

## Keynote 3

### Digital Libraries and Digital Humanities: Some Reflections on their Synergy

Widad El Hadi

University of Lille 3, France.

Email: widad.mustafa@univ-lille3.fr



This paper addresses the relationship between Digital Libraries (DL) and Digital Humanities (DH) with a special focus on global access to knowledge. An overview DL & DH and their interconnections are described. This is followed by an overview of what the humanities need from a global networked infrastructure involving Digital libraries: in other words, what can Digital Humanities learn from Libraries? Conversely, how can Libraries be DH-Friendly? To conclude we propose some research orientations that emerged from comparing DL and DH agendas, and we outline the challenges that still lay ahead. Review paper on the State of the art of Libraries and Digital Humanities. What do the humanities need from a globally networked infrastructure involving Digital libraries? What can Digital Humanities learn from Libraries? What would a DH-friendly Library involve? What are the core elements and activities that would be involving both? The challenges facing both Digital Libraries and Digital Humanities are discussed and some elements on future directions are outlined.

**Keywords** Digital Humanities, Digital Libraries, Knowledge Organization, Virtual Research Infrastructures, Global Networked infrastructures

#### Introduction

Berry (2011) describes that

The digital humanities try to take account of the plasticity of digital forms and the way in which they point toward a new

way of working with representation and mediation, what might be called the digital 'folding' of reality, whereby one is able to approach culture in a radically new way (p.1).

To begin, I will first focus on some definitions of Digital Humanities (DH) drawn from the fast-growing literature in the field. DH is defined as "an important multidisciplinary field, undertaking research at the intersection of digital technologies and humanities. It aims to produce applications and models that make possible new kinds of research, both in the humanities disciplines and in computer science and its allied technologies. It also studies the impact of these techniques on cultural heritage, memory institutions, libraries, archives and digital culture" (Warwick, et al., 2012). The authors of the French Manifesto suggest another definition for Digital Humanities (Dacos, 2010). The authors describe this field as a "community of practice born out of the many issue-related communities that came from the interest directed at the tools relating to various transversal objects (codification of textual sources, geographical information systems, lexicometrics, the digitalization of cultural, scientific and technical heritage, web cartography, oral archives, digital arts and literature and so on), these communities are progressively converging to give birth to the field of digital humanities".

The digital humanities concept has its own history which cannot be reduced to the dimensions of current technological developments (El Hadi et al., 2014). What is new is not the use of information technology for the

humanities but the fact that the various fields they deal with, including arts and social sciences, have gathered around the DH. Its scope can be extended to “building, managing and valorizing data gathering, texts, images of various types or periods belonging to a variety of domains is now important and strategic enough to have led to the birth of a community of practices called Digital Humanities” (El Hadi et al., 2014). What is new is that we are witnessing, as pointed out by Svensson (2010), a rich multi-level interaction with the “digital” that is partly a result of the persuasiveness of digital technology, and all the approaches it involved. Humanists are exploring different modes of engagement, institutional models, technologies and discursive strategies (El Hadi et al., 2014).

The domain of DH embraces not only humanities but also a wide range of theorists and practitioners, those who have been active in the field for decades, and those who became recently involved. Disciplinary experts, computer scientists, library and information studies specialists have all been brought together to consider DH as a discipline of its own to reflect on how it relates to areas of traditional humanities scholarship (Schreibman, et al., 2010).

According to McCarty (2009), to begin with the historical dimension, the Digital Humanities have had a rather interesting history. They started as ‘computing in the humanities’, or ‘humanities computing’, in the early days they were often seen as a technical support to the work of the ‘real’ humanities scholars, who would drive the projects. This involved the application of the computer to the disciplines of the humanities, something that has been described as treating the ‘machine’s efficiency as a servant’ rather than ‘its participant enabling of criticism’ (as cited in Berry, 2011, p.2).

An analysis of the literature shows how “the incorporation of the computer into Humanities research is changing how we can transmit, shape, understand, question and imagine all aspects of the Human record. In the non-

specialist sense, digital technologies offer Humanities scholars new ways of carrying out activities they have always been engaged in, for example, when social media is used in place of the more traditional letter and epistolary network” (Nyhan, 2013, p.1).

Tasovac (2015) sketched the institutionalizing of Digital Humanities in the following : Computers and the Humanities (1966); The Association for Literary and Linguistic Computing (Londonbased, 1973); Association for Computers and the Humanities (US-based, 1987) ; Literary and Linguistic Computing (1986); Alliance of Digital Humanities Organizations (ADHO, 2005); and, Currently, Master’s and PhD programs in many Western universities such as UCL London, King’s College, University of Oxford, University of Lausanne, University of Paris 3 (in some Master programs), University of Lille 3, ENS Lyon, among many other.

