen has reflected:

“[...] 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.”<sup>1</sup>

1. Peursen, Wido T., van . “Editorial.” *Text Comparison and Digital Creativity*. Ed. Weel, Adriaan H., van der, Ernst Thoutenhoofd, and Wido T. van Peursen. Leiden: Brill, 2010. 11. Print.



Cover

British Museum glass ceiling © MonkeyRiverTown

The publication of the *European Strategy Forum on Research Infrastructures (ESFRI) roadmap*<sup>2</sup> has made clear the importance of dedicated Humanities RIs. However, only two Humanities RIs have been funded through this initiative.

This report positions itself in terms of priorities and future research directions for a common strategy on RIs in the Humanities at the European level.<sup>3</sup> The bridging of physical and digital RIs presents opportunities and challenges with implications that are discussed under the following rubrics: researcher input and engagement in making RIs; preservation and sustainability; the evaluation of digital research and its outputs; communities of practice; cultural and linguistic variety (transnational RIs); education and training.

Digital infrastructures are developing rapidly but unevenly, and there is an urgent need for coordination, standardisation and sharing of experience to prevent unnecessary duplication and the atomisation of good initiatives. This Science Policy Briefing identifies already

2. The latest update of the ESFRI roadmap at the time of writing is available at <[http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy\\_report\\_and\\_roadmap.pdf](http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf)>.

3. Note that national efforts in this direction have already been made or are under way. See for instance *Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften*. Wissenschaftsrat, January 2011. Web. (accessed 29/07/2011).

pressing and future needs by focusing on current developments and initiatives with the aim of bringing to the fore intellectual challenges rather than mainly technical or funding issues.

This document is aimed at researchers and information professionals (including librarians, archivists, etc.) as well as the institutions which make decisions of crucial importance to them, such as funding bodies, those responsible for management and administration of research organisations and RIs, selection and promotion committees. It is also addressed to faculty and curriculum accreditation committees responsible for developing courses in the area of RIs.

Work on this report began in early 2009 following the decision of the ESF Standing Committee for the Humanities (SCH) to make RIs in the Humanities one of its top priorities. Having identified RIs as an area of strategic importance, the SCH set up an SCH Expert Group on RIs which has for the last two years been deeply involved in the creation of this paper.

The report you have before you builds on already existing documents pertaining to RIs in the Humanities and especially on the report compiled by an earlier SCH 'Vision Group on RIs' which met during 2007.<sup>4</sup> In addition to a literature review this report also incorporates views put forward by members of different communities involved in Humanities RIs, views expressed at the *Strategic Workshop on Research Communities and Research Infrastructures in the Humanities* organised by the ESF Standing Committee for the Humanities and the above mentioned RI Expert Group in October 2010 in Strasbourg. Many of the case studies featured below are based on papers presented at this workshop. Furthermore, the report was reviewed by some of the leading experts in the field and, where possible and appropriate, their comments have been incorporated.

**Professor Marja Makarow**  
ESF Chief Executive

**Professor Milena Žic Fuchs**  
SCH Chair

**Professor Claudine Moulin**  
Chair of SCH RI Expert Group

4. The work of this group was linked to the SCH contribution to the production of the *Report on the EC-ESF-EUROHORCS survey on research infrastructures* (see <<http://www.esf.org/activities/science-policy/research-infrastructure/the-2nd-european-survey-of-ri.html>>), produced in 2006, and currently available at <[http://ec.europa.eu/research/infrastructures/pdf/survey-report-july-2007\\_en.pdf](http://ec.europa.eu/research/infrastructures/pdf/survey-report-july-2007_en.pdf)>.

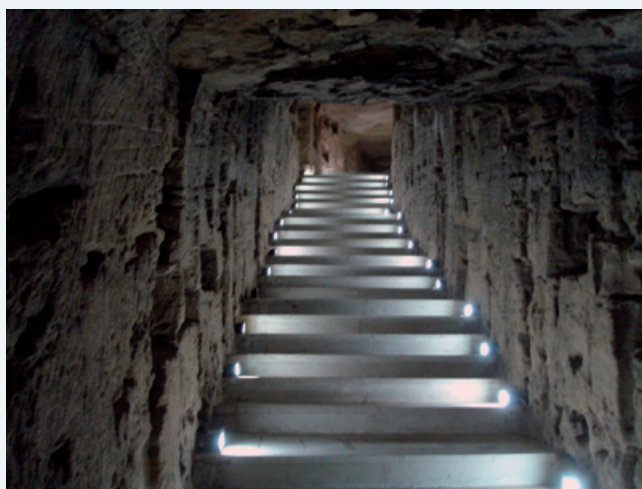
## Introduction

### On the origins of Research Infrastructures (RIs) in the Humanities

To begin with the historical dimension, it is worth noting that it was in the field of Humanities that the idea of an RI was first born. It is not possible here to write a cultural history of RIs but it is most insightful to note that as early as the 3<sup>rd</sup> century B.C., the imperative to collect, organise and conserve the knowledge acquired in the service of the advancement of knowledge gave birth to the first ever 'Information Centre' in the form of the *Mouseion*, a cultural centre, university and library founded in Alexandria under the successors of Alexander the Great. The positive consequences of this ambitious venture soon became obvious, as the production of various RIs in the form of *Grammars* and *Lexica* proliferated.

This early success of RIs in the Humanities was not lost in the period that followed the end of the Roman Empire; rather, it inspired a range of activities which were to foster an emerging system of knowledge and emphasise the importance of collection, organisation and conservation in serving subsequent scholarly development, not exclusively in Humanities. Examples of such activities include, *inter alia*, the creation of medieval libraries and, later on, the formation of art collections.

These early 'databases' provided material for subsequent phases of RIs in the Humanities. To name but a few examples, the advancement of editorial ventures as well as the intentional and systematic collection



**Figure 1.** Ruins of the *Serapeum*, temple dedicated to the syncretic Hellenistic-Egyptian god Serapis, and where the daughter library of the Library of Alexandria, the *Mouseion*, was located. Alexandria, Egypt.

© Daryl Manning

of knowledge on a glossographical and encyclopaedic level opened a line of tradition from Medieval to Modern times. Already towards the end of the early-modern period, we note, for example, the major projects of academies such as the publication of large-scale encyclopaedia as well as museum and collection catalogues; the formation of scholarly disciplines, fundamental advancements in classification and taxonomy and the diffusion of European journals. Some of the more ambitious and, by necessity, long-term ventures were initiated by prestigious scholarly academies. They undertook the systematic categorisation of objects and texts and their dissemination in formal – in many cases still ongoing – multivolume critical editions or serial corpora. These in turn often sparked the development of new research tools, such as extensive indexes, bibliographies, biographical dictionaries, etc.<sup>5</sup> Many archival institutions were also reorganised in a way that allowed historians to study the records in their original groupings; this reorganisation enhanced the general usefulness of the archival records, making them into what we today would call RIs.

Despite the advancements of recent decades, some Humanities researchers have been reluctant to realise and acknowledge the advantages that the application of computing technology to Humanities can bring.<sup>6</sup> Nevertheless, it is increasingly common for researchers to explore the many links that exist between our cultural and material heritage and the development of RI tools, such as comprehensive databases (building on the foundations of ‘traditional’ tools), research often being motivated by the need for secure preservation of endangered data through digitisation.<sup>7</sup> The discipline of Classical Studies, in particular, has been among the earliest adopters of such digital approaches. In order to support the continuation of such vital work, and to attract a new generation of Humanities researchers to

---

5. Examples of ventures which served and still serve a worldwide scientific community include the *Corpus Vasorum Antiquorum* (CVA), the *Corpus of Latin Inscriptions* (CIL), the *Inscriptiones Graecae* (IG) corpus, and in recent years the *Lexicon Iconographicum Mythologiae Classicae* (LIMC), to mention but a few.

6. Notable exceptions include the early-adopter field of computational linguistics (today corpus linguistics and language technologies). This field was built around the use of computers in the processing of natural language and speech data. In this respect it may be considered to be one of the founding layers of what is today understood by the broad term Research Infrastructures in the Humanities. Since language is ubiquitous, it serves also in different Humanities disciplines (and wider) as the basic conveyer of research objects. In this respect, having language data in digital form can be considered one of the first steps towards the development of RIs in the Humanities.

7. See, for example, Crane, Gregory and Melissa Terras (Eds) *Changing the Center of Gravity: Transforming Classical Studies through Cyberinfrastructure*. Spec. issue of *Digital Humanities Quarterly* 3.1 (Winter 2009). Web (accessed 14/07/2011).

carry out such work, a number of ongoing problems need to be addressed. These include the development of a language of common understanding between the computing and non-computing Humanities, the lack of appropriate funding models for such projects and the institutional and professional transformations that are needed to underpin such work.

## Definitions, Taxonomies and Typologies of RIs

Humanities researchers have long been familiar with Research Infrastructures (RIs) and the objects that populate them. Archives, museums, galleries and libraries have always housed collections of physical objects such as archaeological fragments; paintings or sculptures; inscriptions or manuscripts; books and journals, etc. The digital age is compelling us to introduce such physical collections onto the digital plane by digitisation and/or to construct new collections of digital objects as subjects of research in Humanities today. So, with the coming of the digital age, Humanities researchers are extending their views on the nature of research objects, whether physical/analogue or digital; they are reinterpreting and redefining traditional or physical repositories and collections of research objects; and they are reassessing the very nature and definition of RIs. Therefore, the overall definition of RIs in Humanities needs to encompass both physical and digital RIs in order to facilitate new research in established subject areas, and lead to the creation of new subject areas.<sup>8</sup>

Many definitions of RIs have been formulated in recent years; regarding Humanities, it should be stressed that there are special dynamics and aspects that must be considered while doing so. For some researchers, an infrastructure is the technical and operational framework that allows them to collaborate and share data and results; for some it is the content to which access is offered rather than the facilities around it; and for some it is both.

Whereas the *European Strategy Forum on Research Infrastructures* (ESFRI) has focused on a broad definitional approach that spans the disciplines, the present paper would like to concur with, and extend – so as to

---

8. On the relation between physical collections and digital materials see Lynch, Clifford A. “Special Collections at the Cusp of the Digital Age: A Credo.” *Research Library Issues* 267 (December 2009): 3-9. Web. 4 August 2011.

adapt to the pan-European context – the definition of “cyberinfrastructure”<sup>9</sup> in *Our Cultural Commonwealth* report.<sup>10</sup> This inclusive definition evokes both large and small scale, as well as international and national infrastructures:

“the term cyberinfrastructure is meant to denote the layer of information, expertise, standards, policies, tools, and services that are shared broadly across communities of inquiry but developed for specific scholarly purposes: cyberinfrastructure is something more specific than the network itself, but it is something more general than a tool or a resource developed for a particular project, a range of projects, or, even more broadly, for a particular discipline. So, for example, digital history collections and the collaborative environments in which to explore and analyze them from multiple disciplinary perspectives might be considered cyberinfrastructure, whereas fiber-optic cables and storage area networks or basic communication protocols would fall below the line for cyberinfrastructure”. (p. 8)

Colleagues in Sciences and Technology have made a profound contribution to the development of RIs. Disciplines such as Computer Science and Engineering have designed and implemented many fundamental technologies, while also building substantial bodies of knowledge about pertinent social, legal and institutional issues. Key aspects of this are transferrable to Humanities; nevertheless, as the definition above intimates, Humanities researchers have particular requirements of RIs. For example, in the hard sciences, datasets tend to be, in the first instance, generated rather than collected and are homogenous in nature, i.e., numeric. In Humanities, artefacts of, *inter alia*, human culture, expression, interaction and imagination form our ‘datasets’; data tends to be collected and heterogeneous in content and format.<sup>11</sup> Therefore, the considerable advances in RIs that have been made in the Sciences cannot simply be transferred to Humanities ‘out of the box’.

9. In this paper the term “Research Infrastructure” is adopted instead.

10. *Our Cultural Commonwealth: the report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: American Council of Learned Societies, 2006. Web (accessed 14/07/2011).

11. “While in Sciences the raw data constituting the hinterland of research are typically produced and kept by the same people who write the publications, authors in the Humanities are as a rule not those who collect and preserve cultural heritage or provide access to it”, in Hyman, Malcolm, and Jürgen Renn, ‘From Research Challenges of the Humanities to the Epistemic Web (Web 3.0)’ NSF/JISC Repositories Workshop, Max Planck Institute for the History of Science, Berlin, 6 April 2007. Web (accessed 14/07/2011).

Another relevant definition is the following, included in the *Community legal framework for a European Research Infrastructure Consortium* (ERIC) regulation:

“research infrastructure” means facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling Information and Communications Technology-based infrastructures such as Grid computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research. Such infrastructures may be “single-sited” or “distributed” (an organised network of resources).<sup>12</sup>

And the definition proposed by the European Commission Framework Programme 7 (Contract # 262159) project on *Mapping of the European Research Infrastructure Landscape* (MERIL)<sup>13</sup>:

“A European Research Infrastructure is a facility or (virtual) platform that provides the scientific community with resources and services to conduct top-level research in their respective fields. These research infrastructures can be single-sited or distributed or an e-infrastructure, and can be part of a national or international network of facilities, or of interconnected scientific instrument networks.

The infrastructure should:

- offer top quality scientific and technological performance, that should be recognised as being of European relevance
- offer access to scientific users from Europe and beyond through a transparent selection process on the basis of excellence
- have stable and effective management”.

These high level definitions clearly show that an RI cannot be defined in an abstract, absolute and immutable way; rather, it is a term that is adapted for and by different disciplines.

Accordingly, we endeavour to set out below a description of the most common forms of RIs found in Humanities. At an overarching level, four primary layers of RIs can be identified:

12. Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC). *Official Journal of the European Union* 52 (8 August 2009): 1. Web (accessed 14/07/2011).

13. For more details see <<http://www.esf.org/meril>>.

- Physical infrastructures: collections of physical objects/installations/vessels/instruments (these may be single-sited or hosted by more than one institution/country);
- Digital data infrastructures: these comprise single-sited or interconnected data repositories, spread over several institutions/countries;
- E-infrastructures: networks and/or computing facilities spread over various institutions and/or countries. This is the technical backbone of a given RI, and examples include GRID computing, cluster computing, cloud computing and the networks that connect them;
- Meta-infrastructures: conglomerates of independent RIs, residing in different institutions/countries with different data formats and data structures (i.e., resulting from different activities) yet connected using compatible metadata formats or processes, thus enabling access to different data archives.

At this macro-level a number of broad categories of offerings can also be identified, these include:

- access to data and physical/analogue objects;
- access to services;
- access to expertise;
- access to laboratory facilities.

Within these broad categories, a typological framework of RIs in the Humanities may be sketched from a multi-layered and multidimensional perspective.

As reflected in the diagram below, where a set of criteria for defining the RI is presented, on the vertical axis a distinction can be made between RIs that offer access to material that could be considered of ‘primary’ nature (original data/documents) and those that provide enriched data and metadata. These distinctions do not indicate that such categories are discrete or opposed (many examples of RIs that combine both categories may be identified). Furthermore, such a differentiation should not be seen as oversimplifying the dichotomy of the status of primary data as ‘raw’ material; nor should it ignore the potential of enriched data and metadata to itself become raw material and form another RI whether dynamically, as in the case of linked data and semantic web technologies,<sup>14</sup> or with regard to the individual objects that populate the RI, for example, in the case of a digital edition.

Remaining at the vertical level, another significant consideration is the distinction between digitised objects

and born-digital objects. The first are digital abstractions or remediations of physical objects. Inevitably, they represent sets of data that are used as abstractions of the original object for research or other purposes. The second are objects that are intended, produced and exist as digital objects only. A third case is the combination (hybrid), where the same object is available at the same time in both media (e.g., a newspaper published traditionally on paper and digitally on a news portal).

At the horizontal level, different categories can be identified, such as subject-related RIs as opposed to institutional RIs. At a further level there exists a range of RIs that create dynamic interconnections, networks and bridges across different types of resources (e.g., thesauri, co-referencing systems, conceptual models, semantic web frameworks, ontologies). While physical RIs in Humanities and Social Sciences are composed of predominately static collections of objects, newly emerging RIs may give the community a different dimension. RIs may also be composed of collections of objects that are dynamically identified and aggregated, thus enabling the researchers to track changes in (almost) real-time if needed. A step even further might include RIs that are not composed of collections but of services (or tools) that allow researchers to track the kind of data of relevance to their research (e.g., ‘web as a corpus’ approaches, where dedicated specialised corpora are being built from a set of seeding terms and/or URLs). In this way an RI can enable research methods that have not been previously available to Humanities.

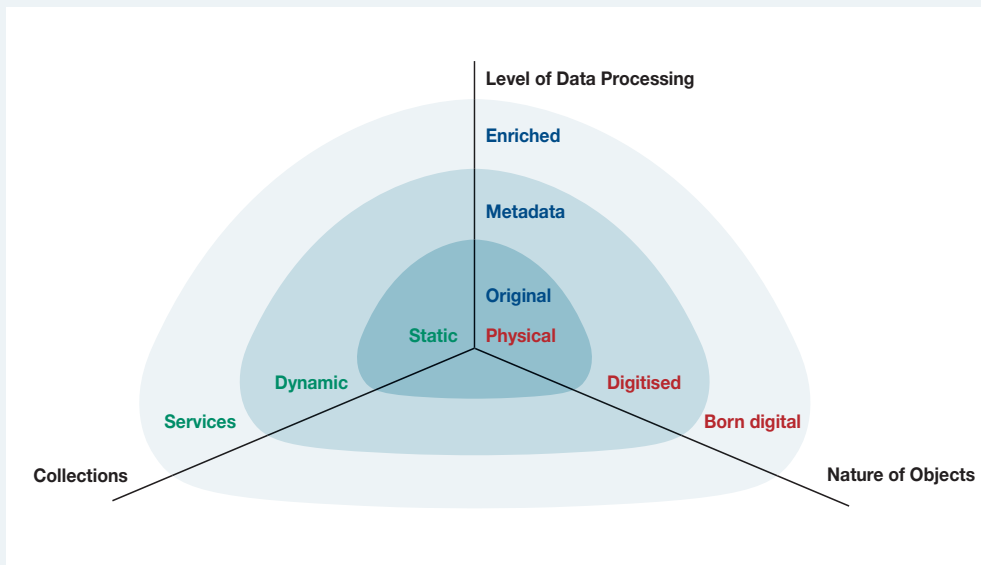
Overarching this open model of the often intersecting characteristics of individual RIs there exists an **ecosystem** of RIs where the following levels of scale or reach of an RI co-exist and even intersect harmonically:

- global or pan-European level;
- national or community-driven level;
- local/institutional level (e.g., competence centres).

It should be noted that some RIs are scaled and targeted to one or more of these categories from the moment of their design and implementation, while some were planned on a smaller scale in the beginning, but by their usage have evolved into larger-scale RIs. The prerequisite for this evolution is (free) Internet accessibility and usage of standards in data repositories and data service.

Within this ecosystem, the corresponding service provider structure caters for the worldwide/European technical infrastructure (e.g., GRID computing) while at the same time encompassing the European/National Competence centres at community level (e.g.,

14. One of the first projects to build a semantic infrastructure for philosophy was the *Discovery Project*: <<http://www.discovery-project.eu/home.html>> (accessed 15/06/2011).



**Figure 2.** A set of concurrent criteria for defining the RI in Humanities. The same representation applies for the local/institutional level, the national/community level, and the pan-European/global level.

DARIAH<sup>15</sup>/ CLARIN<sup>16</sup>). Such an ecosystem of RIs is conceived as being regulated by the interaction of components such as, *inter alia*:

- the quality of data, tools and services, (e.g., scholarly reliability, transparency of standards, and methodologies; the ownership vs the licensing of such data, tools and services together with related issues of Intellectual Property and copyright);
- their dynamic functioning beyond a project-based support;
- and, last but not least, ethical issues.

In order to ensure the sustainability of Humanities RIs a number of key activities require support. These include:

- long-term preservation, including persistent identifiers for digital objects;
- sharing of data, tools and services, through dedicated authentication and authorisation services as well as general (social) networking;
- active collection and development of new data and tools, preventing RIs from becoming obsolete;
- applicable business models, enabling RIs to be sustained and further funded by either their founders or appropriate bodies.

#### CASE STUDY

### MERIL project: Categories of Research Infrastructures relevant for Humanities

The European Commission Framework Programme 7 (Contract # 262159) project MERIL (*Mapping of the European Research Infrastructure Landscape*) started on 1 October 2010. Its main goal is to produce a comprehensive inventory of existing RIs of European importance.

Within this project a set of categories or types of research infrastructure has been proposed. For the purposes of this paper, listed below are those RIs which are relevant for the Humanities, either in the strict sense of being made by researchers in the Humanities and/or made for researchers in the Humanities, or in a wider sense, being 'transversal', thus being able to serve different disciplines, but also being useful for researchers in the Humanities.

#### Humanities-specific RIs

- **Cognitive Sciences facilities**  
facilities for neurological/psychological research on speech/textual, visual, audio, tactile and olfactory stimuli used in linguistics, phonetics, musicology, art history, anthropology, etc.
- **Research facilities for Cultural Heritage objects**  
facilities to do research in and perform restoration and conservation of cultural heritage
- **Music and instrument collections**  
collections of musical scores, recordings, musical instruments and relevant musicological data
- **Literature and text archives**  
text collections/repositories of literary works, databases of analytical data and metadata

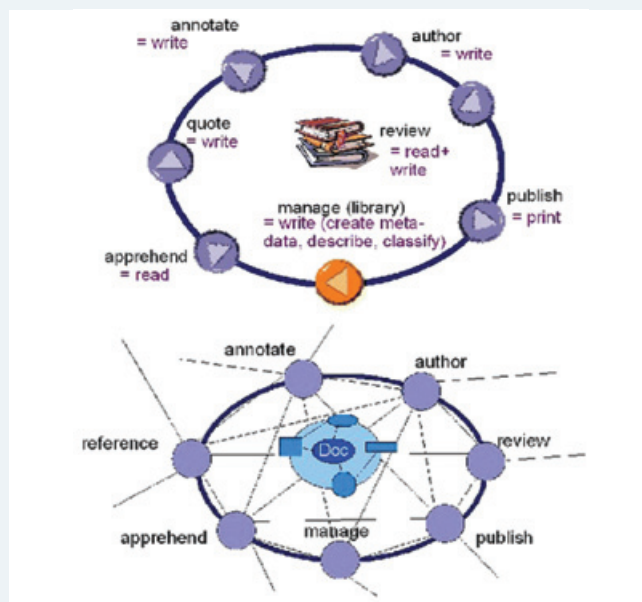
15. See <<http://www.dariah.eu/>> (accessed 14/07/2011).

16. See <<http://www.clarin.eu/>> (accessed 14/07/2011).

- **Language resources, tools and services**  
language resources (corpora, lexica, grammars), tools and services made by language and speech technologies
- **History archives/databases**  
archives of historical documents (texts, maps, pictures, etc.), databases of analytical historical data and metadata
- **Digitised manuscript databases/ collections/ repositories**  
image/text collections/repositories of digitised manuscripts, databases of analytical data and metadata
- **Arts & Art History databases/collections/repositories**  
collections/repositories of works of art/their digital replicas, *in situ* locations, databases of art history data and analyses
- **Archaeology, Anthropology and Ethnology databases/ collections/ repositories**  
findings/digital replicas, *in situ* locations, databases of archaeological/anthropological/ethnological data and analyses

### Transversal RIs

- **Research libraries**
- **Research archives**
- **Large-scale research bibliographies**
- **Education databases/ collections/ repositories**
- **Digital collections/ Data repositories**  
general type digital repositories
- **Analogue audio/visual/multimedia**  
collections/repositories/datasets of analogue recordings
- **Conceptual models, ontologies, thesauri**  
conceptual networks, taxonomies developed in each discipline or for cross-referencing purposes
- **Geographical information systems data facilities**  
computer systems for processing data that are spatially referenced
- **Timeline data facilities**  
computer systems for processing data that are chronologically referenced on macro (dates, years, centuries, millennia) and/or micro (hours, minutes, seconds, parts of seconds) scales
- **Visualisation facilities**  
visualisation tools and services capable of visually presenting data from different sources/ collections/ repositories, etc.
- **Software development centres of competence**  
software development for general and/or specific purposes
- **Natural History collections**  
collections of naturalistic objects of interest for archaeological, anthropological and ethnological research



**Figure 3.** Fig. 1. The traditional scholarly information continuum and Fig. 4. A de-constructionist scholarly information continuum from Gradmann, Stefan, and Jan Christoph Meister. "Digital document and interpretation: re-thinking 'text' and scholarship in electronic settings". *Poiesis & Praxis* 5.2 (2008): Electronic pre-publication, 22 April 2008: <<http://www.springerlink.com/content/g370807768tx2027/fulltext.html>>.

## Bridging Physical RIs in the Humanities with Digital RIs

The Humanities and physical RIs (such as specialist libraries) exist in a symbiotic relationship with one another. Physical RIs collect, order, make accessible and reusable the primary and secondary sources that form the basis of scholarly work in the Humanities and give Humanities scholars access to wide-ranging human expertise, for example, subject librarians or information professionals. Upon completion of the research process – which is often facilitated by such physical RIs via the resources they provide – the research outputs of Humanities researchers, e.g., monographs, journal articles, scholarly editions or reference works such as dictionaries are then absorbed back into the RI and the research and creative cycle starts again.

Research into the application of computing technology to Humanities has been ongoing since the 1940s at least (see footnote 19), but since the advent of the internet in particular, ever increasing proportions of resources – whether text, image or sound – are being remediated in

17. McGann, Jerome, Ed. "Introduction. Sustainability: the Elephant in the Room." *Online humanities scholarship: the shape of things to come*. 26-28 March 2010, University of Virginia. Ed. Jerome McGann. *Connexions*. 8 May 2010. Web. 18 July 2011. <<http://cnx.org/content/col11199/1.1/>>.



digital form. The most immediate implication of this has been elegantly expressed by McGann<sup>17</sup>:

“As with the renaissance sped forward by the printing revolution of the fifteenth century, digital technology is driving a radical shift in humanities scholarship and education. The depth and character of the change can be measured by one simple but profound fact: the entirety of our cultural inheritance will have to be reorganized and re-edited within a digital horizon.”

Physical RIs in the Humanities tend to cater to users far beyond the academic community, including the general public and numerous levels of the educational system. Today, most RIs are engaged in digitising their holdings. Deegan and Tanner<sup>18</sup> have discussed a number of the benefits that digitisation can offer institutions, including making frequently used items in their collections available online as well as the possibility of ‘virtual reunification’ of collections.

Not only do these developments underline the value of digital RIs that, in turn, collect, order, make accessible and reusable our digital datasets (whether remediated or born-digital), but such developments are also resulting in a new kind of symbiosis between physical and digital RIs on the one hand and the Digital Humanities<sup>19</sup> research community on the other. Over the past years a number of initiatives<sup>20</sup> have identified the importance of developing digital objects (including tools), workflows and methodologies that are transferrable, capable of being repurposed and sustainable. Such aims do not invalidate research into the development of specialist digital resources that are created in order to explore a particular research question or to develop new research questions. Notwithstanding this, the potential benefits of moving away from the ‘digital silo’ and ‘disciplinary

silo’ models are clear. Not only do such models result in a fragmentation of research undertakings and outputs but “lack of IT capacity and sustainability are major threats to the continuity of our digital research sources, tools, and results.”<sup>21</sup>

Just as it is desirable for individual researchers and teams of researchers to move away from a ‘digital silo’ model, so too is it desirable for digital RIs to operate in a similar context. Accordingly, participation in a digital RI is becoming a question of survival not only for European research institutions, libraries and archives but also for Humanities itself. While many national RIs own very important electronic archives their patrimony will remain silent without the possibility of interoperability in an open access environment. So too, Humanities in Europe, which is often bound by various national languages, will benefit from greater access to a culturally broader and more varied set of empirical data (see section ‘Cultural and Linguistic variety – transnational RIs’ below). Absence of such data sets threatens to make research in the Humanities too confined to data that is easily available or that reflects narrow national contexts and developments only.

In order to maintain and push forward digital RIs, a number of key changes, both within academia and Humanities itself are required.

RIs for Humanities can be developed only as a result of a multidisciplinary collaboration of expertise from across the disciplines. The following list aims to be neither comprehensive nor proscriptive (especially in terms of the areas that the professions mentioned below might work in) but rather aims to indicate the many disciplines that have an important role to play in RIs for the Humanities. These include:

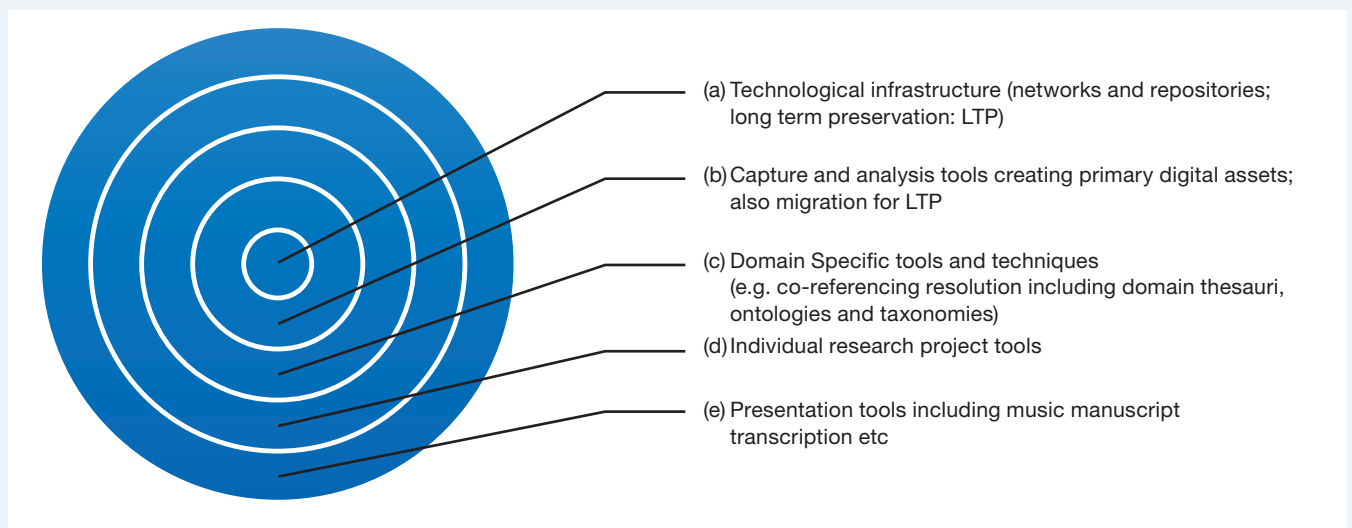
- Library and archive professionals to draw on their knowledge of physical and digital RIs and related issues such as curation and preservation;
- Digital Humanities researchers, adept at working at the interface of Humanities and Computing, for example, to create computational models of humanities sources and engage in other aspects of the digital remediation, analysis, theory and philosophy of sources;
- Computer scientists and software engineers to, for example, design appropriate technical infrastructures (such as networks, architectures, repositories, mechanisms for long-term preservation) and software;
- Information specialists to engage in the kinds of knowledge elicitation and information behaviour stud-

18. Deegan, Marilyn, and Simon Tanner. *Digital Futures: Strategies for the Information Age*. Digital Futures series. London: Library Association Publishing, 2002. 32-33. Print.

19. The field now known as *Digital Humanities* aims to use “information technology to illuminate the human record, and [bring] an understanding of the human record to bear on the development and use of information technology”, Schreibman, Susan, Ray Siemens and John Unsworth. “The Digital Humanities and Humanities Computing: An Introduction”. *A Companion to Digital Humanities*. Eds. Susan Schreibman et al. Oxford: Blackwell, 2004. xviii. Print. Traditionally, it traces its most immediate origins back to 1949, when Father Roberto Busa started the electronic processing of the complete work of St Thomas Aquinas in order to produce an exhaustive index of the lemmatised words. It was thus possible to obtain a basis to better interpret the theological thought of St Thomas. The principal work is the *Index Thomisticus: Sancti Thomae Aquinatis Operum Omnium Indices et Concordantiae*. Stuttgart: Frommann-Holzboog, 1974-1980. Print. This work was also made available in CD-ROM (1990) and then in DVD.

20. See for example the COST Action *InterEdition*: <<http://www.interedition.eu/>>.

21. <<http://www.interedition.eu/>>.



**Figure 4.**

The ‘onion rings’ represent different degrees of disciplinary input, and abstraction from general technologies, into the specific research infrastructure for areas of the Digital Humanities. At the core – level (a) – lies the underpinning technologies, which are generic ICTs to store, manipulate, communicate and maintain any collection of data. At the next level (b) the Digital Humanities assets are constructed so as to support different lines of Humanities enquiries. This requires co-development between Humanities and computing professionals, with each offering specialist skills and together designing the required infrastructure. The digital representations chosen will influence not only the forms of interrogation that can be made of the elements, but also the mechanisms required to provide long-term preservation and accessibility. Thus, digitised images of individual pages of text will have different potential uses from text files of the content. Capturing both may offer the opportunity for additional enquiry, but also requires the long-term preservation of both and maintenance of the link between the image and the text. Level (c) involves the tools that are used to conduct domain specific research. These tools need to access embedded knowledge in the data and the metadata, which is represented by the domain’s ontologies, taxonomies and specialist thesauri to enable the semantic linkages and relationships to be used in formulating queries and answering them. Such tools should also allow humanists to experiment and browse the data in order to evolve novel research methods and interesting lines of enquiry. Many of these can be expected to involve detection and resolution of co-referencing with multiple data sources (i.e., recognising that the same entity is being referred to in multiple sources). These tools require co-development by bringing together humanists in different domain specialisations with computer scientists and information professionals. Level (d) contains tools that are evolved around specific projects, and are therefore more specific. If level (c) involves generic humanities research tools then level (d) may tune them for specific datasets, for example, by using the thesauri that relate to that dataset in content-based searches using natural language tools. Level (e) is the user interface where searches are formulated and the results of investigating relationships are presented back to the user. A number of paradigms for the presentation of results are likely to evolve as Digital Humanities develop more routine currency. Standard ways of presenting relationships and correlations detected in datasets are likely to evolve through experimentation and user education/acceptance of specific paradigms. In the future these interfaces will typically need to inter-relate different types of data sources. Almost all of these levels require genuinely interdisciplinary co-development, rather than the historic model of client/supplier divisions with user requirements analysis, specification and coding.

ies that can make explicit the research processes of the various disciplines of Humanities;

- Humanities scholars to engage actively in collaboration with colleagues listed above and, informed by new developments in technologies, to delineate the needs and requirements of the Humanities community in a continually reflective process.

The bridging of physical and digital RIs presents opportunities and challenges with implications that are treated throughout this paper. The following is a brief and necessarily selective summary of some of the new perspectives that digital RIs and the objects that populate them can bring to the Humanities, drawing briefly on observations from some of the key researchers in the field.

Cohen and Rosenzweig<sup>22</sup> (2004) have identified the seven qualities of digital media that “potentially allow [...] us to do things better” as capacity, accessibility, interactivity, flexibility, manipulability, interactivity and hypertextuality. But a moment’s thought on the nature of our cultural inheritance indicates that the ever increasing levels of capacity of computing and computing infrastructures offer researchers new ways to store, transmit and access the ‘data sets’ of the Humanities. For example, drawing on the capacity of grid computing, projects such as TextGrid<sup>23</sup> support a new kind of scholarly work and collaboration.

22. Cohen, Daniel, and Roy Rosenzweig. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Pennsylvania: University of Pennsylvania Press, 2005. 3-8; cf. their five problem areas 8-13.

23. See also similar efforts in the US within the Project Bamboo, an international partnership of 10 universities currently in its operational phase (2010-2012), funded by the Andrew W. Mellon Foundation to support the collaborative development of humanities research technologies: <<http://www.projectbamboo.org/>>.

## CASE STUDY

Andrea Rapp (TU Darmstadt), Oliver Schmid (TU Darmstadt),  
Michael Bender (Universität Trier/TU Darmstadt) – DE

### TextGrid – Virtual Research Environment for the Humanities

TextGrid<sup>24</sup> is one of the first virtual research environments (VRE) of its kind currently available to the Digital Humanities. It is for the moment available as beta-version software with a focus on text-oriented research. With the release of version 1.0 in June 2011, TextGrid offers a stable infrastructure productively usable in research projects. TextGrid is part of a larger infrastructure, the D-Grid initiative, and is funded by the German Federal Ministry of Education and Research (BMBF).

TextGrid serves as a VRE for philologists, linguists, musicologists and art historians; disciplines that have recently joined include philosophy, Jewish studies, archaeology and the history of science. Built on grid architecture, its web-based platform will provide services and tools for researchers for analysis of textual data in various digital archives – independent of data format, location and software. An important objective of TextGrid is to develop tools that allow enhanced retrieval, manipulation and analysis of data. The focus is not only on letter-based information, but also music notation and digitised art objects.

The core of the TextGrid infrastructure is the so-called TextGridRep, a grid-based repository for reliable storage, sustainable availability and access to research data. It will also ensure interoperability. The independence of data formats is provided by the use of standards, e.g., of the *Text Encoding Initiative* (TEI) which collectively develops and maintains guidelines for the representation of texts (especially in the Humanities and Social Sciences) in digital form. Access to the repository is provided by the TextGridLab (Laboratory), a web-based platform which grants access not only to content but also to scholarly tools and services. As portable software, the TextGridLab will not modify the computer's operating system; for example, it can also be started from a USB drive. Tools will be available globally that previously were available only on a local level.

In this way, a platform can be constructed in which experts, irrespective of time, location or subject, can concentrate on the problems of a constantly evolving research landscape, with the help of up-to-date methods and procedures. The crucial aim of the TextGrid concept is not only an accumulation of universally applicable tools; it is intended to be a platform for the exchange of tools and methods as

24. <<http://www.textgrid.de>>.

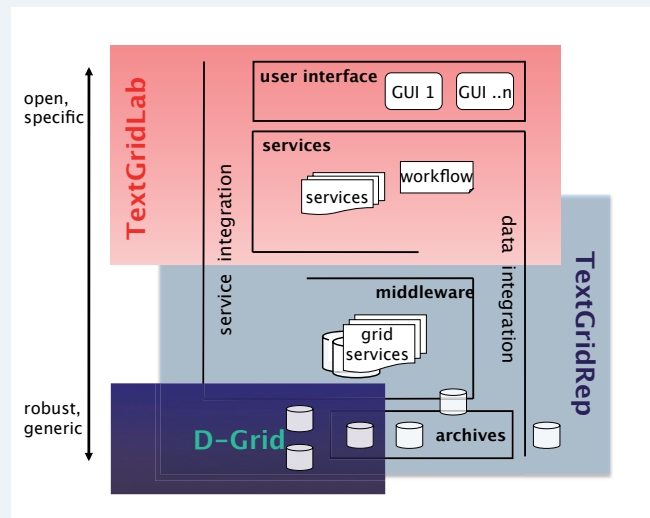


Figure 5.  
© TextGrid Team 2011 <<http://www.textgrid.de/>>.

well as content. In addition, the modular, open source platform can be extended easily for special projects or users' purposes.

The transformative insights into the nature of Humanities provided by the computational modelling of artefacts of cultural heritage offer an especially rich case, among many possible examples. As Willard McCarty<sup>25</sup> has observed, "I celebrate computing as one of our most potent speculative instruments, for its enabling of competent hands to force us all to rethink what we trusted that we knew".