Digital technologies offer Humanities scholars opportunities to use methodologies and techniques that have not traditionally been available to them, for example, image processing and 3D modeling. Theoretical and philosophical changes are also changing as “Digital Humanities research begins to question the gap between making and thinking, or *epistêmê* and *technê*, a gap that has long been in existence and remains highly indicative of Humanities research” (Nyhan, 2013, p.1). As Berry (2011) pointed out,

Few dispute that digital technology is fundamentally changing the way in which we engage in the research process. Indeed, it is becoming more and more evident that research is increasingly being mediated through digital technology. Many argue that this mediation is slowly beginning to change what it means to undertake research, affecting both the epistemologies and ontologies that underlie a research programme. (...) this development is variable depending on disciplines and research agendas, with some more reliant on digital technology than

others, but it is rare to find an academic today who has had no access to digital technology as part of their research activity. Library catalogues are now probably the minimum way in which an academic can access books and research articles without the use of a computer, but, with card indexes dying a slow and certain death (Baker, 1996, 2001), there remain few outputs for the non-digital scholar to undertake research in the modern university (p.1).

Digital Humanities research is usually considered as collaborative and interdisciplinary, crossing the traditional academic borders between Computer Scientists, Engineers, Library and Information Professionals as well as Humanities Scholars (Moulin et al., 2011). The inclusion of the computer in Humanities research is altering not just the scope and possibilities of that research, but the very conditions under which it is carried out (Nyhan, 2013). This idea is also highlighted by Warwick et al., (2008) who showed how DH are linked to the change of scope of research and the role libraries can play in promoting it: "Information resources such as libraries, archives museums and research centres, and the web pages that provide information about them are vital for humanities scholars. The university library website was considered the most important resource, even compared to Google. Secondary finding aids and reference resources are considered more important than primary research resources, especially those produced by other scholars, whose output is less trusted than publications produced by commercial organizations, libraries, archives and museums". (Wawick et al., 2008).

The core activities of DH can be sketched in the following domains: textual editing (digital editions, digital collections), data modeling and creation of structured data (XML), distant reading (as opposed to close reading), statistical methods and processing, data mining, stylometry, visualization, network analysis, and tools and services, including research infrastructures such as CLARIN, DARIAH, TGIS

HumaNum, for instance. These infrastructures are described in section 2.3.4.

### **Digital Libraries and Digital Humanities: in the Continuity of Building Knowledge Organization Systems (KOS) for the Humanities**

Although the relationship between the academic discipline of Knowledge Organization underpinning the Digital Libraries theoretical and methodological foundations and Digital Humanities remains as yet unexplored (Koltay, 2014), it is however possible to highlight common issues and challenges common to the two fields of study. If we go back to the history and nature of DH in order to shed light on some of the related issues linking them to Digital Libraries, worthwhile to bring up the parallel established between Information Science and its relationship with other disciplines. Koltay (2014) pointed out that Information Science has imported knowledge and methods from other disciplines and then exported ideas to different fields such as computer science and management (Buckland, quoted by Koltay 2014, p. 25). The nature of the digital humanities is in this regard similar in that it developed alongside corpus linguistics, something that is situated on the boundary between humanities, social sciences and applied sciences. The first modern manifestation of DH was the early adoption of 'computational linguistics' known today as corpus linguistics and language technologies. This field was built around the use of computers for natural language and speech data processing.

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 Research Infrastructures in the Humanities. (Moulin et al., 2011, p.3).

But the birth of modern (Appendix A) DH is traditionally, traced back to the contribution of Father Roberto Busa: “the 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 (Moulin et al., 2011, p.9).

It is interesting to follow the relationship established by Moulin al. (2011) between early manifestations of Digital Humanities (Appendix B) and of knowledge organization practices, information processing and capturing and scholarly communication and Research Infrastructures (RIs).

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 emphasize the importance of collection, organization 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 of knowledge on a glossographical and encyclopaedical 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 encyclopedia 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 multivolume critical editions or serial corpora. These in turn often sparked the development of new research tools, such as extensive indexes, bibliographies, biographical dictionaries (p. 4).

The use of Knowledge Organization System (KOS) would enable building gateways between resources and systems. Knowledge Organization (KO) as a discipline and as a profession has had a tradition of adopting new technologies (Machine readable Cataloging, for example) and currently more advanced technologies are being used. Descriptive metadata (such as the Dublin Core Metadata Set) has been widely used to support identification and discovery of resources in digital environment. Apart from formal description of the resource, descriptive metadata can carry information about the content (subject of the document) and this subject can be expressed using a term from a KOS. By using this model, it is possible to describe the contents of a textual document, image, audio or dataset. This model of information discovery is especially recommended in the following scenarios (Slavic, 2011; El Hadi, 2013; El Hadi & Favier, 2014; Digital Cultures and Universality in Knowledge Organization (DIGIKO), 2013) - when digital libraries contain heterogeneous resources (sound, images, datasets, texts) that cannot necessarily be retrieved through text retrieval techniques; for digital libraries that may contain resources spanning the entire universe of knowledge, making it hard for verbal indexing methods to contextualize and relate semantic