Aspects of modelling are not new to Humanities researchers, as made clear in McCarty's discussion of analytic modelling, which, he argues, involves analysing how something works "by taking it apart, for example, when a literary critic dissects a poem to understand how it works and what it does".<sup>26</sup> In the Digital Humanities, this activity is extended and transformed when a computer is used to represent the scholar's (rarely fixed) understanding of how the poem works, or how it can productively be taken apart or explored. When using a computer to model a poem, for example, the model must be created within the constraints of computing technology and so complete explicitness and consistency are required. This may appear as a tall, if not impossible task for many aspects of works of imagination and learn-

25. McCarty, Willard. "Attending from and to the machine." Kings College London. 2 Feb. 2009. Inaugural lecture. Web (accessed 14/07/2011).

26. McCarty, Willard. "What's going on?" *Literary and Linguistic Computing*, 23.3 (2008): 256. Print.

ing, as well as for the scholars who seek to model them. Significantly, the challenges of reconciling quantitative, computational methods with interpretative frameworks remain some of the most productive areas of inquiry for Digital Humanities. McCarty's conclusion firmly reiterates the research potential of computer modelling:

"Its great and revolutionary success for the humanities is to force the epistemological question – how is it that we know how we somehow know – and to give us an instrument for exploring that". (*ibid.*)

A common retort, if not defining statement of our age, is that users are suffering from information overload.<sup>27</sup> However, humans have always tried to deal with making information accessible and recoverable in repeatable ways (e.g., categorisation, cataloguing, development of taxonomies, etc.); traditional RIs have played a crucial role in this. As ever increasing portions of our datasets become available in digital form the phrases 'data deluge' and 'information overload' become ever more common. In the Humanities this is no less true, and the concomitant consequences and opportunities for the present-day and future research have been expressed by Gregory Crane<sup>28</sup>:

"As Solon points out in *The History of Herodotus*, there are only about 30,000 days in a human life – at a book a day, we would need 30 generations to read through even a moderate collection of a million books and 10,000 years to cover the 10 million-or-so unique items in the Harvard Library system."

The making available of our cultural heritage in digital form and the sensitive interlinking of those resources opens a new frontier for Humanities research and provides us with opportunities to questions and trace patterns in datasets that are becoming available (and, *caeteris paribus*, accessible) on an unprecedented scale.

"Ultimately, in computer-assisted analysis of large amounts of material that has been encoded and processed according to a rigorous, well thought-out system of knowledge representation, one is afforded opportunities for perceiving and analysing patterns, conjunctions, connections, and absences that a human being, unaided by the computer, would not be likely to find."<sup>29</sup>

27. Cf. in relation to the data deluge/grand challenge issue the EU report *Riding the wave. How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data, A submission to the European Commission, October 2010. Web (accessed 14/07/2011).

28. Crane, Gregory. "What Do You Do With A Million Books?" *D-Lib Magazine* 12.3 (2006). Web (accessed 14/07/2011).

29. Schreibman, Susan et al. *A companion to Digital Humanities*. xviii.

Indeed, rich and robust digital RIs in the Humanities are a support to researchers in addressing also the hard sciences-driven grand challenges of our time such as is the case of historical databases used for medical research<sup>30</sup> or archival resources informing climate change research:

"Hundreds of Royal Navy ships' logbooks have been digitised dating from the 1760s to 1923. The accurate weather information they contain is being used to reconstruct past climate change – hitherto untapped scientific data. Digitising over 20,000 fragile photographs from 1845-1960, the Freeze Frame archive delivers to a worldwide audience some of the most important visual resources for research into British and international polar exploration. Personal journals and official expedition reports provide a narrative to the photographs."<sup>31</sup>

## CASE STUDY

Brett Boble, *US National Endowment for the Humanities, Office of Digital Humanities – USA*

### The Digging into Data Challenge

Traditionally, cultural heritage materials have been the object of study for many researchers in the Humanities and Social Sciences. Books, newspapers, journals, paintings, survey data, census data, music, film, audio, sculpture, epigraphs and other materials form the core dataset for study. The historian might spend years reading old newspapers and books for clues about the past; the archaeologist might study ancient cuneiform tablets to try to interpret what they mean; the art historian might study paintings and sculpture; the economist might study census data and tax records.

But in the past few years, this landscape has changed. Millions of books, newspapers, journals, photographs, audio and video recordings that were once held in libraries, museums and archives are now widely available via the web. New, born-digital data are being created at an enormous rate. What is remarkable about these digital materials is the sheer scale. Never before have scholars had access to such a huge volume of materials. This kind of scale adds new challenges and new opportunities. A scholar of 19<sup>th</sup> century literature

30. An example of this is the *Demographic Data Base* developed at Umeå University and being used particularly for research into genetic diseases. Indeed, information on age at death, cause of death and kinship dating back to the 19th century is provided in this resource, which can therefore be exploited to learn more about the mechanisms behind various genetic diseases: <<http://www.ddb.umu.se/ddb-english/?languageId=1>>.

31. Tanner, Simon. *Inspiring Research, Inspiring Scholarship: The value and benefits of digitised resources for learning, teaching, research and enjoyment*. The Higher Education Funding Council for England (HEFCE), JISC, 2010. 11.; Web (accessed 14/07/2011).

could never hope to read every book published in the 1800s – but a computer can. An historian who is studying World War II could never hope to read every newspaper editorial about the war – but a computer can. A sociologist could never hope to read every letter written by major figures of the Enlightenment – but a computer can.

We have only begun to scratch the surface on how this mountain of data might be used to advance scholarly research. What new knowledge can we acquire? What new questions might the data drive us to ask? How might it help the scholar locate new materials ripe for close reading? How might old theories be questioned and new ones posed?

To address these issues, in 2009, four international research funders from the United States, Canada, and the United Kingdom announced and ultimately made eight awards during the first round of the *Digging into Data Challenge*. Then in 2011, four additional funders (and one additional country, the Netherlands) joined the endeavour and announced round two of the *Digging into Data Challenge* with awards slated to be announced in December 2011.

### What is the ‘challenge’ we speak of?

The idea behind the *Digging into Data Challenge* is to address how ‘big data’ changes the research landscape for the Humanities and Social Sciences. Now that we have massive databases of materials used by scholars in the Humanities and Social Sciences, what new, computationally-based research methods might we apply? As the world becomes increasingly digital, new techniques will be needed to search, analyse and understand these everyday materials. *Digging into Data* challenges the research community to help create the new research infrastructure for 21<sup>st</sup> century scholarship.

Applicants will form international teams from at least two of the participating countries. Winning teams will receive grants from two or more of the funders and, two years later, will be invited to show off their work at a special conference sponsored by the eight funders (Social Sciences and Humanities Research Council of Canada, Netherlands Organisation for Scientific Research, UK Joint Information Systems Committee, UK Arts and Humanities Research Council, UK Economic and Social Research Council, US Institute of Museum and Library Services, US National Endowment for the Humanities, US National Science Foundation).

Notwithstanding the new perspectives that RIs can open to the Humanities it is not proposed that digital RIs will or should replace traditional RIs. The relationship that exists between the digital and analogue resource is complex and dynamic, as Warwick<sup>32</sup> has argued: “It appears that we are moving beyond not printed books and print-based scholarship, but the naive belief that they can easily be replaced by digital resources”. Furthermore, projects such as LAIRAH<sup>33</sup> have demonstrated that far from making library professionals obsolete, the availability of digital resources and research infrastructures are making their services and expertise all the more necessary – as scholars working in digital and analogue formats require all the more advice about judicious selection of resources.<sup>34</sup>

### CASE STUDY

Elisabeth Kieven, *Bibliotheca Hertziana, Max Planck Institute for Art History* – DE/IT

## Research Infrastructures for Historic Artefacts: Knowledge Networks

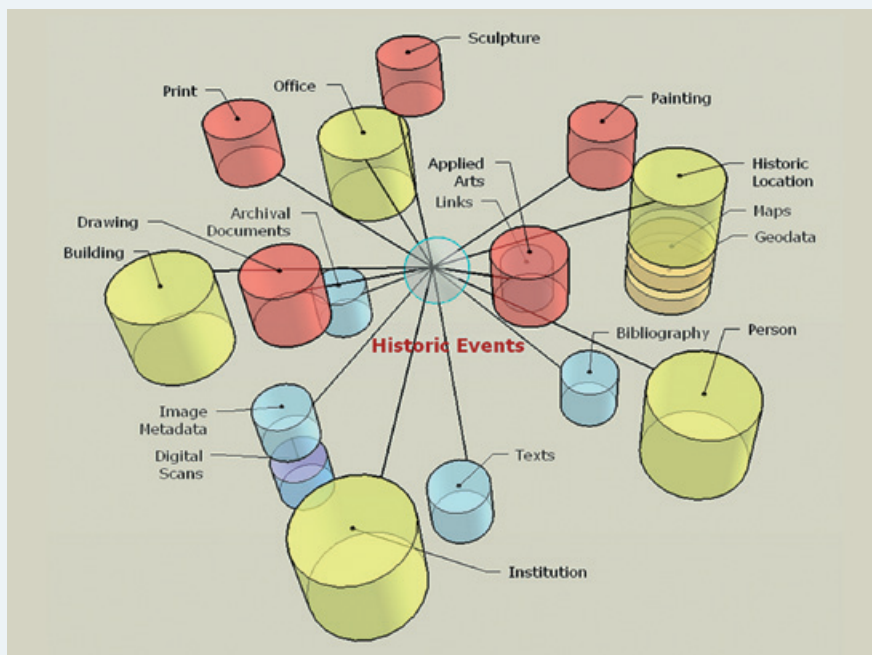
Whereas texts may be published on the internet without detriment to their integrity, works of art – as any real-life object – cannot be represented virtually through digital images and metadata without loss. So far, cultural artefacts have mainly been catalogued digitally in a comparatively simple way, often without adequate visual documentation, such as standalone museum inventories published on the internet.

### Metadata-centred storage

These standard repositories are no longer sufficient for current research purposes, as artefacts are not ‘self-contained’: they carry additional cultural information referring to their production, reception and subsequent history which is of particular interest to scholars. This information itself is subject to interpretation and needs scholarly evaluation before publication. Due to these circumstances, requirements for word-image metadata (standards and data models) and technical infrastructures (storage and retrieval systems) are

32. Warwick, Claire L.H. “Print Scholarship and Digital Resources.” *A Companion to Digital Humanities*. Eds. Susan Schreibman et al. 379. Print.  
33. *Log Analysis of Internet Resources in the Arts and Humanities (LAIRAH)*. AHRC Information and Communication Technology in Arts and Humanities Research Programme, Strategy Projects scheme, 2005-2006. Web. For more details see <<http://www.ucl.ac.uk/infostudies/research/circah/lairah/> (accessed 14/07/2011).

34. Warwick, Claire L.H., Melissa Terras, Isabel Galina, Paul Huntington and Nikoleta Pappa. “Library and information resources and users of digital resources in the humanities.” *Program: Electronic Library and Information Systems*. 42.1 (2008): 5-27. Print.



**Figure 6.** Simplified data model, showing entities interconnected by 'historic event' relations.

more sophisticated and at the same time still less developed than in text-oriented disciplines.

### Knowledge networks

A step in this direction are formal ontologies as the basis for knowledge representation networks is the CIDOC conceptual reference model,<sup>35</sup> which describes actors, objects and relationships. At the moment only a few resources are advancing in these directions.

One example of a database enriched by a conceptual reference model of this kind is the ZUCCARO<sup>36</sup> information system of the Bibliotheca Hertziana<sup>37</sup> (Max Planck Institute for Art History), which unites the traditional repositories of the library and the photographic collection with specialised resources such as the architectural drawings research database Lineamenta or the project ArsRoma which focuses on historic evidence for painters and painting in Rome around 1600.

ZUCCARO is not merely an image repository enriched by metadata. The information is split into basic units, which are interconnected by formal representations of 'historic events', which in turn specify the type of relationship. For example: a person – the Marchese Sacchetti – erects a building at a certain historic moment – the Villa Sacchetti in Rome. He commissions it from another person – the architect and painter Pietro da Cortona. The Marchese belongs to a social 'institution' – an important aristocratic family; as such he is

also a member of certain religious confraternities. The artist is a member of the Roman Academy of St Luke and other social structures. His villa, with various buildings and interior decoration, which today no longer exists, is documented in maps, architectural drawings, *vedute*, prints, descriptions and so forth. Traditional databases cannot represent such a complex network of relationships covering, for example, the social background of clients and artists and the social importance of art and architecture in 17<sup>th</sup> century Rome.

The event-based data model is very flexible and much more suitable for scholarly research:

- Data can be viewed from a great variety of perspectives
- Data can be consulted in various ways which may also differ from the original purpose
- Content can be accumulated and corrected over long time periods, consequently becoming denser and more reliable
- Scientific documentation can be added in a flexible way
- External documents and resources can be linked in easily

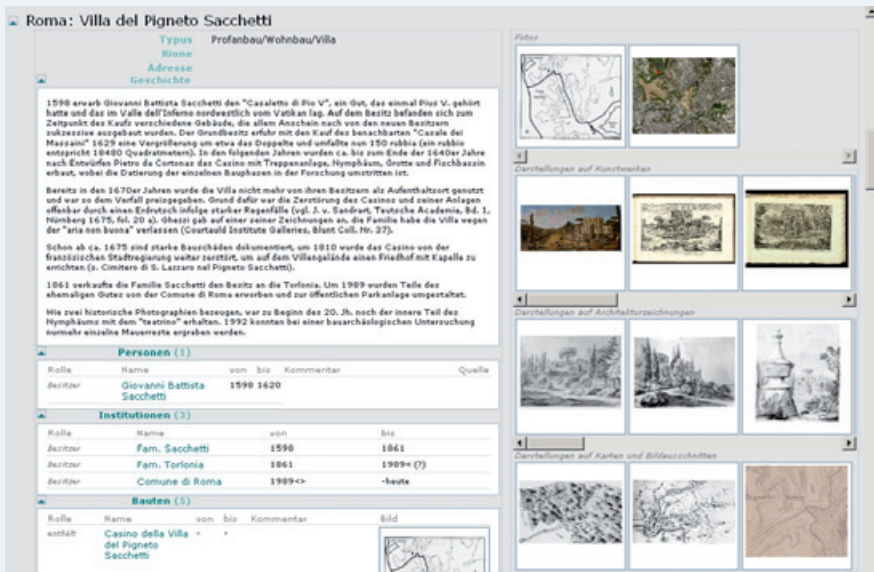
### Data input and retrieval

Complex databases based on ontology models meet a wide range of scientific interests. However, the richness of perspective, which is achieved through a high degree of data segmentation, has a downside. User interfaces which can give access to this wealth of information can be very complicated to devise and to use. This is true for data input as well as for retrieval. Data cannot simply be added to the database and piled up like file cards: every piece of information must be isolated first, and then has to be connected to many other pieces. This is a time-consuming process which in many

35. Cf. <<http://www.cidoc-crm.org/>>.

36. <<http://zuccaro.biblherz.it>>

37. <<http://www.biblherz.it>>



**Figure 7.** Composite data sheet, displaying rich information derived from the corresponding relationships.

cases is not easily automated, if only for subtle variations of spelling. Entry forms which collect much of the data at once are difficult to conceive and to manage.

The retrieval of the data is not an easy process either. How is a question like “I want to see all objects in the Sacchetti collection which the Marchese bought in Florence before a certain date” to be reformulated in a language which the computer system will understand? Either the user himself analyses the question and splits it up into a series of interdependent queries, or the system must be able to do it automatically. Especially for untrained users who are accustomed to a Google-like interface, the first solution is not viable. Practical experience is rare in this field, and getting a timely and satisfactory answer from the system is quite a challenge in terms of database performance, as also the current prototype of ZUCCARO shows. More research on modern, flexible and understandable user interfaces is certainly needed.

### Authority files and scholarly reliability

Artefacts (cultural heritage objects) and other entities (persons, places, roles) must be identifiable through a worldwide system of unique identifiers. In some cases ‘master data’, cared for by big research institutions, is available. Research databases like ZUCCARO can provide mappings from their content to these authorities. At this point more standardisation (unique and persistent identifiers) on authority files is necessary.

‘Copyright’ problems also tend to obstruct modern web-based research and publication procedures. Access to images from the realm of cultural heritage, where most of the originals have been in the public domain for a long time, gets more and more restricted by tightened ‘copyright’ legislation.

### Conclusion

Digital technology has found its way into modern art history, especially in the same domains as in other humanist disciplines. As digital image analysis still has a long way to go before it will be useful for scholarly research, the current challenge is still to build up and manage collections of meta-data. The most promising sector seems to be the worldwide interconnection of data, the consistent separation of single information units and the use of ontologies as patterns for creating knowledge networks.

## Researchers’ Input and Engagement in Producing RIs

In the traditional Humanities, pronounced dividing lines have existed between archivists, cultural heritage professionals, scholars and library and information professions. In the context of digital RIs such dichotomies are now being broken down dramatically, as Humanities researchers engage ever more in the re-mediation of cultural heritage, by participating in digitisation projects, for example. In addition to the interdisciplinary nature of such research as well as the necessary changes in the publication and evaluation cultures in the Humanities, key issues have emerged regarding the research input and engagement of researchers in producing digital RIs in the field of the Humanities.

In the light of the experiences gathered in the past decade the following recommendations can be defined as of major and essential importance.

- The establishment of partnerships and productive alliances across communities and institutions (scholarly community/libraries/ archives/museums) based on equality, a shared working vocabulary and mutually beneficial collaboration (such a cooperation platform can be perceived as being an infrastructure in itself).<sup>38</sup>
- The identification of shared obstacles to alliances between librarians, archivists, curators, on the one hand, and the academic community on the other. Possible obstacles include: potential or perceived conflict between a library's request for free, reusable data and researchers seeking to add value/retain recognition of work. This aspect is particularly important when the research activity, aimed at in-copyright publication (digital or analogue), is unlikely to be used as an 'open access' resource within an RI for an entire community of scholars.
- The pursuance, where appropriate, with private sector partners (publishing houses, software companies, internet service providers, companies producing multimedia and digital systems, etc.) of collaborations that could further the creation of RIs.
- The enhancement of the networked dimension of RIs and also the potential for the inclusion of non-experts.<sup>39</sup> This includes fostering the social dimension of RIs, especially in terms of outreach beyond academia and the engagement of the wider public ("citizen science"),<sup>40</sup> e.g., through the participation of non-academic staff in crowd-sourcing projects.<sup>41</sup>
- The dissemination of scientific results of research products arising from/facilitated by RIs within the single communities of scholars, enhancing the advantages derived from the use of materials and services available in an RI.<sup>42</sup>

38. Examples of these within the libraries community are the European Library in Europe (<http://www.theeuropeanlibrary.org/>) and the Hathi Trust in the US (<http://www.hathitrust.org/>) which also includes research institutions.

39. See, for example, Terras, Melissa. "Digital curiosities: resource creation via amateur Digitization." *Literary and Linguist Computing* 25.4 (2010): 425-438. Print.

40. As far as engaging the public in science is concerned, Liz Lyon quotes: "We are now seeing a veritable resurgence in citizen science with the social culture of the Web beginning to influence and radically change the way science is performed." Lyon, Liz. *Open Science at Web-Scale: Optimising Participation and Predictive Potential. Consultative Report*. Joint Information Systems Committee (JISC), UK, UKOLN, Version V1.0, November 2009, p.25. Web (accessed 4 August 2011). <<http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf>>.

41. See for instance the *Transcribe Bentham: A Participatory Initiative*: <<http://www.ucl.ac.uk/transcribe-bentham/>>.

42. An interesting initiative in this direction and in line with a sort of 'experimental turn' in the Humanities is the *Stanford Literary Lab* created to "discuss, design, and pursue literary research of a digital and quantitative nature": <<http://litlab.stanford.edu/>>.

In order to achieve these aims on a broad scale and at a scholarly and public level, existing models of good practice should be disseminated to provide information and education to researchers on how to build new and effective community infrastructures. Inherent to this issue is that of the peer recognition of digital driven scholarly research and its equal evaluation in comparison to traditional research (see infra, 'Evaluation of Digital Research and its Outputs').

#### CASE STUDY

Elton Barker (*The Open University*), Stefan Bouzarovski (*University of Birmingham*), Chris Pelling (*Christ Church, Oxford*), Leif Isaksen (*University of Southampton*) – UK

### HESTIA (the Herodotus Encoded Space-Text-Imaging Archive): an interdisciplinary project

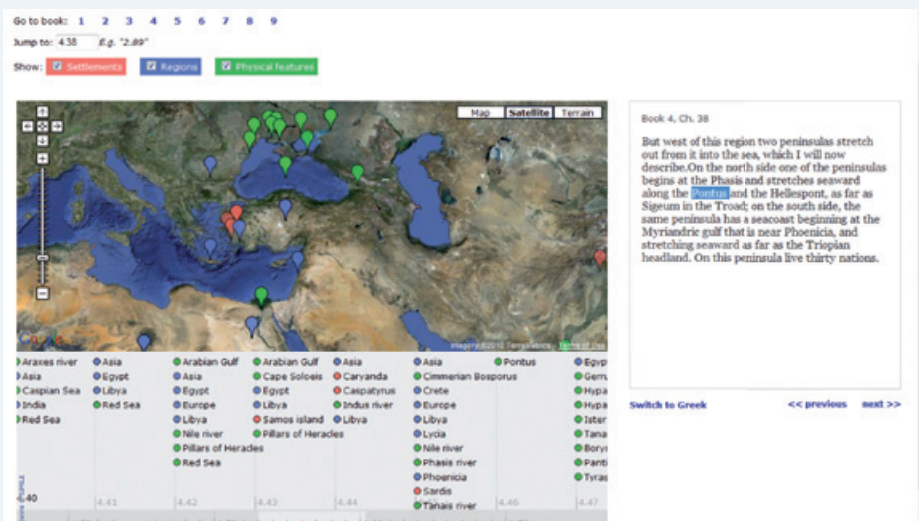
HESTIA<sup>43</sup> uses digital technology in combination with close textual study to investigate the geographical concepts through which Herodotus's *Histories* describe the conflict between Greeks and Persians. This short case study draws on the experience of HESTIA to consider three of the main themes addressed at the ESF strategic workshop in October 2010 (interdisciplinarity, repurposing of data and text vs non-text), the research scope afforded by the collaboration and some closing remarks towards building a research infrastructure.

#### Interdisciplinarity

HESTIA is an interdisciplinary team involving the authors of this case study: two classicists, a social geographer and a digital humanist. The way in which this team came together was serendipitous but underpinned by a number of institutional contexts, including an interdisciplinary university environment and a funding agency (the Arts and Humanities Research Council, UK) looking to support early career researchers. The IT component gave our proposal to that body an important additional dimension, which in combination with the interdisciplinary outlook must have contributed to the successful award. But it should be noted that having a technical consultant greatly assisted the project: by embedding the expertise in the team itself the project avoided the usual problems with university computing services, often expected to offer assistance without any knowledge of the academic subject.

43. See <<http://www.open.ac.uk/Arts/hestia/>>.





**Figure 8.** A 'timeline' map of Herodotus, with places visualised corresponding to their mention in the text, after Nick Rabinowitz <<http://code.google.com/p/timemap/>>.

## Repurposing of data

When we submitted the proposal, we anticipated marking up the text of Herodotus to capture place information. Fortunately the *Perseus Digital Library*<sup>44</sup> makes available classical texts with place information already encoded (by computer), which may be used and adapted by being released under a *Creative Commons Attribution Non-Commercial Share-Alike* licence. Perseus has followed HESTIA with keen interest precisely because, by using one of their texts, we have demonstrated a scholarly value to their work. Indeed, we will be handing a hand-verified Herodotus text back to Perseus, thereby promulgating the mutually beneficial exchange of data, knowledge and expertise.

## Texts vs non-texts

With the text in digital form, the project is able to capture the great majority of the places that Herodotus mentions and display them in various web-mapping technologies. The aim here, however, is not only to introduce Herodotus's world to an internet audience, though dissemination is important. Maps are also used innovatively, not as products of ideology but as tools with which to interrogate spatial concepts. In fact, it is through the visual medium that a better sense of the ways in which space is narrativised in the Histories is to be gained.

## Research scope

HESTIA utilised its interdisciplinary framework to trial different methods of extracting, visualising and interrogating the connections that Herodotus draws between two (or more) places, thereby allowing the analysis of the Histories' embedded narrative networks. The twin methodologies trialled were:

- A qualitative analysis of one stretch of narrative (book 5) based on the geographical principles of movement and transformation, and including the geographical concept of the 'proxy' (the peoples, individuals or even non-human agents that may be linked to a physical space).
- Database-generated network maps. Given the time-consuming and highly complex nature of the qualitative analysis, a key alternative strategy has been to use the database to generate rapid networks (based on the simple co-presence of terms within sections of the text) to flag up potential links between locations that may repay further study.

These two alternative strategies were intended to complement, challenge and inform each other, but in themselves have offered few possibilities for data exchange. It is a desideratum that in the future the two approaches should be brought together by using text mining techniques to develop a typology based on the automatic generation of networks.

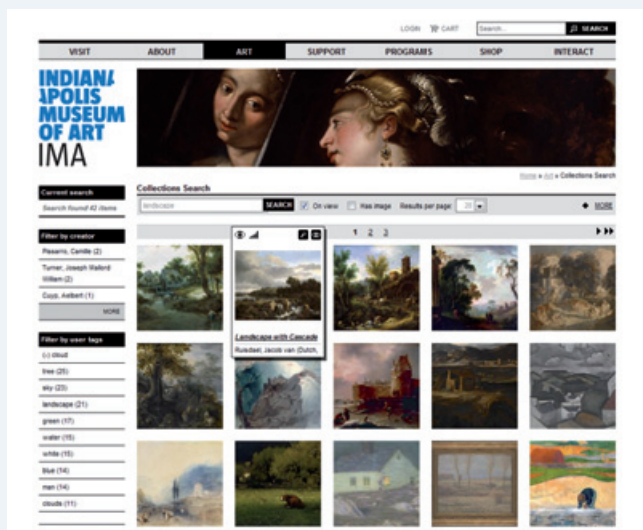
## Integrating extant resources

HESTIA is in contact with a number of other projects, including Perseus (see above) and *Pleiades* – an online gazetteer of ancient places using Uniform Resource Identifiers (URIs) as non-ambiguous identifiers of place. The collaborations demonstrated the desirability of linking different datasets. In fact EB and LI have since gained funding from JISC (*Joint Information Systems Committee, UK*)<sup>45</sup> for the *Pelagios project*<sup>46</sup> which, as an international consortium working to establish generic processes for referencing ancient places in digital documents (texts, tables, images, etc), offers a 'bottom up' paradigm to building a research infrastructure.

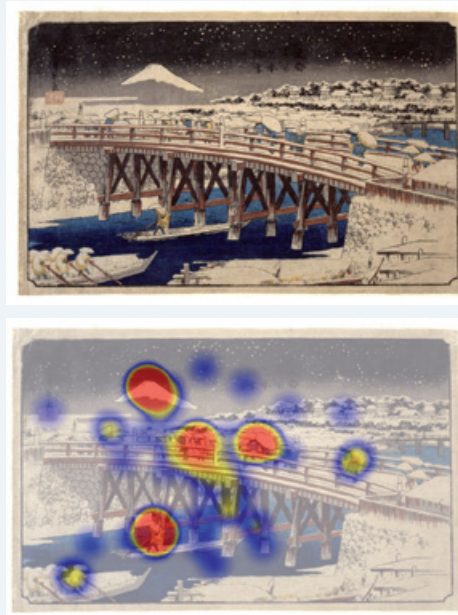
45. See <<http://www.jisc.ac.uk/>>.

46. See <<http://pelagios-project.blogspot.com/>>.

44. See <<http://www.perseus.tufts.edu>>.



**Figure 9.** Indianapolis Museum of Art collection search for 'landscape' filtered by tags for 'cloud' <<http://www.imamuseum.org/art/collections/search>>.



**Figure 10.** Utagawa Hirōshige (Japanese, 1797-1858) – Nihonbashi in the Snow; Click density overlay showing locations of taggers' 'first impressions'.

## CASE STUDY

Robert Stein, *Indianapolis Museum of Art – USA*

### Crowd-Sourcing Art History: Research and Applications of Social Tagging for Museums

The advent of low-cost techniques for digitisation and scanning has resulted in an ever-increasing number of datasets for scholarly research. The widespread availability of online text and media databases has changed the nature of scholarship in the Humanities. As a result, scholars have many new opportunities for discovery, but are also beginning to recognise new challenges related to the size and scope of available resources. While computer systems facilitate the creation of research infrastructures capable of processing these comprehensive datasets, there are still many tasks for which computers cannot match the innate intuition of human perception.

Recent work by the Steve.Museum project<sup>47</sup> has sought to understand the ways that user-generated data can enhance existing knowledge, improve access to online collections, and engage visitors and enthusiasts in contributing valuable information to Humanities datasets. The current Steve.Museum dataset consists of 86,720 objects from 21 institutions and has gathered some 478,000 user-generated descriptive tags.

47. Steve: *The Museum Social Tagging Project*; see <<http://www.steve.museum>> and <<http://tagger.steve.museum>>.

Early research by the project team (Trant, 2007) demonstrated that these social tags offer substantial new contributions to existing collection documentation. Museum staff judged 88% of tags to be 'useful' and nearly 100% of tags were 'useful' when they occur 2 or more times for a single image. Anecdotally, the project team found that while the project research featured over 4,000 taggers, even a small number of taggers can produce a large amount of valuable data.

More recently, the project has continued its research collaborations examining how techniques from Computational Linguistics might augment tag datasets, creating links between lexically or conceptually similar tags, and by examining similarities and differences between user tags and keywords extracted from extended text. Initial results demonstrate that applications of simple morphological and lemmatisation techniques can merge over 20% of one-word tags<sup>48</sup>. Other techniques are being applied to multi-word tags to determine whether parts-of-speech or bi-grams analysis helps disambiguate the meaning of these tag-phrases.

The project is also creating tools that can be easily inte-

48. Klavans, Judith, Robert Stein, Susan Chun and Raul David Guerra. "Computational Linguistics in Museums: Applications for Cultural Datasets." *Museums and the Web 2011, 6-9 April 2011, Philadelphia, US*. Eds. J. Trant and D. Bearman. Toronto: Archives & Museum Informatics, 31 March 2011. Web (accessed 14/07/2011). <[http://conference.archimuse.com/mw2011/papers/computational\\_linguistics\\_in\\_museums\\_applicati](http://conference.archimuse.com/mw2011/papers/computational_linguistics_in_museums_applicati)>

grated into web-based systems supporting the needs of museums and scholars. *Figure 9* demonstrates how social tags, from the Steve.Museum website, are used in the online collection search of the Indianapolis Museum of Art<sup>49</sup> (IMA). Tags are indexed by the museum's collection search engine and provide points of access for search results and faceted browsing of art objects. *Figure 9* shows a collection search for 'landscapes' filtered by objects tagged 'cloud'.

In addition to collecting descriptive words about objects, tagging systems can also support many different kinds of data. Vector tagging in Steve may include geographic locations, colour values, or pixel locations. The IMA conducted a small tagging experiment to discover what elements in an artwork are the first to capture the attention of viewers. *First-Impressions*<sup>50</sup> was featured on the IMA's blog and asked visitors to click on the point in an image that first caught their eye. Click-statistics were recorded for all respondents and displayed as heat-mapped overlays. *Figure 10* shows the results of click-tagging on one Japanese wood-block print from the museum's collection. Notice how clicks are more concentrated over major elements in the image, but still manage to capture all the important features.

---

## Digital Research in the Humanities: who is responsible for RIs?

The web is now part of the daily life not only of young 'digital residents' but more and more the workplace of the humanist researcher, including the older generation of 'digital immigrants'. The latter often use the web and email for information, communication and to prepare publications. But many are far from able to take advantage of the full potential of the new research and publication facilities it offers.

The breakthrough for mass usage of digital devices came with the introduction of smartphones and tablets and the new way the standardised applications were distributed. It is currently common for the general user to choose across thousands of apps for special purposes. Similar facilities are not, however, available to the research community, especially those working in the Humanities. On the contrary, there are many hindrances to the use of digital research in the daily lives of Humanities researchers. There are no established standards; the researcher may lack a critical mass of digital data and material; and, last but not least, there are few incentives in academia to undertake digital research.

There are numerous examples of databases, tools and services in Humanities, but their lack of coherence is a significant problem. As a result, there is an urgent need for standards for metadata, for the organisation and interlinking of data and texts (semantic web) and for (Open Access<sup>51</sup> and Permanent Access)<sup>52</sup> publishing in text and data repositories.

The following is a brief summary of current challenges in this area:

- Digital research in Humanities is mainly project-driven; small scattered research groups are working to short timescales.
- Digital data and documents are volatile: they need long-term preservation.
- Digital objects in Humanities have to be able to be consulted for a long period: they need institutions responsible for maintaining them for future generations of researchers.

---

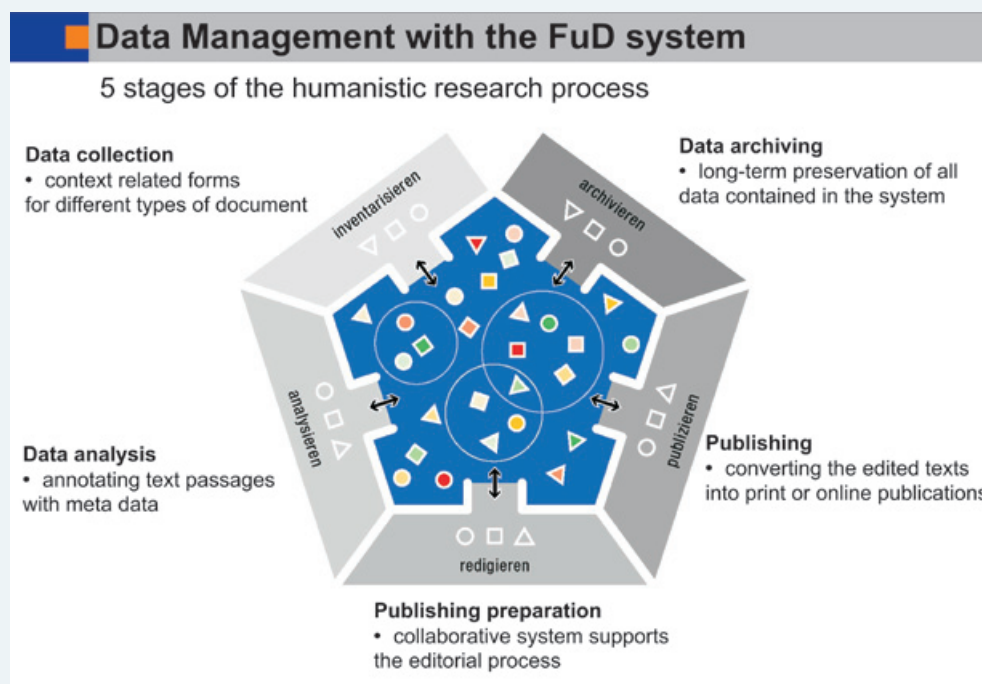
51. Open Access "refers to the practice of granting free Internet access to research articles. As all research and innovation builds on earlier achievements, an efficient system for broad dissemination of and access to research publications and raw data can accelerate scientific progress". See: <<http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1294&lang=1>>.

52. See especially the *Alliance for Permanent Access* (APA): <<http://www.alliancepermanentaccess.org/>>.

---

49. Indianapolis Museum of Art: Collection Search, <<http://www.imamuseum.org/art/collections/search>>.

50. Indianapolis Museum of Art: First Impressions, <<http://www.imamuseum.org/interact/first-impressions>>.



**Figure 11.** INF-Project Research Network and Database System (FuD) – CRC 600 Strangers and Poor People <<http://www.fud.uni-trier.de/>>.

### But who is responsible?

Every researcher in the Humanities in Europe must be assured of finding a service provider for their digital research activities, for short-term accessibility as well as long-term preservation of data and publications. A key dimension to the delivery of this objective is the concept of a Europe-wide Research Infrastructure for the Humanities with strong (virtual) centres, as developed in DARIAH and CLARIN, and designed to support the development, promotion and implementation of shared protocols and standards.

Additional intermediary services are often research community driven<sup>53</sup> and can provide an RI with generic and specialised software for research communities. A good example for the design of a research environment is TextGrid (discussed earlier) with its fundamental grid layer with middleware and archival functionalities, a repository layer and a set of tools in the so-called ‘laboratory’ level.

These central, community-oriented services can be complemented by local or institutional activities, providing services for the complete research workflow in universities and research institutions. The Trier FuD system can be seen as a model for a local research environment, providing collaborative services for the complete Humanities research process: data collection and

analysis; preparation of publications; and publishing and archiving.

In summary, the development of a new culture of digital research within the Humanities requires a multi-faceted approach:

- Advocacy is needed to strengthen the acceptance of digital research, publications and the development of data.
- The character of research as a social activity requires fostering and support.
- A new academic recognition system must begin to recognise the scholarly value of electronic editions and publications; to review them in highly ranked journals; and to evaluate them as research contributions.

53. An example of this is the *CERL Thesaurus* and *Portal for Written and Printed Cultural Heritage*: <<http://www.cerl.org/web/>>.

## Preservation and Sustainability

The report *Reinventing research? Information practices in the Humanities* (2011)<sup>54</sup> aimed to establish how Humanities researchers “find and use information, and in particular how that has changed with the introduction of new technologies” (p. 6). It reflects on the stereotype of the Humanities researcher that characterises them as working alone on a small number of highly specialised topics, predominantly carrying out research in libraries and archives and rarely engaging with web-based technologies and content. Its findings include the following:

“In each of our case studies we found researchers working with new tools and technologies, in increasingly collaborative environments, and both producing and using information resources in diverse ways. There is richness and variety within humanities information practices which must be recognised and understood if we are to provide the right kind of support for researchers.” (p. 6)

An important consideration, in addition to the increasing uptake of RIs by researchers, is the scale of digital information that is being produced each year. Gantz and Reinsel have found that

“Between now and 2020, the amount of digital information created and replicated in the world will grow to an almost inconceivable 35 trillion gigabytes as all major forms of media – voice, TV, radio, print – complete the journey from analog to digital.”<sup>55</sup>

In such a context preservation and sustainability are a *sine qua non* of RIs. As ever increasing proportions of cultural heritage artefacts, whether text, image, sound or multimedia format, are digitised and made available, researchers need to be sure of the accessibility, authoritativeness, context, stability and longevity of such resources. This is also true of the analytical and interpretative tools, dynamic media such as mash-ups and web services, social networking, collaboration platforms and services such as grid and cloud computing that are increasingly being used by researchers. The issues of sustainability and accessibility are central also to the researchers who invest significant time and expertise into the development of such resources, as well as to

the funding bodies and organisations which make available the resources to support and enable such research to be carried out.<sup>56</sup> Nevertheless, *The New Renaissance*<sup>57</sup> report states that

“Today, only 22% of the cultural heritage institutions which digitise collections have long-term preservation plans in place. This means that the European investments in digitisation are in danger.” (7.1.2)

The issues of sustainability and preservation are among the most complex issues facing the creators, users, institutions, authorities and bodies that legislate for RIs.

*Key challenges* include:

- Maintenance and preservation of the objects that populate RIs (whether content or tools that scholars use to interrogate those objects), as well as RIs themselves following the completion of funding awards. Technological obsolescence of both hardware and software.<sup>58</sup>
- Emerging technological challenges relating to the preservation of objects created via complex digitisation techniques (for example, 3D<sup>59</sup> imaging, scanning

56. See, for instance, the final report of the Blue Ribbon Task Force: *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, National Science Foundation (NSF), Andrew W. Mellon Foundation, Library of Congress, Electronic Records Archives Program of the National Archives and Records Administration, Council on Library and Information Resources, US, Joint Information Systems Committee (JISC), UK, February 2010. Web (accessed 4 August 2011). <[http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)>. The same group submitted a Grand Challenge recommendation on “Sustainable Knowledge Infrastructure” to the U.S. Office of Science and Technology Policy “to ensure that the knowledge of today is available for use tomorrow, while fostering innovation for sustainable growth and creating high-quality jobs.” See <<http://brtf.sdsc.edu/>>.