content. The presence of a systematic knowledge structure allows the semantic relationships between such concepts to be made explicit, facilitates navigation of the collection and automatic search expansion (Slavic, 2013; DIGIKO, 2013). Lack of such a structure prevents semantic content integration; - Then there is the fact that digital libraries often integrate or become connected through integrated subject gateways. In these situations, the content to be searched may exist in different languages, occur in different subject areas, and be indexed by different local knowledge organization systems. More importantly, digital repositories often merge digital resources originating from research, library and educational domains. Integration of information and alignment of concepts from these different areas can only be achieved by superimposing a knowledge organization scheme. But developing well adapted semantic technology remains one of the challenges. Over the past ten years a number of W3C workgroups in the area of Semantic Web development have been showing a keen interest in knowledge organization systems traditionally used in the bibliographic domain (thesauri and classifications) and their power in capturing and communicating meaning. This is especially the case with W3C development groups in the area of vocabularies, web ontology standards, linked data and more recently the W3C Library Linked Data Incubator. The mission of the Library Linked Data incubator group is to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities— focusing on Linked Data— both in the library community and beyond, by building on existing initiatives and identifying collaborative tracks for the future, (DIGIKO, 2013).

The group has explored the ways and means according to which existing building blocks of librarianship, such as metadata models, metadata schemas, standards and protocols for building interoperability - encourage libraries to re-orient their approaches to data interoperability towards the Web, also reaching

to other communities. It has also envisioned these communities as a potential major provider of authoritative datasets (persons, topics...) for the Linked Data Web. As these evolutions raise the need for shared standardization efforts within the library community around the Semantic Web standards, the group has also sought to refine this need by expressing requirements for standards and guidelines, and by offering a way forward so that the library community can contribute to further Web standardization (DIGIKO, 2013).

### **Libraries as Infrastructure for Digital Humanities**

An extensive literature about the links and core elements between Knowledge Organization and Digital Humanities (many conferences are held on this topic) has been highlighted by several authors, (Barret, 2014; Moulin et al., 2001; Warwick et al., 2008, 2012; Calhoun, 2014; El Hadi & Favier, 2014). Instead of trying to be exhaustive in carrying out this comparison I would rather focus on the key themes that have recently emerged.

The common core activity for Digital Humanities and memory institutions such as libraries, archives, and museums is digitizing the representations of cultural and historical documents, images, and artifacts (Barret, 2014; Moulin et al., 2001; Warwick et al., 2008, 2012; Calhoun, 2014; El Hadi & Favier, 2014). Most of these resources are delivered online to users. The most manifest common core issues can be summed up as follows:

The emergence of Digital Libraries in the early 1990s was a turning point and a critical component of the world-wide shift to networked information. Libraries are considered as an important part of the DH infrastructure (Svensson, 2010). Resource digitization became the basis for both Digital Library (DL) holdings and Digital Humanities (DH) research. Libraries began to create and deliver digital representations of cultural and historical documents and images to improve access to the material they held. The DH therefore began to be

identified with projects, characterized by several things:

- The handling of a big volume of data.
- The use and development of digital tools allowing the processing, exploitation and scientific distribution of these data. Examples of these common issues can be found in the achievements of DH which are also common practices for the DL. Libraries can focus on a wide range of targets: complete and extend a paper publication, gather and offer resources on a given theme, reconstitute scattered original collections/ sets such as the Charles V library (Appendix C) virtually reconstituted through the Europeana Regia program, (for more details see Barret (2014) who gave several examples drawn from the French context.
- Another domain engaging both DH and DL are collections and digital corpora. This development appeared relatively early in the history of Humanities Computing (HC), with the constitution of collections or digital corpora of study, at first centered on paper documents, which later extended images and sounds. The resulting sets can come from either the digitalization of a pre-existent material (mainly documents), as is the case for most of current digital libraries, or from the gathering of materials that were initially digital, such as the internet archives collected at the Bibliothèque Nationale de France (BNF).

The processing and the analysis of data, which were at the heart of Humanity Computing (as mentioned above), remain an important segment of DH. Fundamental for the textual analysis, they concerned firstly philology, linguistics and their related domains.

### **Research Cyber-Infrastructures**

Today, Research Infrastructures (Appendix D) moved to what is called "Research Cyber-infrastructures". This concept is defined by

Rockwell (2011) in his blog as: "Anything that is needed to connect more than one person, project, or entity is infrastructure. Anything used exclusively by a project is not". This type of infrastructure is essential for setting the place of humanities within the digital realm.

In the same way Babeu (2011) presents the components of cyberinfrastructure as follows: the network, discipline-specific software, data collections, tools, expertise/best practices, and standards. -At its core, cyberinfrastructure is made up of extensive and reusable digital collections, but each of the categories mentioned above is also vital to the success of a cyberinfrastructure.

*Researching cyber-infrastructures for the humanities.* Historically, Humanities researchers have long been familiar with Research Infrastructures (RIs) and the objects that populate them such as archives, museums, galleries and libraries where collections of physical objects such as archaeological fragments; paintings or sculptures; inscriptions manuscripts books and journals were kept. An infrastructure is thus considered as the technical and operational framework that allows researchers to collaborate and share data and results, (Moulin et al., 2011). Many definitions of RIs have been formulated over the years. Regarding Humanities, it should be stressed that there are special dynamics and aspects that must be considered while defining this type of RIs.

The European Strategy Forum on Research Infrastructures (ESFRI) has focused on a broad definitional approach that spans the disciplines, Moulin et al., (2011) adapted this definition and extended it in order to account for an international context in which cyber infrastructures may grow. Their inclusive definition evokes both large and small scale, as well as international and national infrastructures: "the term cyber infrastructure 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: cyber infrastructure 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 cyber infrastructure, whereas fiber-optic cables and storage area networks or basic communication protocols would fall below the line for cyber infrastructure" (Appendix E).

Science and Technology researchers have made significant contributions 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. The main features of their work can be transferrable to Humanities; however, specific requirements of RIs should be integrated, as Moulin et al., (2001) noted. One of the main aspects to consider concerns the nature of datasets bases which tend to be generated in the hard sciences, and to be homogenous in nature, i.e., numeric. In Humanities, data tend to be collected and heterogeneous in content and format. Therefore, the considerable advances in RIs that have been made in the Sciences cannot simply be transferred to Humanities 'out of the box' as Moulin et al., (2011), argued. *The Common Forms and Features of RIs for the Humanities*. Moulin et al., (2011, p.6) gave a description of the most common forms of RIs found in Humanities and identified, four primary layers of RIs:

- "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/analogous objects;
  - Access to services;
  - Access to expertise;
  - Access to laboratory facilities"

*Categories of Research Infrastructures relevant for Humanities Defined by the European Commission Framework Program (Appendix F)* Within this program a set of categories or types of research infrastructure has been proposed and listed by Moulin et al., 2011, p.6). Here is a summary of the most important elements. *Humanities-specific RIs*. Cognitive Sciences 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
- 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
- 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 (Moulin et al., 2011, p.7)

#### *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

*Research Infrastructures Relevant for Humanities, Examples ISIDORE (France)* ISIDORE (Appendix G), is a platform enabling research and access to the digital and digitized data from the on research human and social sciences (SHS). ISIDORE harvests metadata and indexes the digital data by enriching the resources with the terms of scientific reference tables. ISIDORE (Appendix G) stands in French for (Intégration de services, interconnexion de données de la recherche et de l'enseignement). This research platform, was set up within the framework of TGIR HumaNum and offers a unified access to SHS resources: archival storage units (HAL-SHS, theses.fr, TEL) digital libraries (Gallica) and electronic publishing platforms (Persée, Cairn, Revues.org).



*TGE-ADONIS, TGIR Huma-num (France).* The Humanities and social sciences disciplines followed in the footprints of science and technology. In France, for instance, recent institutional programs such as TGE-ADONIS, currently called TGIR Huma-num, stands for DH, as in French it is “Humanités Numériques” (Huma-Num, 2015). This infrastructure aims at achieving an SHS research “digital turn” (Berry, 2011). Today’s technological offer supply proposals remain basically oriented towards data processing, conservation, research data, interoperable access and interoperability between repositories.

*High Throughput Humanities e-Research (HiTheR) (UK).* High Throughput Humanities e-Research (HiTheR) was focused on the creation of a digital system for managing the Nineteenth Century Serials Edition (NCSE) corpus. The NCSE contains around 430,000 articles that originally appeared in roughly 3,500 issues of six 19th Century periodicals. The project investigated the use of grid technologies and high throughput computing to provide more intuitive ways of searching the NCSE’s large corpus. Specifically, the project set up a prototype campus grid and used it for carrying out text processing on this corpus. The project tied in with campus grid activities at King’s and the National Grid Service (UK). The website set up to carry out these investigations is being developed further as part of the Forging Restful Services for e-Humanities (FReSH) project at CeRch (King’s College London, 2015).

*DARIAH (Europe).* DARIAH, the Digital Research Infrastructure for the Arts and Humanities, aims to enhance and support digitally-enabled research and teaching across the humanities and arts. DARIAH is a network of people, expertise, information, knowledge, content, methods, tools and technologies coming from various countries (DARIAH, 2015).

*CLARIN, Common Language Resources and Technology (Europe).* The CLARIN project is a large-scale pan-European collaborative effort aimed at making language resources and

technology readily available for the whole European Humanities (and Social Sciences) community. This includes coordinating the development of appropriate resources. Amongst other things, CLARIN will offer scholars tools for computer-aided language processing (CLARIN, 2015).

The European Resources Infrastructure that CLARIN will create is based on an open European Federation of strong service centres and repositories that jointly provide (CLARIN, 2015):

- knowledge of existing language resources
- coordinated creation of, archiving of, and access to such resources,
- access to services and tools that would allow scholars to utilize such resources, and
- bundling of and access to expertise related to specific language processing problems (CLARIN, 2015).