57. *The New Renaissance, Report of the ‘Comité des Sages’, Reflection group on bringing Europe’s Cultural Heritage online*. European Commission, 2011. Web (accessed 14/07/2011).

58. “The foremost American authority on the longevity of various media, the National Institute of Standards and Technology (NIST), still cannot give a precise timeline for the deterioration of many of the formats we currently rely on to store precious digital records” (Cohen, Daniel and Rosenzweig 222). As far as scientific software development, including documentation, updating, maintenance and dissemination are concerned, the ESF Forward Look report *European Computational Science Forum: The ‘Lincei Initiative’: from computers to scientific excellence* (published in 2009 and available at <<http://www.esf.org/index.php?id=3011>>) argued for the need of a coordinated and sustained effort across national research funding agencies.

59. One of the challenges 3D brings is the sheer volume of data to be handled, but there are also very specific challenges concerning metadata of the artefacts, semantics of shape, digital provenance of the models and long-term preservation. Although some of these areas are also challenges to digital texts, images and videos, they become highlighted when working with 3D. Indeed, 3D models are normally produced by post-processing of the original scanned or image data. For instance, studies of surface detail of archaeological objects may clearly be influenced by the methods used in the original capture(s) and yet if these methods are documented with the object the study may still be fruitful. Whilst very similar considerations

54. Bulger, Monica, et al. *Reinventing research? Information practices in the humanities*. Research Information Network, UK, April 2011. Web (accessed 14/07/2011).

55. Gantz, John, and David Reinsel. *The Digital Universe Decade – Are You Ready?* IDC iView, May 2010: 1. Web (accessed 14/07/2011).

and multimedia digitisation) and dynamic, Web 2.0 content.<sup>60</sup>

- Reconciliation of the agendas of the various commercial, private and public bodies that participate in RI development.<sup>61</sup>
- Adherence to (not to mention development of) copyright law and other relevant legal issues.<sup>62</sup>

## CASE STUDY

Comité des Sages Reflection group on bringing Europe's Cultural Heritage online

### New Renaissance Report and its Implications

The key recommendations of the *New Renaissance* report (7.7.1) provide a useful diagnosis of the field:

1. "Preservation is a key aspect in digitisation efforts. Digital preservation is also a core problem for any born digital content. The organisational, legal, technical and financial dimensions of long term preservation of digitised and born digital material should be given due attention.
2. To guarantee the preservation of the European digital

apply to all digital assets, the complexities of acquiring 3D models from many materials means that there are currently many technical approaches being taken to cater for the physical properties of the objects. In addition, the necessity to detect and maintain the information about both the segmentation of a large 3D model and the inter-relation between its parts (e.g., of a statue or a building) stretches the performance of current systems in both representation of collections of artefacts and searching capabilities. At an even deeper level, shape embodies semantics that are easily recognisable to the human mind but very much more difficult to analyse from detailed surface models by an automated system. There is currently no agreed vocabulary of shape elements which would underpin the detection of such meanings from surface characteristics, although application such as face recognition and computer games interfaces may be leading the way in the case of gesture, for instance. 3D-COFORM ("Tools and Expertise for 3D Collection Formation" – <<http://www.3d-coform.eu/>>) is a large-scale integrating research project co-funded from the EC Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231809 (2008-2012). The objective is to improve 3D digitisation tools and processes to make 3D documentation a practical alternative for tangible cultural heritage. It builds on many of the challenges recorded in the EPOCH Research Agenda for the use of ICTs with tangible cultural heritage: <[http://public-repository.epoch-net.org/publications/RES\\_AGENDA/final\\_res.pdf](http://public-repository.epoch-net.org/publications/RES_AGENDA/final_res.pdf)>.

60. The POCOS project, funded by JISC in the UK, is an additional example of research into the preservation of complex objects: <<http://www.pocos.org/>>.

61. "Commercial content often functions only in combination with proprietary software, creating a major problem for archival institutions. Standardised and well documented file formats can be handled more easily than proprietary formats, but open documentation is not always in the interest of software companies." (*The New Renaissance* report, section 7.1.7.2)

62. For an overview of the European scenario see *The New Renaissance* report section 7.1.6; for an overview of the American legislation see especially *Our Cultural Commonwealth* 19-20.

cultural heritage, a copy of digitised or born digital cultural material should be archived at Europeana [<http://www.europeana.eu/>]. For in-copyright works the deposit site would be a dark archive functioning as a safe harbour.

3. To avoid duplication of effort by companies operating across borders and by the cultural institutions a system could be envisaged by which any material that now needs to be deposited in several countries would be deposited only once and then passed on to every cultural institution that has a right to it under its respective national deposit law.
4. Copyright and related legislation has to enable the cultural heritage institutions responsible for preservation to create archival copies and to make file conversions for archival purposes.
5. Persistent identifiers must be implemented in each digital object archived in cultural institutions. A reliable resolution service for persistent identification of digital objects must be developed and maintained on European level, preferably linked to Europeana."

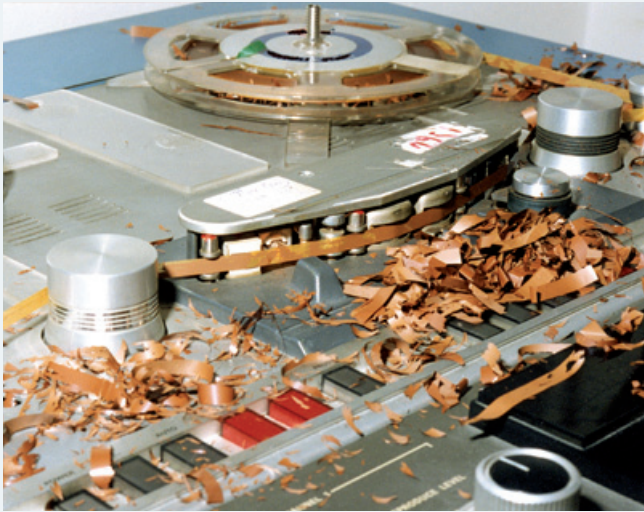
The observations made by the *Reinventing Research* report<sup>63</sup> amplify these issues with respect to the practical obstacles encountered by Humanities scholars:

"A majority of our scholars are not consistent in citing the digital resources they use, because of concerns about the legitimacy of online resources, and about disappearing links. There is also a lack of agreed citation standards for long-term persistent referencing and easy access. This is being addressed by the use of DOIs [Digital Objects Identifiers], but researchers are still not consistently using them. Building systems that include recommended citations and DOIs will help both scholars who are trying to follow a chain of citations, and those interested in understanding the impact of research."

The implementation of resolutions to these challenges will include the following measures:

1. More systematic research in the preservation and sustainability of complex, multimedia resources and dynamic content, with support from funding bodies.
2. New and more systematic collaborations with archivists, given the 35 trillion gigabytes of information that is projected to have been reached by 2020.
3. The fostering of cultures related to open access in the Humanities (the availability of an institutional repository does not inevitably foster such a culture among scholars to use it).

63. Bulger et al. 76.



**Figure 12**

A rare case of tape deterioration – but soon ALL audiovisual carriers, even those in perfect condition will be lost, as replay equipment will vanish

## CASE STUDY

Dietrich Schüller, Consultant to *Phonogrammarchiv, Austrian Academy of Sciences – AT; UNESCO, Working Group on Information Preservation*

### Audiovisual Research Documents

Audiovisual recordings are mentioned throughout this publication as important sources for many disciplines. Ever since their availability, they have been systematically produced for research purposes, specifically in the realm of Humanities. They are the primary sources for our present knowledge in wide parts of linguistics (ethno-linguistics, dialectology, endangered languages), ethno-musicology and folk music research, parts of social and cultural anthropology, contemporary history and political studies, oral and visual history, and several more. Their potential role for Digital Humanities, the need for their annotation and incorporation into databases, and their long-term retrievability from trusted repositories is convincingly displayed elsewhere in this study, directly and indirectly.

However, the prerequisite for all these visions is their mere physical existence in digital form, which for these documents is less trivial to organise than for most others. It is not so much the physical and chemical instability of audiovisual carriers that threatens their further availability: it is the rapid obsolescence of analogue and dedicated, proprietary digital audio and video formats since their replacement by ‘true’ computer file formats, and the frightening speed of disappearance of replay machines for these formats, including spare part supply and professional service. There is unanimous agreement that a time window of only 15 years – if at all – is left for replaying all pre-IT audio and video formats, in order to transfer their contents into digital repositories.

The methodology of safeguarding these documents by

extracting their contents from the originals and storing them as file formats in an IT environment by subsequent migration was developed for audio in the early 1990s: video archiving has been following on this trail since around 2000.<sup>64</sup> Successful implementation needs dedicated equipment, high-level expertise in vanishing technologies and their maintenance, and critical masses of several thousands of carriers in each format to make investment into equipment and experts viable.

Radio and television archives as well as wealthy national (audiovisual) archives and libraries<sup>65</sup> will – to sufficient extent – be able to solve signal extraction and its ingest into trusted repositories. However, as the EU-funded project *TAPE (Training for Audiovisual Preservation in Europe)*<sup>66</sup> has confirmed, many audiovisual collections held in small libraries, archives and museums are without any custodial care, let alone a realistic strategy for their safeguarding in the long term.<sup>67</sup>

In the context of this paper, more frightening even is the situation of audiovisual research materials. An estimated 80% of these primary sources forming the basis of our present knowledge in many disciplines are kept in research institutions of generally limited financial resources, which aim to optimise their research output rather than the preservation of their primary sources for re-study and dissemination to the scientific community. Often these materials still sit on the desks of the researchers that recorded them.<sup>68</sup>

64. Standards set by the International Association of Sound and Audiovisual Archives (IASA – <http://www.iasa-web.org/>): Schüller, Dietrich, Ed. *The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. Version 3 (=IASA-TC 03). International Association of Sound and Audiovisual Archives Technical Committee, 2005. Web (accessed 14/07/2011); and Bradley, Kevin, Ed. *Guidelines on the Production and Preservation of Digital Audio Objects*. Second edition (= IASA-TC 04). International Association of Sound and Audiovisual Archives Technical Committee, 2009. Web (accessed 14/07/2011).

A standard on Production and Preservation of Digital Video Objects (IASA-TC 06) is forthcoming.

65. For the typical, mainly professional materials stored by these archives, methods of fast transfer have been developed by the PrestoPRIME project family: <<http://www.prestoprime.org/>>. However, not all commercial practices are applicable for research materials.

66. See <<http://www.tape-online.net/>>.

67. Klijn, Edwin, and Yola de Lusenet. *Tracking the Reel World: A survey of audiovisual collections in Europe*. European Commission on Preservation and Access, January 2008. Web (accessed 14/07/2011). <[http://www.tape-online.net/docs/tracking\\_the\\_reel\\_world.pdf](http://www.tape-online.net/docs/tracking_the_reel_world.pdf)>.

68. Schüller, Dietrich. *Audiovisual research collections and their preservation*. European Commission on Preservation and Access, March 2008. Web (accessed 14/07/2011). <[http://www.tape-online.net/docs/audiovisual\\_research\\_collections.pdf](http://www.tape-online.net/docs/audiovisual_research_collections.pdf)>.

In view of the limited time available, this situation calls for a concerted approach on European and national levels. Apart from awareness-raising, as academia is generally not aware of the dramatic situation, the solution lies in the organisation of cooperative models and – because of the exceptional character of the threat – in the funding of transfer and annotation of contents into digital repositories. Traditional carrier-based materials can be successfully handled only by public and/or private competence centres. For present and future file-based documents, computer centres could take care of data preservation, while decentralised research institutions and individuals deliver data in accordance with protocols for files and metadata.<sup>69</sup>

Any failure to safeguard audiovisual primary source materials would severely infringe scholarly principles and trustworthiness, as neither the validation of secondary literature nor the re-examination under different and future research interests would be possible.

While CLARIN and DARIAH are only partly concerned with audiovisual records, the unusual urgency to counteract this otherwise foreseeable loss calls for the installation of a special programme within ESFRI aimed at their safeguarding.

## Evaluation of Digital Research and its Outputs

As a fundamentally new form of research in the Humanities, digitally driven research and the building of RIs bring a crucial change to both a century-long tradition of publication culture (in, for example, printed journals, monographs, editions, bibliographies, surveys, etc.) and to the way these published research outputs are evaluated.

Digital publications, both in the sense of born-digital and scholarly-driven, remediated digital resources, have still a way to go in terms of the recognition and consideration given to them by traditional scholarly communities. This is also true of the assessment of individual research and career progression. Thus, there is a vital need, especially regarding the rights of young researchers, to rethink acknowledgement and reward for digital scholarship. Symptomatic of this is that some

69. An effective model has been developed at the Max Planck Institute for Psycholinguistics for DoBeS (Dokumentation bedrohter Sprachen; <<http://www.mpi.nl/DOBES>>). However, because of the precarious situation of replay machines and their constant need for professional maintenance, such models are hardly applicable to traditional carrier-based originals.



Figure 13.  
NINES home page

online research indices do not accord the same status to the research products of Digital Humanities as they do to traditional Humanities. Neither do they accord digital scholarship parity of assessment with traditional monothematic disciplines. Whether they are seen as two sides of the same coin or not, both digital resources and digital scholarship deserve to be assessed using fair, open and comparative criteria. Accordingly, the following points are crucial for the further acceptance of digital driven research.

1. It is necessary to secure and support the evolution of a new kind of research culture that accepts the process character of digital publications in their various modes of existence (e.g., digital editions, research databases, digital archaeological surveys, conceptual models, etc.). This should be given highest priority.
2. Furthermore, and accompanying this process of recognition, it is necessary to establish clearing and authoritative mechanisms (including peer review)<sup>70</sup> to secure scholarly reliability and recognition of research across international and interdisciplinary collaborations.
3. Another issue is that of building interdisciplinary tools and teams where all contributing specialist roles and competences are recognised and rewarded.
4. In order to change attitudes towards digital research, advocacy is of great importance, e.g., for emerging cross-disciplinary fields and relevant academic recognition, and identifying authorities, best practice examples and ambassadors in the field.

70. Important work already undertaken in this area includes the *Guidelines for Evaluating Work with Digital Media in the Modern Languages*, Modern Language Association Committee on Information Technology, 2002. Web (accessed 14/07/2011); and chapter 9 of the *European Peer Review Guide. Integrating Policies and Practices into Coherent Procedures*. European Science Foundation Member Organisation Forum, March 2011. Web (accessed 14/07/2011).



5. Last but not least, improved and agreed systems for an evaluation of RIs and in general of alternative research outputs should be put in place on an international level, respecting the heterogeneity and the multidimensionality of research in the Humanities. Regarding RIs, such systems should be grounded on the definition of expectations associated to their development as well as on the economic and social benefits they bring about.

## CASE STUDY

Dana Wheelles, *NINES*, University of Virginia – USA

### Communities of Practice: NINES Peer Review <<http://www.nines.org/>>

*NINES (Networked Infrastructure for Nineteenth-Century Electronic Scholarship)* is a scholarly organisation devoted to forging links between the material archive of the 19<sup>th</sup> century and the digital research environment of the 21<sup>st</sup>. Since 2005, one of our primary goals has been to serve as a peer-reviewing body for digital work in the (long) 19<sup>th</sup> century, with an emphasis on British and American literature. The heads of three Editorial Boards (Americanist, Romantic and Victorian) manage the review of our submissions, gathering evaluation materials and reader reports unique to every digital project we evaluate.<sup>71</sup> Projects that have been vetted and accepted into NINES are then added to our own site's search index via a common metadata format (in our case, RDF).<sup>72</sup> Both as an organisation and a website, NINES works to direct attention to excellent scholarly work online, and to ensure that intellectual labour involved in it is rewarded.

To submit a site for peer review, the project lead must fill out a form requesting information about the site in question, its institutional affiliations and its technical infrastructure.<sup>73</sup> As an aggregator – that is, a site that brings a multitude of other sites into conversation with each other independent of their own locations on the web – NINES requires a measure of stability and sustainability from its partner sites. Peer review by NINES cannot be a single, isolated event, but rather the beginning of an ongoing conversation between our organisation and the projects themselves. As more material is added to any given site, NINES requires another batch of metadata about the new objects to stay up to date, and, in the event of a site's significant expansion or re-structuring, NINES may determine that it is necessary to examine the resource anew.

71. <<http://www.nines.org/about/scholarship/boards.html>>.

72. <<http://www.nines.org/search>>.

73. <<http://www.nines.org/about/scholarship/rdf.html>>.

Evaluation of digital projects necessitates the recruitment of two (if not more) reviewers: one as an expert in the topic to review the content, and another capable of assessing the technical aspects of the resource. In the years since NINES was founded, more and more scholars have developed skills in text encoding, database structure and interface design, a fortunate situation that allows for more insightful and useful reader reports. However, getting usability feedback from traditional scholars can be invaluable, especially if they make up a large portion of the intended audience. We offer a basic set of General Guidelines and Peer Review Criteria for NINES Content<sup>74</sup>, and frequently refer to the MLA's Guidelines for Editors of Scholarly Editions<sup>75</sup>.

As members of the NINES federation of projects, contributors benefit from a community of users dedicated to scholarly work online. Partnership with our organisation can also provide an extra level of stability for projects dependent upon grant money: it makes a stronger case for a project's relevance to the field and to a wider audience and demonstrates a support system that single, isolated sites cannot manage. Peer review by NINES is not simply a stamp of approval; it is an agreement to a long-lasting and dynamic relationship with the project going forth.

## CASE STUDY

Gudrun Gersmann (*German Historical Institute, Paris*),  
Lilian Landes (*Bavarian State Library, Center for Electronic Publishing*) – DE/FR

### The Writings of Others: Review Culture at a Turning Point

The project *recensio.net*<sup>76</sup> emerged from the thought that reviews exclusively released in print can nowadays be considered as anachronisms. The speed of publication plays a much bigger role in reviewing than it does with regard to articles and monographs. In the course of traditional publishing, often years go by before a reviewer is found, the text is written, edited, typeset, printed and delivered.

The situation is furthermore aggravated by the fact that the young generation of scholars is more and more focusing on online research. Articles and reviews in print are struggling for perception at a higher rate than they used to.

In this respect, *recensio.net* responds to a request which has most recently been more and more expressed

74. The guidelines can be downloaded at: <<http://www.nines.org/about/scholarship/9s-guidelines.doc>>.

75. <[http://www.mla.org/resources/documents/rep\\_scholarly/cse\\_guidelines](http://www.mla.org/resources/documents/rep_scholarly/cse_guidelines)>.

76. <<http://www.recensio.net/>>.

by scholars: It is true that online review journals such as the *sehempunkte* were able to successfully establish themselves within the scholarly communities over the last couple of years. It is also true that many small journals also publish their printed reviews additionally on their website. But together all this currently leads to maximum confusion. Readers (who at the same time are scholars, reviewers and reviewees) can't cope with this scattering of reviews all over the place. A further problem lies within the nature of historical research, which is still based around a national perspective, and rarely takes notice of new publications beyond its own national boundaries.

The internet offers the potential to curb these difficulties, and also to comply with the recent, profound transformation of scholarly culture from result- to process-orientation, which in the medium term will also establish itself within the Humanities.

As part of a pilot project (supported by the Deutsche Forschungsgemeinschaft – DFG), the Bavarian State Library (BSB), the Institute for European History (Mainz) and the German Historical Institute (Paris) have developed *recensio.net* as an online platform which on the one hand gathers 'classic' reviews from journals, and on the other hand acts as a testing medium for new, collaborative forms of academic reviewing. The goal of this 'two-pillar-concept' is to meet the requirements of both the established and the younger generation of scholars.

The first pillar is focused on the cooperation with journals whose reviews are currently published in print or online. Towards these, *recensio.net* acts as a service provider, gathering the reviews in Open Access on one platform, making them searchable in full text and tagging them with metadata, in order to make them accessible via content browsing. The cooperating editorial offices still work independently; the platform perceives itself as an Open Access aggregator.

The second pillar complies with our belief that Web 2.0 elements and a faster, more particular form of writing on writings will find its place next to the classic scholarly review: authors of monographs and articles can present the core statements of their research. Platform users can leave comments and thus give their view on the publication in general, or relate to certain aspects. With regard to the busy daily routines at university, this approach might lower the inhibition to become involved in reviewing, and, at the same time, it is supposed to facilitate interdisciplinary views on new publications. A sociologist, for instance, would hardly write a 'classic' review of a historical book, but a comment on a section regarding his area of expertise might indeed result in a productive discussion which can take place directly on *recensio.net*, as the author of a presentation is informed about incoming comments.

A further essential point is the cross-border approach of the platform: making scholarly data from different language areas available is a decisive element. *recensio.net* has three navigational languages (English, German, French). Reviews and presentations can be written in all European languages; the search function includes a language filter.

Linking special platforms with common scholarly research instruments is indispensable in the time of cross-linked working: therefore, in the course of 2011, all review objects will be linked to the entry of the reviewed title within the library catalogue of the BSB.

Time will tell if European historians are ready for a 'monitored Web 2.0'. There is a surprisingly great interest on behalf of the journals in supplying their reviews. Only a few weeks following the launch of *recensio.net*, that much can be said for sure.

---

## Communities of Practice

This section aims to provide some orientation for communities who aim to develop an RI and how they might go about it. So too it aims to give some guidance to established, national-level communities who may wish to interlink with other communities, both at the national and pan-European level. Here the aim is to draw attention to some established communities of practice<sup>77</sup> as well as RIs and resources that either intersect with or underpin the domain of RIs in the Humanities. The objective is also to disseminate some existing models of good community of practice that emerging initiatives may consider appropriate to model themselves on. This section is not intended to be exhaustive, nor can it be. Likewise, some examples of good community of practice are chosen without implying that these are the only examples of good community of practice that exist.

### Pan-European infrastructures

The *HERA Survey of Infrastructural Research Facilities and Practices for the Humanities in Europe* (2006) was

---

77. Etienne Wenger has discussed and defined communities of practice in the following way: "Communities of practice are formed by people who engage in a process of collective learning in a shared domain of human endeavor: a tribe learning to survive, a band of artists seeking new forms of expression, a group of engineers working on similar problems, a clique of pupils defining their identity in the school, a network of surgeons exploring novel techniques, a gathering of first-time managers helping each other cope. In a nutshell: Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly". Wenger, Etienne. *Communities of practice: a brief introduction*. June 2006. Web (accessed 14/07/2011). <<http://www.ewenger.com/theory/>>.

based on a consultation of cultural and research institutions in thirteen European countries. It provides an overview of the shape and scope of the European RI landscape available to Humanities researchers in 2006.<sup>78</sup> While reflecting on the potential of Humanities RIs, it also identifies a range of issues that need further attention if RIs are to realise their potential, for example, in order to combat the linguistic and scientific fragmentation that exists at the pan-European level. From the outset, it is argued that

“collaboration at the European level is needed to secure cooperation of parallel initiatives, to improve standards and methods and to develop joint and more cost-effective solutions for research infrastructures.” (p. 1)

The issue of fragmentation was acknowledged to be of continuing high importance by the participants of the *Strategic Workshop on Research Communities and Research Infrastructures in the Humanities* (Strasbourg 2010) organised by the ESF Standing Committee for the Humanities.<sup>79</sup> The publication of the *European Strategy Forum on Research Infrastructures* (ESFRI) roadmap has taken an important first step towards solving such problems. Nevertheless, only two Humanities RIs have been funded through this initiative as described below.

CLARIN<sup>80</sup>, *The Common Language Resources and Technology Infrastructure*, has partners from numerous European countries. Its target audience is linguists, but it is of relevance to all the disciplines of the Humanities and Social Sciences that study language. It aims to provide an RI that offers a single point access to data, tools, resources and human expertise pertaining to language:

“the user will have access to guidance and advice through distributed knowledge centres, and via a single sign-on the user will have access to repositories of data with standardised descriptions, processing tools ready to operate on standardised data, and all of this will be available on the internet using a service oriented architecture based on secure grid technologies.”<sup>81</sup>

DARIAH, *Digital Research Infrastructures for the Arts and Humanities*, was also included in the first ESFRI

roadmap (2008) and has partners from various European countries. It works with researchers, information professionals and the cultural heritage sector in order to:

- “Explore and apply ICT-based methods and tools to enable new research questions to be asked and old questions to be posed in new ways
- Improve research opportunities and outcomes through linking distributed digital source materials of many kinds
- Exchange knowledge, expertise, methodologies and practices across domains and disciplines”<sup>82</sup>

As part of their deliverables, both CLARIN and DARIAH publish a wealth of material relevant to both emerging and established communities of practice. The material includes technical guides, reports and newsletters.<sup>83</sup> Their websites also provide a mechanism through which institutions may apply to participate in them, in various capacities.

These pan-European RIs interlink with the global perspective through projects such as CHAIN

“a forum for discussion, with a very light-weight organisational structure, with fluid membership and boundaries, no budget, and meeting only when necessary. CHAIN participants have resolved to work together on advocacy for improved infrastructure, and on aligning our infrastructure initiatives to allow the maximum interoperability of services”<sup>84</sup>

and CenterNet:

“an international network of Digital Humanities centers formed for cooperative and collaborative action to benefit Digital Humanities and allied fields in general, and centers as humanities cyberinfrastructure in particular.”<sup>85</sup>

78. See <<http://heranet.info/hera-era-net-2004-2009-workplan>>.

79. For the programme and workshop presentations, see <<http://www.esf.org/research-areas/humanities/strategic-activities/research-infrastructures-in-the-humanities.html>>.

80. Both CLARIN and DARIAH have recently completed their preparatory phases, for further information see: <<http://www.clarin.eu/executive-summary>> and *DARIAH Newsletter* 8 (Spring 2011): 1-3. Web (accessed 14/07/2011).

81. <<http://www.clarin.eu/external/index.php?page=about-clarin&sub=0>>.

82. <<http://www.dariah.eu/>>.

83. See the respective web pages at <<http://www.clarin.eu/external/index.php>> and <[http://www.dariah.eu/index.php?option=com\\_docman&Itemid=200](http://www.dariah.eu/index.php?option=com_docman&Itemid=200)>.

84. <<http://www.arts-humanities.net/chain>>.

85. <<http://digitalhumanities.org/centernet/about/>>.

## CASE STUDY

### Internationally Recognised Good Practices, Guidelines and Standards: Some Examples

#### Text Encoding Initiative

The Text Encoding Initiative (TEI)<sup>86</sup> publishes internationally recognised XML-based guidelines. It is “a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics”. TEI is endorsed by agencies such as the NEH, AHRC and the EU’s Expert Advisory Group for Language Engineering. Its impact has extended beyond the Humanities. One of the most influential projects in Digital Humanities, it has had a formative influence on the development of XML, the *lingua franca* of global data-interchange. TEI facilitates community involvement in a number of ways: both its Board and Council are staffed by library and cultural heritage professionals; information professionals; academics from across the Humanities; administrators; they are elected by the membership of the consortium. Through its ‘Special Interest Groups’ it provides members with “a forum for people working in a particular area, or with a specific set of concerns, to exchange opinions and build consensus”.<sup>87</sup>

#### CIDOC-CRM

The CIDOC *Conceptual Reference Model* is an official ISO standard that is developed by the International Council of Museums. It sets out “definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation”.<sup>88</sup>

In addition to the standards discussed above, a broad overview of the standards currently available across the cultural heritage sector has been given by Riley.<sup>89</sup>

#### Organisations

The *Alliance of Digital Humanities Organisations* (ADHO) “promotes and supports digital research and teaching across all arts and humanities disciplines, acting as a community-based advisory force, and supporting excellence in research, publication, collaboration and training”.<sup>90</sup> It supports a wide range of initiatives, such as publication and training. It fur-

thermore coordinates the activities of three organisations: the *Association for Literary and Linguistic Computing*<sup>91</sup> (ALLC), the *Association for Computers and the Humanities*<sup>92</sup> (ACH), and the *Society for Digital Humanities/Société pour l'étude des médias interactifs*.<sup>93</sup>

#### Peer reviewed journals include:

- *LLC, The Journal of Digital Scholarship in the Humanities* published by Oxford University Press, “an international journal which publishes material on all aspects of computing and information technology applied to literature and language research and teaching”.<sup>94</sup>
- *DHQ, Digital Humanities Quarterly*, “an open-access, peer-reviewed, digital journal covering all aspects of digital media in the humanities”.<sup>95</sup>
- *Digital Studies/Le champ numérique*, “a refereed academic journal, publishing three times a year and serving as a formal arena for scholarly activity and as an academic resource for researchers in the Digital Humanities”.<sup>96</sup>
- *Informatica Umanistica*, a venue to reflect on the connections between computer sciences and humanities. It aims at reflecting on the methodologies of these disciplines and on the new approaches that emerge when the two domains interact.<sup>97</sup>

#### Key initiatives include:

- *Humanist*: “an international online seminar on humanities computing and the Digital Humanities. Its primary aim is to provide a forum for discussion of intellectual, scholarly, pedagogical, and social issues and for exchange of information among participants”.<sup>98</sup>
- The ACH and the Chronicle of Higher Education’s ProfHacker sponsored *Digital Humanities Questions & Answers*.<sup>99</sup>
- The annual, international conference *Digital Humanities* sponsored by ADHO.
- *Digital Humanities Now* “a real-time, crowdsourced publication. It takes the pulse of the Digital Humanities community and tries to discern what articles, blog posts, projects, tools, collections, and announcements are worthy of greater attention”.<sup>100</sup>

86. <<http://www.tei-c.org/>>.

87. <<http://www.tei-c.org/Activities/SIG/rules.xml>>.

88. <<http://www.cidoc-crm.org/>>.

89. Riley, Jenn. *Seeing Standards: A Visualization of the Metadata Universe*. Indiana University Libraries White Professional Development Award, 2009-2010. Web (accessed 14/07/2011). <<http://www.dlib.indiana.edu/~jenlrile/metadatamap/>>.

90. <<http://digitalhumanities.org/>>.

91. <<http://www.allc.org/>>.

92. <<http://www.ach.org/>>.

93. <<http://www.sdh-semi.org/>>.

94. <<http://llc.oxfordjournals.org/>>.

95. <<http://digitalhumanities.org/dhq/>>.

96. <[http://www.digitalstudies.org/ojs/index.php/digital\\_studies/index](http://www.digitalstudies.org/ojs/index.php/digital_studies/index)>.

97. <<http://www.ledonline.it/informatica-umanistica/>>.

98. <<http://www.digitalhumanities.org/humanist/>>.

99. <<http://digitalhumanities.org/answers/>>.

100. <<http://digitalhumanitiesnow.org/about/>>.

## CASE STUDY

David J. Bodenhamer (*Indiana University Purdue University Indianapolis*), Ian N. Gregory (*Lancaster University*), Andreas Kunz (*Institute of European History, Mainz*) – US/UK/DE

## Geographical Information Systems and Spatial Humanities

*Geographical Information Systems* (GIS) software combines a database management system (DBMS) with a computer mapping system. The conventional DBMS provides data on *what* resources there are but is unable to say *where* they were located. The GIS also stores a location for every data element in the form of a point, line, polygon (which represents a zone or area) or a pixel. The resulting structure incorporates information on *what*, *where* and potentially *when*. Rather than a mere mapping system, GIS is in fact a database management system explicitly designed to store, manipulate, visualise and analyse data that incorporates a spatial reference.

Since the 1990s, there has been a growing interest in how GIS – a technology that emerged from the Earth sciences, computer science, and the military – can be used by historians and others who research the past.

This field has become known as Historical GIS, a field that now has a large and growing literature.<sup>101</sup> Much early Historical GIS research was associated with the use of quantitative or cartographic sources. More recently, there has been an increasing interest in how other sources, particularly texts, can be incorporated into GIS and used across the full range of Humanities disciplines, including archaeology, literary studies, linguistics, classics and religious studies. There also has been a re-awakening of interest in the importance of space in understanding the past, a development known as ‘the spatial turn’. As a consequence, what started as Historical GIS has broadened to become *Spatial Humanities*,<sup>102</sup> a field in which spatial technologies can be used to enhance our understanding of the geogra-

101. See, for example: Knowles, Anne Kelly, and Amy Hillier, Eds. *Placing History: How maps, spatial data and GIS are changing the practice of history*. Redlands, CA: ESRI Press, 2008. Print.; Gregory, Ian N., and Paul S. Ell. *Historical GIS: Technologies, methodologies and scholarship*. Cambridge studies in historical geography. Vol. 39. Cambridge University Press: Cambridge, 2007. Print.; Gregory, Ian N. *A Place in History: A guide to using GIS in historical research*. Oxbow: Oxford, 2003. Print.; Knowles, Anne Kelly, Ed. *Past Time, Past Place: GIS for history*. Redlands, CA: ESRI Press, 2002. Print. See also the *Historical GIS Research Network*: <<http://www.hgis.org.uk>>.

102. Bodenhamer, David J., John Corrigan, and Trevor M. Harris, Eds. *The Spatial Humanities: GIS and the future of humanities scholarship*. Bloomington: Indiana University Press, 2010. Print.

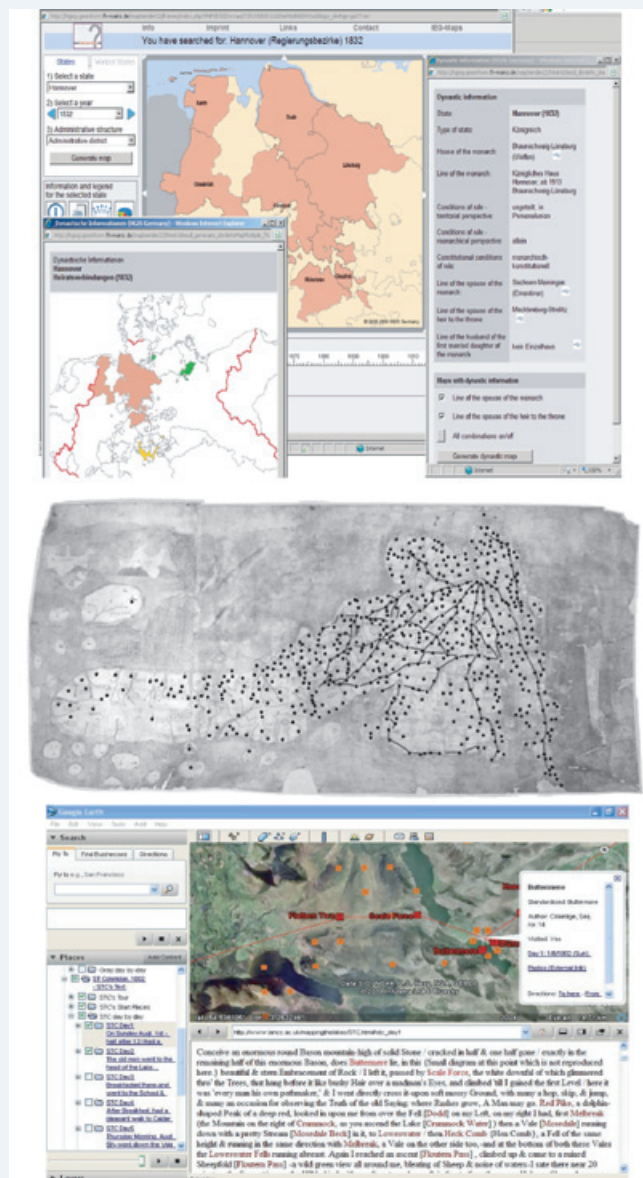


Figure 14.

Examples of GIS in the Humanities:

**Top** – Nineteenth century German dynasties using HGIS Germany (<http://www.hgis-germany.de>);

**Middle** – A GIS of the Gough Map of Medieval Britain ([http://www.qub.ac.uk/urban\\_mapping/gough\\_map/](http://www.qub.ac.uk/urban_mapping/gough_map/));

**Bottom** – The Mapping the Lakes GIS of Lake District literature (<http://www.lancs.ac.uk/mappingthelakes>). The image of the Gough Map is reproduced by kind permission of The Bodleian Libraries, University of Oxford [MS. Gough Gen. Top. 16 (Gough Map)]. Our thanks also to Dr. K. Lilley (Queen's University Belfast) for assisting with this.

phies inherent in the problems confronted by all Humanities disciplines and to use the wide range of evidence employed by humanists, including cartographic, textual, image-based, statistical and multimedia formats.

Developing the Spatial Humanities requires inter-disciplinary collaboration in at least two ways: Humanities scholars must collaborate with technical experts from fields such as geographical information science, library science, and computer science; and they must collaborate across disciplines within the Spatial Humanities. Researchers who wish to create and analyse GIS databases on topics as diverse as Ming dynasty China, ancient Rome, Civil War America or 20th century Europe face similar challenges and can learn much from each other.

To improve research collaboration and develop the Spatial Humanities requires action in four key areas:

- data resources
- training
- methodological development
- applied scholarship.

The time and expense of creating GIS databases is often a major barrier to entry to the field. Significant investment has been made to create national and regional Historical GIS databases, most of which incorporate administrative boundaries as they change over time. Gazetteers also exist to provide coordinates for each place-name within a source. But in each instance, the need for enhancement exists, especially to address problems such as temporal change, multiple names and different languages. Methods for extracting place-names from texts, especially unstructured texts, also need to be developed to speed the use of this commonly used evidence within a GIS. Finally, it is important that scholars have access to a comprehensive inventory of the wide range of existing GIS databases so these resources can be publicised, disseminated, preserved and integrated and re-purposed as required. The most effective use of these resources, however, will require improved metadata standards to accommodate the record's location in time and space.

The advent of Google Earth has meant that the simplest geo-technologies are now extremely intuitive and easy to use; however, more powerful GIS software is still relatively complicated. Further software training should be made available, especially to postgraduates and junior academics. Beyond this, however, there is a relative lack of awareness of what GIS and spatial approaches have to offer to the Humanities. Workshops, expert meetings, networks and conferences are important for developing this understanding. While much potential exists for the analysis of GIS data over time and space, significant methodological developments

will be needed to ensure the best use of Humanities data within a GIS. Three issues can be identified here:

- better techniques for the quantitative analysis of spatio-temporal data;
- enhanced methods for the spatial analysis of qualitative data;
- and improved visualisation of complex spatio-temporal material, especially for uncertainty or ambiguity.

Despite the challenges, GIS professionals and humanists who use geo-spatial technologies already are turning advances in data development and methodologies into new knowledge about Europe's societies, economies and environments. It is important that research remains focused on these areas in the long term. The approach to doing this will be to conduct and encourage exemplar research and to ensure conferences, workshops, and seminars are available to allow the approaches and results of this research to be disseminated widely.

#### CASE STUDY

Ewa Dahlig-Turek, *Institute of Arts, Polish Academy of Sciences and Humanities* – PL

### Between Research and Public Use: Music-Related Research Infrastructures

Due to its subject of study, which is an acoustic phenomenon of music and its cultural, social, cognitive and many other aspects, musicology needs RIs able to deal with various types of data such as:

1. music audio recording;
2. music text (i.e., written music);
3. music-related text (i.e., written information on music);
4. image (e.g., historical iconographic sources important in the study of musical instruments);
5. video recording (e.g., musical performance and its context).

Available Information Technology solutions cover all these areas, and music-related RIs usually combine a few types of data.