Taking into consideration these elements and achievements, a propitious moment for collaboration between libraries and the humanities was highly acclaimed by many THATCamp “The Humanities and Technology Camp”. The Saint Malo’s THATCAMP called for more cooperation between DH and libraries. In line with Saint Malo’s, initiative (October 2013), the meeting came out with a proposal from Olivier Le Deuff giving a first reflection on “the relations between digital humanities and libraries: debate on competences, trends in profile evolution dialogue between libraries the research world, the use of information by researchers” (THATCamp Paris, 2010). These declarations and initiatives open the way for a possible collaboration that merges with a reality on the ground. The feedback experience from the librarians show that researchers, while deserting the central libraries, have been turning to documentation specialists to carry out DH projects. The declaration highlighted the Library’s positive assets, namely its experience in computerization, electronic resources digitalization and other services, (Barret, 2014).

The digital technology is today widely present in libraries of higher education and research where it came in with computerized document processing and spread significantly due to the development of electronic resources and digitalization programs: acquisitions, cataloguing, data dispatching, and management of periodicals. The whole book circuit is managed today through the integrated system of library management (SIGB) which in the 1990s replaced the first independent modules of commercial computing (Barret, 2014).

### **Overview of What the Humanities Need from a Global Network Infrastructure Involving Digital Libraries**

*The Virtual Research Environments (VREs)* Crane (2007) states that Digital libraries are far more than simple digital surrogates of existing conventional libraries. They are, or at least have the potential to be, complex Virtual Research Environments (VREs), (as cited in Dunn et al., 2008). This concept is defined by the UK Joint Information Systems Committee (JISC) Virtual Research Environments Programme, as comprising: "A set of online tools and other network resources and technologies interoperating with each other to support or enhance the processes of a wide range of research practitioners within and across disciplinary and institutional boundaries. A key characteristic of a VRE is that it facilitates collaboration amongst researchers and research teams providing them with more effective means of collaboratively collecting, manipulating and managing data, as well as collaborative knowledge 'Creation' (Joint Information Systems Committee (JISC), 2015)". Judith Wusteman (2008) traced back its origins and links with Learning Environments (VLEs):

"Virtual Research Environments (VREs) will increasingly change the face of research and librarians have a crucial role to play in their development and use. In the future, VREs should be as fundamentally integrated into the university library information space as Virtual Learning Environments (VLEs) are

now. If VREs are to achieve their potential in supporting researchers, librarians need to ensure that they are involved at this early stage of their evolution so that they can influence their development" (p.68).

Yet researchers and e-infrastructure providers in the humanities lack the resources to realize this full potential. The need for service-oriented architectures that, at the very least, deliver textual resources from digital libraries to the researcher's desktop has been highlighted and a remote access to content is considered to be the most basic requirement. Within such VREs, the concepts of 'reader' and 'author' need to be rethought (Dunn & Blanke, 2008).

Web 2.0 technologies, allied with service-oriented data delivery services, would enable readers to interact creatively with texts, by (for example) selecting elements from different libraries, or using customization tools remotely, to annotate, aggregate, compare and structure text according to their own research needs. In other words, placing digital libraries within global infrastructures would allow the reader/user to break down the distinction between library A, library B and their own desktop, (Dunn & Blanke, 2008).

Integrating text mining tools in VREs is essential: a VRE that is based on digital library infrastructure will have to include not just text, but software tools that allow users to analyze, retrieve elements of and search those texts in ever more sophisticated ways (Dunn & Blanke, 2008).

*What are the Benefits of the VREs for the Humanities?* Historically the humanities have had far less funding and e-infrastructure than the sciences, and researchers in the humanities dealing with digital texts have felt understandably inhibited by this (Dunn & Blanke, 2008). The workshop, reported for in focused on four areas that will become critical as the e-science agenda impacts on the established textual humanities. The major benefits and

challenges of VREs for the Humanities can be sketched as follows (Appendix I).

### Other Facilities Sustaining VREs

- a) *Textmining*. The standard approach in humanities research is not to solve a problem by testing one hypothesis against another, but rather to enrich the object of study by repeated observation and reporting. Data mining tools and their accompanying visualizations, which facilitate pattern finding across a wide range of data, can definitely play a role in this process
- b) *Interfaces*. The online availability of a wide range of digital data has resulted in a corresponding increase in various kinds of tools for retrieving and handling the items in a collection. Interface design researchers have worked on systems intended to help users' access to digital images, work with electronic text files and apply data mining algorithms to a variety of problems, both in the sciences and in the humanities.
- c) *Semantic Annotation of Texts within Textual VREs*. Semantic Annotation (SA) means adding meaningful structures to document resources. It is particularly useful for making computers communicate with each other more effectively. SA can be one of the elements fostering systems and resources interoperation (El Hadi, 2015).