Among RIs typical of musicology, digital music editions are very important. They widely replace conventional music scores as they offer incomparably more flexible and efficient ways of accessing musical information.

Examples of these are:



- ECOLM – *Electronic Corpus of Lute Music*<sup>103</sup>; Department of Music, King’s College, London; a repository of full-text encodings of music sources for the Western-European lute, with graphical images from manuscripts.
- CMME – *Computerised Mensural Music Editing Project*<sup>104</sup>; Utrecht University; an online corpus of electronic editions with accompanying software tools.
- OCVE – *Online Chopin Variorum Edition*<sup>105</sup>; Danish Royal Library and Royal Holloway, University of London; a new type of ‘dynamic edition’ allowing the users to construct a unique edition of their own.

To store and analyse music, musicologists need tools appropriate to one of the forms of music representation, i.e., either music notation or recorded sound. The widely used so-called ‘Western’ music notation requires translation into machine-readable codes, variety of which has been developed. The codes range from very simple to very complex. Examples include:

- The Humdrum Toolkit developed by David Huron (USA)<sup>106</sup>; more complex, and therefore more versatile and universal than EsAC (the Essener Assoziativ-Code developed by Helmut Schaffrath, Germany) which, however, is still popular as the easiest encoding method for most of the European monophonic music.
- The music ‘esperanto’ of today is MusicXML<sup>107</sup> – the format used for interactive music publishing online.

Musical codes allow building large digital corpora of European folk music, used either for individual research, or publicly available online (e.g., Meertens Institute in

Amsterdam).<sup>108</sup> Encoded music collections form a basis for the studies belonging to a broad category of Music Information Retrieval (with annual conferences of the *International Society for Music Information Retrieval*),<sup>109</sup> and also for Cognitive Modelling.

While due to the specificity of the encoding languages the use of such RIs is limited to specialists, RIs based on recorded sound are available for the broad public as well. The invention of the phonograph (1877) marked the beginning of the era of music archives established at different scholarly centres, many of which contain truly unique material. Due to independent methods of data archiving incompatible for networking purposes, archives’ content has remained underexposed for decades. Today music archives are an active group of musicological institutions cooperating to create publicly available repositories which, however, meet requirements of research communities in terms of their contents, quality and metadata structure, and thus play a double role of public collections and RIs. Examples include:

- DISMARC (*DIScovering Music ARChives*, 2006-2008)<sup>110</sup> is a European initiative to build a network of archives sharing their metadata and sound recordings in the internet. Instead of imposing a common structure, DISMARC offers an effective system that allows data to be imported in its native format and subsequently mapped to DISMARC’s general scheme. Thus, with 30 languages implemented, it offers a possibility of simultaneous searching in the user’s native language, and finding the output in all the accessible data.
- *EuropeanaConnect* (2009-2011)<sup>111</sup> applies all the DISMARC solutions to provide music audio contents for *Europeana*.

103. See <<http://www.ecolm.org/>>.

104. See <<http://www.cmme.org/>>.

105. See <<http://www.ocve.org.uk/>>.

106. See <<http://www.musiccog.ohio-state.edu/Humdrum/>>.

107. See <<http://www.recordare.com/musicxml/>>.

108. See <<http://www.liederenbank.nl/>>.

109. See <<http://www.ismir.net/>>.

110. See <<http://www.dismarc.org/>>.

111. See <<http://www.europeanaconnect.eu/>>.

## Cultural and Linguistic Variety – Transnational RIs

A key motivation for this paper is the belief that good RIs are needed as strongly in the Humanities as in the so-called hard sciences. In both cases, good RIs allow scholars to access, order, analyse, store and reuse data in ways otherwise impossible. Therefore, they pave the way for innovative research (even if they are not the only route to this).

In many cases, however, digital RIs were developed earlier in the hard sciences than in the Humanities and currently receive a larger proportion of funding. One important factor is that the objects of study – an author, a society, etc. – are more likely to be culturally bound in the Humanities, usually studied through the written (often hand-written) word, which calls for time-consuming transcription. As a consequence, scholars in the Humanities tend towards more qualitative than quantitative work, analysed, published and disseminated in specific national languages.

It is, however, obvious that Humanities scholars have an interest in asking questions that are not confined to specific cultural areas. Indeed, the Humanities have a responsibility and a unique position for answering more general questions about human development, such as how languages function and how societies have developed over time. Theoretically informed, comparative and transnational research gains from access to large datasets that are not all from the same culture and are multilingual. Being able to access data from various linguistic areas is thus crucial.

To make this possible, the construction of RIs involves sensitivity to the need for many-level translations (translation taken in the broad sense of the word). This may include the transcription, digitisation and encoding of handwritten texts in order to make them machine-readable, the further addition of metadata and, often, English translations, especially for data retrieved from sources in so-called lesser-used languages. A good example of the need to establish ‘translation codes’ is the so-called HISCO-system<sup>112</sup>, developed in the Netherlands, with the purpose of making historical labour data comprehensible and comparable.

112. See <<http://www.iisg.nl/research/hisco.php>>.

### CASE STUDY

Gunnar Sivertsen (*Nordic Institute for Studies in Innovation, Research and Education*), Istvan Kenesei (*Research Institute for Linguistics, Hungarian Academy of Science*), Nigel Vincent (*British Academy*), Milena Zic-Fuchs (*ESF Standing Committee for the Humanities*) and Sir Roderick Floud (*ESF Standing Committee for the Social Sciences*) – NO/HU/UK/ESF

### Towards Comprehensive Bibliographic Coverage of the Scholarly Literatures in the Humanities and Social Sciences

In reaction to the inadequacy of existing bibliographic/bibliometric indices, the development of a *European Reference Index for the Humanities* (ERIH)<sup>113</sup> has been supported by various ESF Member Organisations (MOs) with the aim of making visible and accessible the world-class research published by Humanities researchers in the European languages. ESF Standing Committee for the Humanities took responsibility for ERIH in 2002.

### For a better developed Research Infrastructure for scholarly communication and publishing

Tools for searching and accessing the research literature are an important part of the research infrastructure in the Humanities and Social Sciences as they are in other research domains. Access to the journal articles themselves, or at least their bibliographical metadata, their abstracts and information about the authors, responds to the need of every scholar to search for relevant literature. Relevant solutions for a solid bibliographic coverage of the Humanities and Social Sciences on the international level already exist in several instances at the national level. However, integration and standardisation of procedures and data sources at the international level are missing.

Following a discussion amongst representatives of ESF MOs (2010) on ERIH, its results so far, and the MOs’ expectations with regard to its future, a working group was set up to provide a short internal report with recommendations.

In this internal report released to ESF MOs in September 2011, the group made a number of recommendations especially as far as the framework of a better research infrastructure is concerned. First of all, organisations representing research in the Humanities and Social Sciences

113. See <<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities.html>>. The first phase of the project was completed with the publication in late autumn 2007 and in early 2008 of 14 ERIH Initial Lists, covering academic journals. The ERIH Revised Lists published in 2011 are the result of the second round of the ERIH project, following the revisions of the ERIH Initial Lists.





**Figure 15.**  
The threads of *European History Online* transcend national and disciplinary perspectives

are encouraged to define criteria and standards that the library information systems will have to meet in order to be regarded as a well-functioning part of the research infrastructure for scholarly communication and publishing in an age of internationalisation and electronic information.

### The 'Bibliographic Approach'

The working group considers the original aim of ERIH – increasing the visibility of the Humanities by documenting, reviewing and developing the infrastructure for scholarly communication and publishing – as still valid and very important, and suggests further steps to be taken in this direction along a path which is referred to as the 'bibliographic approach' to create a resource for scholars and potentially improve research assessment.

In opposition to some of the current limits of the ERIH lists, this approach takes into consideration the need to build a robust system and a suitably consistent and reliable methodology to cover scholarly publications not only in *journals* but also in *books* and in both the *Humanities and the Social Sciences* with searchable bibliographic references to the *publications themselves*, not only their publication channels.

The aim of the bibliographical approach is thus to make all the scholarly literatures – in print or electronically distributed – searchable and accessible across countries. The

bibliographic references would thereby mirror as fully as possible the output of European research in the Humanities and Social Sciences and show what diversity of European research actually means. This diversity would be reflected not only in national languages but also in national traditions within specific disciplines. Visibility and availability would be gained at the same time.

The idea of the bibliographic approach will first need to be discussed at the level of research funders. In addition, relevant partners and stakeholders for collaboration – for example major academic libraries and their information systems as well as the main commercial suppliers to such systems – will be identified and involved in any eventual future developments.

The problem is not only a problem of linguistic variety, but rather one of different ontologies and taxonomies. The world is categorised and made meaningful in different ways in different cultures. Conceptual clarity and, as a consequence, reliable interpretation and comparison of cultural products presupposes high awareness of problems of varying taxonomies. Otherwise, the promises of more adventurous projects and higher quality in results are compromised by insufficient translations on

many levels as the cases below exemplify. Such questions have been tackled by Humanities computing researchers for many years. While numerous elegant strategies have already been developed, for example, by the *Text Encoding Initiative* (TEI, discussed above), such issues remain important research questions in the field.

It is important to combine an ambitious approach to these matters with a pragmatic one. While the complete digitisation and translation of the cultural heritage is unrealistic for the foreseeable future, this does not mean that we should not endeavour to develop new RIs in the Humanities that reach across linguistic borders.

#### Example 1

### One year – but how many seasons?

In most European languages, the year is divided into four parts, reflecting the annual rhythm of agriculture. But in cultures where agriculture is not the basis of living, the year is subdivided differently, reflecting other work rhythms. The Sámi rein herding culture of northern Europe provides an example of this.

English	German	Sámi	Time period
Spring	Frühling	Gidá	(c. 25 April to 25 May)
		Gidágiesse	(c. 25 May to Midsummer)
Summer	Sommer	Giesse	(c. Midsummer to 24 August)
		Tjaktjagiesse	(c. 24 August to Mid-September)
Autumn	Herbst	Tjaktja	(c. Mid-September to 14-15 October)
		Tjaktjadálve	(c. 14-15 October to Christmas time)
Winter	Winter	Dálve	(Christmas-time to end of February)
		Gidádálve	(beginning of March to c. 25 April)

#### Example 2

### Similar occupation – but varying contexts and connotations

Depending on differing social realities and historical trajectories, the words that designate approximately the same phenomenon may nevertheless evoke different ideas. This may, in its turn, lead to confusion or a false sense of agreement on what is being compared.

Some of the innumerable words designating a ‘person who makes a living at least partly through agriculture’ will illustrate this.

Swedish 18 <sup>th</sup> -century <i>bonde</i>	word evokes notions such as ‘taxpayer’, ‘land-holder’, sometimes also ‘holder of political rights’
English 18 <sup>th</sup> -century <i>farmer</i>	word evokes notions such as ‘market-oriented rural entrepreneur’
Russian 18 <sup>th</sup> -century <i>krest’ianin</i>	word evokes notions such as ‘non-noble person’, ‘person doing manual (not necessarily agricultural) work’, sometimes also ‘unfree person’

The implementation of these objectives has significant implications for funding agencies in particular:

1. Funding agencies need to encourage and maintain digitisation programmes and development of meta-data and tools for translating national language sources.
2. Funding and commercial agencies (such as those engaged in mass digitisation initiatives)<sup>114</sup> need to pay attention to the complexity of language and chronology issues when making funding decisions, in order to avoid a situation where only data from modern, English-language sources is digitised.
3. Funding agencies need to engage in making the linguistic diversity of the European research output and publication landscape visible, accessible and recognised.
4. Funding agencies need to participate in a close dialogue with the research community in order to identify the research-driven needs for RIs in a multi-lingual and multicultural context.

#### CASE STUDY

Joachim Berger, *Institute of European History, Mainz – DE*

### EGO | European History Online

*EGO | European History Online*<sup>115</sup> is a transcultural history of Europe in the early modern and modern period. It is published by the Institute of European History (IEG) in Mainz in

114. The most known example of this is Google Books: <<http://books.google.com/>>. It is interesting to note that Google has been recently engaged in supporting directly digital Humanities research. Following the result of its Digital Humanities Research Awards and the positive response from the community in July 2010, a second set of awards focusing on European universities and research centres was announced in November 2010. Twelve projects led by fifteen researchers at thirteen institutions across Europe and dealing with different languages were awarded. Quoting the announcement made by the Engineering Manager for Google Books, Jon Orwant: “Projects like these, blending empirical data and traditional scholarship, are springing up around the world. We’re eager to see what results they yield and what broader impact their success will have on the humanities.” <<http://googlepolicyeurope.blogspot.com/2010/12/more-researchers-dive-into-digital.html>>.

115. <<http://www.ieg-ego.eu>>.

cooperation with the University of Trier's Centre for Digital Humanities.

1. As a transnational research and publication structure, EGO aims to transcend disciplinary boundaries within the Humanities. While affirming the legitimacy of the perspectives provided by each separate discipline, the project brings them together under the overarching viewpoint of communication and transfer. Transfer processes that extend across individual, familial and local realms are at its core. EGO traces these transfer processes in and between, amongst others, the spheres of religion, law, politics, art, music, literature, economics, technology, military history, science and medicine. Thematic 'threads' group separate articles into a modular structure arranged thematically and methodologically. These threads are transdisciplinary, transnational and multithematic; they join, on a common (online) platform, the perspectives of different historical disciplines and their international authors. While different authors are of course guided by particular disciplinary perspectives, their combination allows the users to acquire a cross-disciplinary perspective.
2. EGO pursues a multilingual approach that acknowledges the need for a workable meta-language / *lingua franca* in the Humanities but at the same time does justice to the linguistic variety of national academic cultures in Europe: EGO-articles are accepted in English and German. All major contributions are translated by native speakers and published in both languages. In addition, authors may publish their article in their native language. Users are invited to consult both the original and the translation in order to trace differing argumentative patterns and conceptual peculiarities of the respective languages – the linear, examples-driven narrative in Anglo-American scholarly writing, and the dialectic, intricate structure of German academic prose.
3. By taking full advantage of the multi-media potential of the internet, EGO creates a multidisciplinary network. This network is established, on the one hand, via internal links to media published within EGO and, on the other, via links to external images, textual sources and biographical data digitalised or published on the internet, as well as – in the notes – scholarly literature and other academic resources online. The dynamic EGO publication infrastructure thereby merges and groups thematically the range of international online resources on European history. While these external resources represent all national traditions relevant to the history of Europe, EGO makes them accessible to a transnational academic community via a bilingual user interface.

## Education and Training

With increasing frequency over the past twenty years scholars have reflected on the role of computing in the Humanities, and how academia might best respond to the particular changes that it heralds. An area of pressing and ongoing discussion continues to be that of education and training. At the pan-European level, from 1996-2000, for example, the Socrates/Erasmus network ACO\*HUM "aimed at developing an international dimension for investigating the educational impact of new technologies in humanities disciplines".<sup>116</sup> Numerous articles have been published in international journals discussing the form that such curricula might take and the relationship between these and traditional Humanities curricula (see, for example, de Semdt,<sup>117</sup> Unsworth,<sup>118</sup> Terras,<sup>119</sup> Liu,<sup>120</sup> Aarseth<sup>121</sup>).

Here the aim is not to address such debates, rather to reiterate the urgency that must be given to developing educational and training programmes in the area of computing and the Humanities. Furthermore, the attempt is not to address particular technologies, rather to reiterate de Smedt<sup>122</sup> that

"[m]ore important than the use of machines are new ways of thinking that are based on computational methods. Central to these are the formalization and quantification of scholarly problems".

A key aspect of this is the pressing need to identify methodologies, approaches and research questions that go beyond individual disciplines as well as the recognition of the emergence of new generations of scholars who have both Humanities and Computing/Engineering expertise. These cohorts of scholars operate in an essentially interdisciplinary framework, are capable of communicating effectively with both traditional and non-traditional Humanities, and are experts in the

116. <<http://gandalf.aksis.uib.no/AcoHum/>>.

117. Smedt, Koenraad, de, et al. *Computing in humanities education: a European perspective*, SOCRATES/ERASMUS thematic network project on Advanced Computing in the Humanities (ACO\*HUM). Bergen: University of Bergen, 1999. Web (accessed 14/07/2011).

118. Unsworth, John. "What is Humanities Computing, and What is Not?" *Jahrbuch für Computerphilologie* 4 (2002). Eds. Georg Braungart, Karl Eibl and Fotis Jannidis. Paderborn: mentis. Web (accessed 14/07/2011).

119. Terras, Melissa. "Disciplined: Using Educational Studies to Analyse 'Humanities Computing.'" *Literary and Linguistic Computing* 21.2 (2006): 229-246. Print.

120. Liu, Alan. "Digital Humanities and Academic Change." *English Language Notes* 47 (2009): 17-35. Print.

121. Aarseth, Espen. "The field of Humanistic Informatics and its relation to the humanities." *Human IT* 1.4 (1997). Web (accessed 14/07/2011).

122. Smedt, Koenraad, de. "Some Reflections on Studies in Humanities Computing." *Literary and Linguistic Computing* 17.1 (2002): 89-101. Print.



**Figure 16.**  
A DRAPler Project Screen

application of computational methods and techniques to Humanities problems. In order to maintain, let alone drive forward the state of the art of RIs and the objects that populate them (to say nothing of the numerous other roles of computing in the Humanities) it is essential that accredited, third level-degrees are developed, which see the percolating down of such cutting-edge research competence into the third level curriculum.

Given that modern-day technology is impacting every aspect of scholarly life, it is becoming increasingly necessary for scholars who wish to avail of such research outputs, by, for example, using RIs or interrogating computational models of Humanities works, to be able to avail of appropriate training. This may involve the use of specialised tools or, in relation to ethical issues such as the definitions of ownership versus licensing of data, of process and instruments and their connection to intellectual property issues.

In essence, it is not enough to seek to educate future digital humanists only; some aspects of digital research methods must be directed at all students of Humanities. This educational context needs to be addressed and basic versus advanced training identified.

At the heart of this issue lies the *raison d'être* of such education: how can such education and training foster truly interdisciplinary research (and the interdisciplinary research that is needed to develop 'fit for purpose' RIs)? Such education is both dependent on and strives towards fostering the intellectual and institutional conditions necessary for humanists, computer scientists, those in the cultural heritage and library and information science sectors, as well as emerging hybrid disciplines such as Digital Humanities to come together, as equal partners, in a truly interdisciplinary knowl-

edge exchange. Furthermore, it is essential that such communities can participate in a knowledge exchange relating to the development of curricula and training programmes.

Three levels of training and education in Digital Humanities may be identified:

- **Fundamental:** This involves educating and training the traditional scholar in the fundamentals of digital literacy and the use of digital tools and RIs in their research. Our understanding of 'digital literacy' includes the critical understanding of digital objects that is essential to their appropriate use in any research project, for example, that metadata is rarely neutral but that, for instance, markup or text encoding is an interpretation of text.<sup>123</sup>
- **Intermediate:** This may involve teaching programmes and projects which, in the terms of De Smedt (2002), "use strongly computer science-based methods (such as database technology, applied to an information analysis of some specific problem area)", "[...] computation-intensive methods (such as statistics) to gain scholarly results, which could not be gained without the tools employed." In addition to such considerations, an intermediate programme may also include the training of individuals with multidisciplinary expertise who can act as mediators between the traditional and non-traditional communities for example those known as "#alt-ac (alternative to the academy) professionals".<sup>124</sup> A further key aspect is the training of people to work with interdisciplinary awareness, whether in communication or in outcomes.
- **Expert:** Such programmes are aimed to produce expert graduates who, in de Smedt's terms, "can develop new humanities information technology: basic and applied research in text encoding, human language industries, information retrieval, computer games, digital art, etc." The full realisation of expert programmes requires looking beyond the applied technological aspects to include the theorisation and philosophy of the application of computing to all aspects of the Humanities.