*How can Libraries be DH-Friendly?* To render a Digital library DH-friendly, Tasovac (2015) stressed the importance of anchoring digital scholarly methods in Digital library workflows. This means developing tools and services for different types of scholarly primitives in modular, dynamic, research driven digital collections. Texts in digital libraries must be machine-computable, not only machine-deliverable but should be a kind of service as opposed to a fixed object in an interactive method (enabling Optical Characters

Recognition and Natural Language Processing) as opposed to a static element.

Digitized texts or digitally-born texts are to be considered a "service". This means they should include the following features: metadata, full text search, retrieving identifiable fragments of texts and constant identifiers.

The concept of "World DH-friendly digital libraries" developed by Tasovac (2015) is quite interesting. A DH-friendly digital library would:

1. provide access to machine-readable and machine computable texts
2. develop well-documented APIs for both metadata and full-text access
3. offer layers of automatic and semi-automatic structural, linguistic and semantic annotations with support for crowd-sourced corrections
4. integrate dictionary and other type of reference services to be queried directly from the text Tasovac (2015).

Such a Digital Library would then be considered as an Infrastructure for the Humanities. This setting is comparable to the elements of VREs described above.

### Challenges and Research Perspectives Involving DL & DH

I will here focus on a selection of some crucial issues, namely, Interoperability between heterogeneous systems and resources, Humanities scholars information seeking behaviors, and Digital and Scholarly communication.

#### *Interoperability Between Heterogeneous Systems and Resources*

One of the crucial issues and challenges facing Digital Libraries is interoperability, (Favier & El Hadi, 2013; El Hadi & Favier, 2014; El Hadi, 2015). This issue became increasingly important as digital library projects, publishers, professional indexing and online services brought content online and demand grew for unified access to

content locked up in separate systems with separate interfaces.

#### *Humanities Scholars Information Seeking Behaviors*

One of the areas of interest in Digital Humanities in relation with information and knowledge organization is to look at Humanities scholars information seeking behaviors, i.e. how they use information whether digital or not. Recent studies showed that Humanities scholars have different information needs than scientists (Terras, 2012). Until recently, more resources have been designed for the majority of users who are not from the humanities. Developing digital libraries for this population is vital for future research in the humanities.

#### *Digital and Scholarly Communication*

As Coble et al., (2014) posited, both Scholarly communication and Digital Humanities resist simple definitions. They suggested a broad definition (Appendix J) accounting for the synergy between the two fields:

“Scholarly communication is the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use. This system includes both formal means of communication, such as publication in peer-reviewed journals, and informal channels, such as electronic listservs” (p.1137).

Needless to say how relevant Scholarly communication is for the Digital Humanities agendas. Both activities are proliferating in academic libraries. Cole et al., (2014) described the forms of activities engaging digital humanists and librarians in the following:

“Digital humanists—a group that includes librarians and non-librarians based in libraries, as well as scholars and practitioners without library affiliation—have developed forms of scholarship that demand and introduce complementary innovations focused on infrastructure, modes of dissemination and evaluation, openness, and

other areas with implications for scholarly communication. Digital humanities experiments in post-publication filtering, open peer review, middle-state publishing, de-centering authority, and multimodal / nonlinear publication platforms are discussed in the context of broader library scholarly communication efforts” (p.1136).

#### **Discussions and Recommendations**

Examining the role of libraries in the digital humanities means thinking about the positioning of libraries towards research and the repositioning of their missions and of their identity. The problem posed by DH coincides in this respect with a more general reflection begun by the profession since the impact and the massive distribution of digital technology. Because DH tends to question the usefulness of libraries, be it in their physical presence as well as in their mediation role, digital technology has been perceived at first as one of the major factors of the crisis affecting libraries. This perception is however changing, and libraries having understood its unavoidable function are beginning to think of it not any more as an obstacle but as a springboard towards the future. Offering to libraries the possibility of taking advantage of the expertise they have acquired through the use of electronic resources Digital Humanities is indeed pointing in that direction. While the digital technology tended to keep researchers away from reading rooms by providing them with means to work remotely, DH appear as a means to re-conquer this public (Barret, 2014).

Libraries have an obvious interest in taking their place in digital humanities and real assets to offer for research in SHS. Warwick et al., (2008), conducted a study on the use and importance of information resources, physical research centers and digital finding aids in scholarly research. They came out with the following findings:

“Information resources such as libraries, archives museums and research centers, and the web pages that provide information about them are vital for



humanities scholars. The university library website was considered the most important resource, even compared to Google. Secondary finding aids and reference resources are considered more important than primary research resources, especially those produced by other scholars, whose output is less trusted than publications produced by commercial organizations, libraries, archives and museums”, (p.1).

A survey conducted by Barret (2014) on the experience of American libraries which have set up means of supporting DH or which have even chosen to work directly with them, underlines the wealth of the perspectives opened by this new trend. The projects that are at present being developed in France stress the fundamental role to be played by libraries on the topic of tool appropriation and digital use. Setting up a true collaboration with the researchers, other than the factual and informal one which is at present common, presupposes however to succeed in bridging the gap which too often still separates information professionals from the researchers, for historic, structural, cultural and social causes (Palermi et al., 2002).

It is now necessary to demonstrate that humanities research is as important to society as scientific research, and is deserving of more consideration from funders. Digital technologies are spreading and developing and this is a favorable moment for collaboration between DL and DH. As Hocky predicted in 2008: “Future generations of scholarship in the arts and humanities will depend upon the accessibility of a vast array of digital resources in digital form” (Hocky et al., 2008, p.69) quoted by Terras (2012), in (Wrawick et al., 2012).

## References

- Babeu, A. (2011). Rome Wasn't Digitized in a Day: Building a Cyber infrastructure for Digital Classics: Council on Library and Information Resources. Retrieved 19 Oct, 2015, from <http://www.clir.org>.
- Barret, E. (2014). *Quel rôle pour les bibliothèques dans les humanités numériques?*, Mémoire Master 2, Sous la direction d'Isabelle Westeel Directrice du SCD de l'Université Lille 3.
- Berry, D. M. (2011). The computational Turn: Thinking the Digital Humanities. *Culture Machine*, 12(2).
- Borgman, Ch. (2000). The invisible library: Paradox of the global information infrastructure. *Library Trends*, 57(4), 652-675.
- Calhoun, K. (2014). *Exploring Digital Libraries: Foundation, practice, prospects*: Facet Publisher, London.
- Common Language Resources and Technology Infrastructure (CLARIN). (2015). CLARIN: Common Language Resources and Technology Infrastructure. Retrieved 19 Oct, 2015, from <http://eudat.eu/communities/clarin-common-language-resources-and-technology-infrastructure>.
- Coble, Z., Potvin, S., & Shirazi, R. (2014). Process as Product: Scholarly Communication Experiments in the Digital Humanities. *Journal of Librarianship and Scholarly Communication*, 2(3), 11.
- Crane, G., Fuchs, B., & Iorizzo, D. (2007). The Humanities in a Global e-Infrastructure: A Web-Services Shopping List. Paper presented in *UK e-Science All Hands Meeting*.
- Dacos, P. M. (2010). Manifesto for the Digital Humanities. Retrieved 19 Oct, 2015, from <http://tcp.hypotheses.org/41>.
- Digital Cultures and Universality in Knowledge Organization (DIGIKO). (2013). *Pré-Projet*

- DIGIKO*, un projet de recherche (Digital Cultures and Universality in KO) proposé avec quatre partenaires (la France en tant que pilote, le Royaume-Uni, les Pays Bas et l'Allemagne, conjointement avec un panel d'experts internationaux. Une version actualisée a été présentée durant la conférence Knowscape, à La Hague, le 23 octobre 2013.
- Digital Research Infrastructure for the Arts and Humanities (DARIAH). (2015). What is DARIAH? Retrieved 19 Oct, 2015, from [www.dariah.eu](http://www.dariah.eu).
- Dunn, S., & Blanke, T. (2008). Next Steps for E-Science, the Textual Humanities and Vres'. *D-Lib Magazine*, 1(2).
- El Hadi, W. M., & Favier, L. (2014). Bridging the Gaps between Knowledge Organization and Digital Humanities. *Advances in Knowledge Organization*, 14.
- El Hadi, W. M. (2013). Faceted classifications as a model for universal classifications. *SRELS Journal of Information Management*, 50(6), 721-734.
- El Hadi, W. M. (2015). Cultural Interoperability and Knowledge Organization Systems, Keynote address in *ISKO-Brazil Conference*, Marília (São Paulo Brazil), 17-19 September 2015.
- Favier, L. & El Hadi, W. (2013) "L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir ?", in F. Papy (ed), *Problématiques émergentes dans les sciences de l'information*. Paris, Hermès.
- Huma-Num. (2015). À propos de Huma-Num. Retrieved 19 Oct, 2015, from <http://www.huma-num.fr/>.
- Joint Information Systems Committee (JISC). (2015). Publications. Retrieved 19 Oct, 2015, from [http://www.jisc.ac.uk/publications/publications/pub\\_vreroadmap.aspx](http://www.jisc.ac.uk/publications/publications/pub_vreroadmap.aspx)
- King's College London. (2015). High Throughput Humanities e-Research (HiThER). Retrieved 19 Oct, 2015, from <http://www.kcl.ac.uk/innovation/groups/cerch/research/projects/completed/hither.aspx>
- Koltay, T. (2014). Information Science in the Mirror of the Digital Humanities: Some Epistemological Observations. *Zagadnienia Informacji Naukowej (ZIN). Information Studies*. 52(2), 23-31.
- Moulin, C., Nyhan, J., Ciula, A. (2011). *Research Infrastructures in the Humanities*, European Science Foundation, Strasbourg.
- Palermi, R., & Polity, Y. (2002). Dynamiques de l'institutionnalisation sociale et cognitive des sciences de l'information en France. *Les origines des sciences de l'information et de la communication: regards croisés. Presses Universitaires de Septentrion, Villeneuve d'Ascq*, 95-123.
- Rockwell, G. (2011). Building a Humanities Cyberinfrastructure. Retrieved from <http://blogs.ischool.utexas.edu/f2011dh/tag/cyberinfrastructure/>.
- Schreibman, S. (2004). Ray Siemens, and John Unsworth. *The Digital Humanities and Humanities Computing: An Introduction. A Companion to Digital Humanities*. S. Schreibman, R. Siemens, and J. Unsworth (Ed.): Oxford, UK: Blackwell.
- Slavic, A. (2011). Classification revisited: a web of knowledge. *Innovations in information retrieval: perspectives for theory and practice*. A. Foster and P. Rafferty P. (Ed.). London: Facet, 23-48.
- Slavic, A. (2013). Classification in knowledge transfer between languages, domains and collections in humanities: from legacy collections to the Web of linked data. Paper presented in the *Presentation at European Seminar*, Ecole Doctorale, 12th April.
- Svensson, P. (2010). The landscape of digital humanities. *Digital Humanities*, 4(1).