123. See, for example, Sperberg-McQueen, C. Michael, Claus Huitfeldt, and Allen Renear. "Meaning and interpretation of markup." *Markup Languages: Theory & Practice 2.3* (Summer 2000): 215-234. Print. (a preprint is available at <<http://cmsmcq.com/2000/mim.html>>).

124. They are a generation of young researchers with traditional and technical expertise who are employed in higher education bodies but rarely in academic positions. They play a crucial role in the challenging work of mediating between the groups discussed in this paper. See especially the open-access collection #alt-academy: a media commons project, edited by Bethany Nowviskie: <<http://mediacommons.futureofthebook.org/alt-ac/>>.



**Figure 17.**  
Computer graphic simulation of the interior of a Çatalhöyük house (Grant Cox, ACRG).

## CASE STUDY

Shawn Day and Faith Lawrence,  
*Digital Humanities Observatory – IE*

### DHO:DRAPler: A Research Infrastructure for Irish Digital Humanities

The *Digital Humanities Observatory (DHO)*<sup>125</sup> is an all-island Digital Humanities collaboratory working with Irish, European and international partners to further e-scholarship. It is a knowledge resource providing outreach and education activities on a broad range of Digital Humanities topics to raise the level of Digital Humanities scholarship in Ireland. At the outset of operations in 2008, the DHO faced two immediate challenges. The first was the lack of familiarity amongst Humanities researchers with Digital Humanities methods, techniques and best practices. The second was lack of awareness amongst researchers of projects in Ireland similar to their own, and or a means to alert researchers to others practising similar techniques and methods and lead to collaboration.

The DRAPler research infrastructure (*Digital Research and Projects in Ireland*)<sup>126</sup> was conceived to remedy these problems, by providing a searchable and browsable online

database to identify, share and discover the types and nature of Irish Digital Humanities projects. It enables users to conduct faceted browsing by subject, funder, temporal periods, geospatial area, metadata standard, encoding methods and the methodology and techniques employed. DRAPler employs SKOS taxonomies developed jointly with the AHRC *ICT Methods Network* to provide future interoperability with similar project databases Europe-wide.

#### Impact

DRAPler currently holds comprehensive information on 71 Digital Humanities projects in Ireland with new projects being added.

DRAPler supports three specific objectives:

1. Discovery of researchers using similar techniques and methods who may be consulted and collaborated with for the purpose of knowledge sharing and expertise;
2. Discovery of collections and data stores that may be relevant to a researcher's own work by identifying initiatives within that subject area;
3. Discovery of projects that focus on Ireland or Irish studies, on the island of Ireland or further afield.

Individual project principals propose their project for submission. DHO staff vets applications to ensure they meet criteria established for inclusion. If so, the project is approved and project information is populated by the project and published. If not applicable the submitter is

125. <<http://dho.ie/>>.

126. <<http://dho.ie/drapier>>.

advised and alternative means of dissemination are suggested. Maintaining currency of project information rests with project contributors to ensure it is kept accurate.

### A Tour of DRAPler

Users of DRAPler can browse the project collection by choosing specific facets of research interest. Lists on the project browser screen present all possible criteria and the number of projects matching those criteria.

Users are presented with a list of projects and short descriptions that match the criteria they have indicated interest in.

Selecting a particular project provides full project information as well as a screen shot and the means to both view the project website (if applicable) as well as contact information to reach staff involved in the project

### Conclusion

DRAPler has bridged the physical research infrastructure through provision of the digital RI that enables researchers working in degrees of academic isolation to identify others working in similar areas and using similar techniques. It allows knowledge and skill sharing not previously possible to drive opportunity for collaborative practice.

The development of DRAPler's focus on projects has led to the development of *DHO:Discovery* that facilitates the discovery of individual digital research objects created by these projects. *DHO:Discovery* is a natural extension of the principle of serendipitous discovery. With *DHO:Discovery* users are able to search across aggregated research collections to discover artefacts that would hitherto not have been collectively matched. Additionally, using advanced data visualisation techniques researchers can visually drill down through immense collections of data to find specific objects of interest.

In the future DRAPler will contain case studies attached to project information to share knowledge about best project practice.

### CASE STUDY

Graeme Earl, *Archaeological Computing Research Group*,  
University of Southampton – UK

## Archaeological Visualisation and Multidisciplinary Research

Training in multidisciplinary research remains poorly developed, and this frequently stems from disciplinary infrastructures that limit interaction between subject areas and a bias toward subject-specific research funding, review and publication. The origins for this are well understood, and include funding structures, the history of disciplinary interactions, and the formulation of research metrics. Arguably the researcher who excels in one tight domain has little stimulus to move sideways, despite the proven benefits of such activity to the broader academic community. In this section I shall introduce the infrastructural challenges that are involved in developing multidisciplinary activities.

The multidisciplinary landscape within which training and research practice occurs can be generalised into four distinct layers. First, local disciplinary research groups. My own research is focused within the Archaeological Computing Research Group (ACRG) that brings together fifty staff, researchers and students working in archaeological computing. Their day-to-day interactions are facilitated digitally by social networking services, shared repositories and other digital research 'spaces', and blogging. In physical terms they have a laboratory and research rooms. Above this disciplinary layer institutional structures are designed to provide training and to identify research synergies. This technological infrastructure is primarily associated with scholarly communications, and means to discover research adjacencies across campuses – both areas which the University of Southampton is currently funded to investigate. Beyond the institution come project infrastructures, characterised by their targeted but ephemeral nature, requiring digital mechanisms for communication and sharing, but still generally dominated by physical interactions. Finally, project, group and individual multidisciplinary activity fit within systems generated at the national and international level, characterised by short-term networking projects. Here digital infrastructures are of crucial significance but risk over-generalising. In terms of training they provide significant opportunities for interactions between junior scholars.

Such multidisciplinary interaction cannot proceed without institutional and research council support. Our work at the interface of digital technologies and cultural heritage is facilitated by the University of Southampton Digital Economy University Strategic Research Group (USRG). USRGs are designed to provide a framework for multi-

disciplinary activity, including regular seminars, targeted internal sandpits, digital communications and mobilisation of university-wide resources and contacts. Its links to training come primarily via university initiatives to broaden curricula and to encourage and facilitate multidisciplinary study at all levels. For instance, our work in visualisation of important archaeological sites such as Çatalhöyük and Portus would be impossible without infrastructures for collaboration with computer science, engineering and other disciplines. In addition, specific multidisciplinary initiatives such as the Web Science and Complex Systems Simulation EPSRC Doctoral Training Centres provide targeted postgraduate training and a wider influence on research and scholarly interaction. In effect they challenge existing disciplines and mint wholly new ones, with attendant benefits in scope and creativity, and problems stemming from the lack of disciplinary history.

The grand challenge for multidisciplinary research remains the efficient and variable balancing of discipline specificity and wider interactions. Training lies at the core of this, in the development of research professionals versant in multiple disciplinary-vocabularies and prepared for the inevitable inconsistencies of divergent domains. However, the multidisciplinary agenda will not progress through training alone, without a consistent policy of investment in multidisciplinary peer review, exposure of research practice to encourage serendipitous developments, and institutional structures that encourage rather than inhibit mixing. It would be counter-productive to encourage multidisciplinary infrastructures to the extent that the lone researcher model fails to operate, but the many examples of poorly peer-reviewed, poorly structured, repetitive multidisciplinary research activities suggest that a robust system will not emerge without deliberate investment. Current funding council and institutional initiatives support this appraisal. Digital infrastructures such as scholarly communications that enable rapid appraisal of a domain of study are one crucial component, but must be supported by policies at all levels encouraging innovative, ambitious work at the interface as much as the core.

-----

## Conclusions: Priorities for Policy and Research

As conclusions to this report, a set of priorities and future research directions for a common strategy on RIs in the Humanities at the European level can be defined as following:

### State of the Art and Needs

- The production of a detailed inventory of **current research activities and future needs** in terms of infrastructural support (e.g., standards/tools/licences) based on a survey/study of inside as well as outside the mainstream of European and/or national projects and infrastructure developments (e.g., within Academies, research units in universities and their departments, libraries).
- The fostering of **partnerships** across communities and institutions (scholarly community/libraries/archives/museums/private institutions and, where appropriate, public bodies and the commercial sector) to achieve better coordination of funding and activities at European level in the fields of cultural heritage and scientific production.
- The recognition and establishment of **RI ecosystems** where multiple levels of RIs (global/pan-European; community-driven; local/institutional) co-exist and collaborate harmoniously. Within such ecosystems, the development of *multilingual, multimodal and multidimensional* frameworks should be supported.
- The development of **higher education programmes** and training opportunities at fundamental, intermediate and expert levels.

### Physical and Digital RIs

- The constant examination of recurring, perceived and actual **challenges** in integrating physical research infrastructures in Humanities and digital research infrastructures and hence the identification of means (and support thereof) to **bridge 'physical' with 'digital'**.
- Increased support for **interdisciplinary profiles** that are able to act as 'translators' between Computer Science/Engineering and Humanities approaches and traditions.

### Strategic Directions

- The development of RIs that build on existing communities and research questions, in order to facilitate research **beyond mono-disciplinary interests** and

across different communities. Infrastructures have the potential to act as ground for cross-fertilisation between the Humanities and other sciences.

- The development of RIs that connect the ‘silos’ of isolated project-based data and resources so as to reach a reliable and dynamic **correlation** across resources and data responding to the complexity of scholarly material and relevant interpretative process.
- The fostering of research programmes that identify and promote **good practices** with respect to interoperability, usability and collection curation (including sustainability and preservation) within and across national boundaries. Indeed, standard development and education together with the development of reference software implementations to encourage their adoption should be considered part of an overall infrastructure development strategy for the Humanities.
- The development of RIs that offer **open access** to processed as well as original data, with regulation by funding agencies of public access to research sources and outcomes of publicly-financed projects (*public access to publicly-funded research* including permission to reuse data).
- The development of RIs in the Humanities that reach across linguistic borders and provide access to large **multilingual** datasets from different cultures to inform comparative and transnational research.
- The development of **legal instruments** (copyright regulations and relevant policies) on a national and international level in order to respond to and steer the digital environment potential according to an open access policy.
- The **sustainability** of RIs through long-term funding commitment beyond thematic or project-based models and their development.
- The deposit of data in **certified repositories**, in order to avoid in future non-accessibility or even loss of primary source materials.

## Partnerships and Networking

- The establishment of partnerships and productive alliances **across communities and institutions** (scholarly community/libraries/ archives/museums and private institutions) based on equality. Cooperation agreements, joint access policies, coordination of depositing rules (as regards, notably, processed data in databases and other forms of research output) are all elements that can only be tackled by the Library, Archive and Information Studies sector and research organisations together.
- The identification of **obstacles to alliances** to date between **RIs providers** (e.g., librarians and archivists) and the academic community (e.g., potential or perceived conflict between libraries’ requests for free, reusable data and researchers seeking to add value and retain recognition of their work).
- The fostering of relevant partnerships with the private sector (e.g., the development of innovation intensive platforms) to contribute to **innovation society** and ‘smart economy’ strategies with potential to foster social and economic benefits, including market innovation.
- The enhancement of the **networked dimension** of RIs (e.g., across CLARIN/DARIAH and other international infrastructures).

## Academic Recognition

- The change towards a **culture** of recognition that accepts the process-oriented character of digital publications.
- The establishment of more comprehensive clearing and authoritative **mechanisms** (including peer review) to secure scholarly reliability and recognition of research across international and interdisciplinary collaborations.
- The fostering of **interdisciplinary tools and teams** where all contributing specialist roles and competences are recognised and rewarded.
- The implementation of **targeted advocacy** (e.g., for emerging cross-disciplinary fields and relevant academic recognition), including the identification of and linkages across appropriate actors.



### Dissemination and Outreach

- The development of RIs that have outreach built into their management plans in order to foster the emergence of a **new culture** that looks beyond established academic circles. It is the **social dimension** of RIs that ensures their good functioning: technological changes are bound to bring innovation only when intellectual and cultural resources are interacting and performing at the best of their potential.
- The demonstration and dissemination of **scholarly results** of research products arising from/facilitated by RIs.
- The dissemination of existing models of good **community of practice** in order to provide education to researchers on how to build new and effective community infrastructures and use existing ones.

### Evaluation of RIs

- The implementation of improved and agreed **systems for evaluation** of RIs and in general of research outputs that are alternative to traditional ones.
- The expansion of current evaluation requirements to include, when appropriate, expectations associated to the development of RIs as well as to their consequent **economic and social benefits**.
- Proper recognition and credit of **interdisciplinary contributions** across research project partners by the European funding agencies and academic professional organisations.
- Proper recognition, credit and career perspectives to foster and promote a **new generation of young researchers** who will be able and willing to take on the challenges discussed in this document.

## References

For a full list see:

<<http://www.esf.org/research-areas/humanities/strategic-activities/research-infrastructures-in-the-humanities.html>>.

## Index of Case Studies

MERIL project: Categories of Research Infrastructures relevant for Humanities	7
TextGrid – Virtual Research Environment for the Humanities	11
The Digging into Data Challenge	12
Research Infrastructures for Historic Artefacts: Knowledge Networks	13
HESTIA (the Herodotus Encoded Space-Text-Imaging Archive): an interdisciplinary project	16
Crowd-Sourcing Art History: Research and Applications of Social Tagging for Museums	18
New Renaissance Report and its Implications	22
Audiovisual Research Documents	23
Communities of Practice: NINES Peer Review Service	25
The Writings of Others: Review Culture at a Turning Point	25
Internationally Recognised Good Practices, Guidelines and Standards: Some Examples	28
Geographical Information Systems and Spatial Humanities	29
Between Research and Public Use: Music-Related Research Infrastructures	30
Towards Comprehensive Bibliographic Coverage of the Scholarly Literatures in the Humanities and Social Sciences	32
EGO   European History Online	34
DHO:DRAPier: A Research Infrastructure for Irish Digital Humanities	37
Archaeological Visualisation and Multidisciplinary Research	38

## List of Authors and Contributors

This ESF Science Policy Briefing has been written by the following experts:

- Professor Claudine Moulin (Chair of the editorial group), Trier Center for Digital Humanities, Universität Trier, Germany – SCH member (LU)
- Dr Julianne Nyhan, Trier Center for Digital Humanities, Universität Trier, Germany/ Department of Information Studies, University College London, United Kingdom
- Dr Arianna Ciula, Humanities and Social Sciences Unit, European Science Foundation, France
- Professor Margaret Kelleher, An Foras Feasa, National University of Ireland Maynooth – SCH member (IE)
- Professor Elmar Mittler, Göttingen University, Germany
- Professor Marko Tadić, Department of Linguistics, University of Zagreb – SCH member (HR)
- Professor Maria Ågren, Department of History, Uppsala University – SCH member (SE)
- Professor Andrea Bozzi, Istituto Linguistica Computazionale, CNR – SCH member (IT)
- Professor Kristin Kuutma, Institute of Cultural Research and Fine Arts, University of Tartu – SCH member (EE)

The following experts contributed to the Science Policy Briefing (including Case Studies) with their knowledge and expertise:

- Professor David Arnold, Faculty of Management and Information Sciences, University of Brighton – UK
- Dr Elton Barker, The Open University – UK
- Mr Michael Bender, Institut für Sprach- und Literaturwissenschaft, Darmstadt University of Technology – DE
- Dr Joachim Berger, Institute of European History, Mainz – DE
- Mr Brett Bobley, US National Endowment for the Humanities, Office of Digital Humanities – USA
- Professor David J. Bodenhamer, School of Liberal Arts, Indiana University Purdue University Indianapolis – USA
- Dr Stefan Bouzarovski, Department of Geography, Earth and Environmental Sciences, University of Birmingham – UK
- Professor Ewa Dahlig-Turek, Institute of Arts, Polish Academy of Sciences and Humanities – PL
- Mr Shawn Day, Digital Humanities Observatory – IE

- Dr Graeme Earl, Archaeological Computing Research Group, University of Southampton – UK
- Professor Gudrun Gersmann, German Historical Institute, Paris – FR/ DFG committee on electronic publications – DE
- Dr Ian N. Gregory, Faculty of Arts and Social Sciences, Lancaster University – UK
- Mr Leif Isaksen, School of Electronics and Computer Science, University of Southampton – UK
- Professor Elisabeth Kieven, Bibliotheca Hertziana, Max Planck Institute for Art History – DE/IT
- Dr Andreas Kunz, Institute of European History, Mainz – DE
- Dr Lilian Landes, Bavarian State Library, Center for Electronic Publishing – DE
- Dr Faith Lawrence, Digital Humanities Observatory – IE
- Professor Louisa-Irene Loukopoulou, Institute for Greek and Roman Antiquity, National Hellenic Research Foundation – GR
- Professor Christopher Pelling, Christ Church, Oxford – UK
- Professor Andrea Rapp, Institut für Sprach- und Literaturwissenschaft, Darmstadt University of Technology – DE
- Mr Oliver Schmid, Institute of Economics, Darmstadt University of Technology – DE
- Dr Dietrich Schüller, Phonogrammarchiv, Austrian Academy of Sciences – AT/UNESCO Working Group on Information Preservation
- Mr Robert Stein, Indianapolis Museum of Art – USA
- Ms Dana Wheeles, Networked Infrastructure for Nineteenth-Century Electronic Scholarship, University of Virginia – USA
- Professor Jost Gippert, Vergleichende Sprachwissenschaft, Institute for Linguistics, Universität Frankfurt/Main – DE
- Mr Steven Krauwer, Utrecht Institute of Linguistics, Utrecht University – NL (ESFRI CLARIN)
- Professor Rūta Marcinkevičienė, Center of Computational Linguistics, Vytautas Magnus University – LT (SCH member)
- Dr Cristina Marras, Istituto per il Lessico Intellettuale Europeo e Storia delle Idee, CNR, Roma – IT
- Professor Louise McNally, Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona – ES (ESFRI SSH working group 2007-2008)
- Professor Jan Christoph Meister, Department of Language, Literature and Media, University of Hamburg – DE
- NeDiMAH Steering Committee, ESF Research Networking Programme <<http://www.esf.org/nedimah>>
- Professor Franco Niccolucci, Science and Technology in Archaeology Research Centre, The Cyprus Institute – CY
- Professor Christian-Emil Ore, Unit for Digital Documentation, Faculty of Humanities, University of Oslo – NO (CIDOC-CRM)
- Professor David Robey, Oxford e-Research Centre – UK
- Professor Tomojy Tabata, Graduate School of Language and Culture, University of Osaka – JP

**In addition, the following experts gave invaluable input following a wide consultation process:**

- Mr Arno Bosse, Research and Development Department, Göttingen State and University Library, Georg-August-Universität Göttingen – DE
- Dr Milena Dobрева, Computer and Information Sciences Department, University of Strathclyde – UK
- Dr Peter Doorn, Data Archiving and Networked Services, Royal Netherlands Academy of Sciences – NL (ESFRI DARIAH)
- European Library Management Committee c/o the National Library of the Netherlands, The Hague – NL

**This ESF Science Policy Briefing has been prepared under the responsibility of:**

the Standing Committee for the Humanities (SCH)

- Chair: **Professor Milena Žic-Fuchs**
- Head of the Humanities and Social Sciences Unit: Dr Nina Kancewicz-Hoffman
- Science Officer: Dr Arianna Ciula
- Editorial support: Mr Diego De La Hoz Del Hoyo, Science Officer

This document was revised after comments from external peer reviewers.

The European Science Foundation (ESF) was established in 1974 to provide a common platform for its Member Organisations to advance European Research collaboration and explore new directions for research. It is an independent organisation, owned by 78 Member Organisations, which are research funding organisations, academies and learned societies from 30 countries. ESF promotes collaboration in research itself, in funding of research and in science policy activities at the European level.

**European Science Foundation**  
1 quai Lezay-Marnésia • BP 90015  
67080 Strasbourg cedex • France  
Tel: +33 (0)3 88 76 71 00  
Fax: +33 (0)3 88 37 05 32  
[www.esf.org](http://www.esf.org)

ISBN: 978-2-918428-50-3  
September 2011 – Print run: 3000