THATCamp Paris. (2010). The Humanities and Technology Camp. Retrieved 19 Oct, 2015, from <http://tcp.hypotheses.org/category/thatcamp-paris-2010>.

Tasovac, T. (2015). Not All Texts Were Born Equal: Toward DH-friendly Digital Libraries, European One-Day Seminar, University of Lille 3, Lille 27 May 2015.

Warwick, C., Terras, M., Galina, I., Huntington, P., & Pappa, N. (2008). Library and information resources and users of digital resources in the humanities. *Program*, 42(1), 5-27.

Warwick, C., Terras, M., Nyhan, J. (Eds). (2012). *Digital Humanities in Practice*. London: Facet Publisher.

Wusteman, J. (2008). Editorial: virtual research environments: what is the librarian's role? *Journal*

## Appendices

Appendix A. "Modern" is used here in contrast with a very early history of the DH reported for by Moulin et al., (2011, p. 3) who traced back the history of Research Infrastructures (RIs), as one of the major components of DH. "It was in the field of Humanities that the idea of RIs 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 3rd century B.C., the imperative to collect, organize 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".

Appendix B. The use of this term DH, may appear here as anarchism but its scope and coverage at that time can be compared to its modern and current understanding

Appendix C. Charles the Fifth, Library, <http://www.europeanaregia.eu/en/historical-collections/library-charles-v-family>, the reconstruction of a model of the princely libraries. The library had gathered and organized in the tower of the Falconry in the Louvre, is not only fabulous one collection of books - more than 900 volumes in 1380 - for the king and for his councilors.

Appendix D. Digital RIs were developed earlier in the hard sciences than in the Humanities and currently receive a larger proportion of funding.

Appendix E. *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 (as cited in Moulin et al., 2011, p. 5).

Appendix F. The European Commission Framework Programme, MERIL project (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.

Appendix G. ISIDORE,

<http://www.rechercheisidore.fr/> is a platform enabling research and access to the digital and digitized data from the on research human and social sciences (SHS). ISIDORE harvests metadata and indexes the digital data by enriching the resources with the terms of scientific reference tables

Appendix H. THATCamp stands for "The Humanities and Technology Camp." It is an "unconference": that means, an open, inexpensive meeting where humanists and technologists of all skill levels learn and build together in sessions proposed on the spot. The most interesting feature of the THATCamp is that it's open and online: participants make sure to share their notes, documents, pictures, and other materials from THATCamp discussions before and after the event on the web and via

social media <http://tcp.hypotheses.org/category/thatcamp-paris-2010>.

Appendix I. The Arts and Humanities e-Science Support Centre (AHeSSC) at King's College London organized a workshop at the 2007 e-Science All Hands Meeting in Nottingham

([http://www.allhands.org.uk/news/textgridws\\_call.cfm](http://www.allhands.org.uk/news/textgridws_call.cfm)), entitled "Text and Grid: Research Questions for the Humanities, Sciences and Industry". This report summarizes the main points that emerged from that workshop, and outlines a medium-term research agenda for how that process of engagement can proceed. Four papers were given in the workshop. The first, presented by Dolores Iorizzo and Brian Fuchs of Imperial College, gave an overview of what the humanities need from a global grid infrastructure in reference to Crane, G., Fuchs,

B., Iorizzo, D. 2007: The Humanities in a Global e-Infrastructure: A Web-Services Shopping List. UK e-Science All Hands Meeting 2007, Nottingham, UK, September 2007. The major ideas related to VREs I sketched in this paper are drawn from Crane et al., (2007), as cited in Dunn & Blanke (2008)

Appendix J. Suggested by the Association of College and Research Libraries, (ACRL), Scholarly Communications Committee, 2003. The ACRL Scholarly Communication Initiative was distinct from the similarly-titled Scholarly Communication Institute, funded by the Mellon Foundation and developed by the Council on Library and Information Resources, University of Virginia Library, and Dartmouth College Library (see also Cole et al., 2014, p.1137)