FEEDING DIGITAL HUMANITIES: THE DM2E AND JUDAICA EUROPEANA PROJECTS

Dov Winer

Scientific Manager, Judaica Europeana

European Association for Jewish Culture, London, UK

and MAKASH Advancing ICT Applications, Israel

dov.winer@gmail.com

Paper presented at the EVA Conference on Electronic Imaging and the Visual Arts

Florence, 7-8 May 2014

Abstract

ICT support for humanities scholarship can benefit from: (1) Increased availability of relevant digitised contents through digital libraries initiatives. (2) The development of tools that enable researchers to enrich and analyse such contents and integrate them in their scholarly workflow. (3) Virtual Research Environments that integrate contents, tools and infrastructures in specific communities of practice. Judaica Europeana seeks to enhance Jewish Studies by integrating such developments with the support of the DM2E project. Judaica Europeana has to-date aggregated more than 5 million digital cultural objects while DM2E has built tools and communities that enable humanities researchers to work with manuscripts in the Linked Open Data Web.

INTRODUCTION

We live now in a ubiquitous ICT environment that has a deep impact on the way research in the Humanities – History, Literature, Philosophy, Archaeology, and the Arts – is carried out. Here we review initiatives that are concerned with the provision of digitised contents and adequate computerized tools that would enable scholars to operate in such environments. Our content focus relates to a specific area - that of Jewish Studies. We conclude this review arguing that a critical element for the effectiveness of such contents and tools is the establishment of Virtual Research Environments informed by the know-how required for the successful operation of Communities of Practice (CoP).

DIGITAL HUMANITIES

Digital humanities projects are using tools like 3-D mapping, electronic literary analysis, digitization, and advanced visualization techniques in interdisciplinary research that aims to shed new light on humanities research. With online publishing and virtual archives, creators and users experiment and interact with source materials in ways that yield new findings, while also facilitating community building and information sharing.

The concept of scholarly primitive has been very productive in the development of digital humanities tools. This concept was first introduced by John Unsworth [1] to refer to "some basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation." He suggests the following primitives: Discovering, Annotating, Comparing, Referring, Sampling, Illustrating, Representing. Other initiatives have adopted this concept and refined it for their purposes. They include the Bamboo Project Scholarly Practice by Mazover (2008-2013) [2]; the OCLC Scholarly Information Practices by Palmer et. al (2009) [3] ; the report by Blanke and Hedges (2013) [4] on building institutional infrastructure for the humanities at King's College London; and the ongoing work of the Virtual Competence Center of DARIAH by Constantopoulos and Munson (2013) [5].

The project Digital Manuscripts to Europeana (DM2E) defined the requirements for the tools it is developing to support scholarly work in a Semantic Web Linked Data environment building upon the above mentioned approaches as well as upon previous work of Bamboo and Research Space. DM2E incorporates the model for the scholarly work flow developed by Stefan Gradmann [6, 7]. Data from Europeana as well as from other sources (including the original digitised content in the case of DM2E) are made part of a digital curation workflow enabling a number of specialized functions that cannot reasonably be implemented in a generic, multi-purpose platform like Europeana. This includes functions such as semantic annotation, text collation, text mining, data linking, combining digital content and metadata and augmenting scholarly knowledge based on further contextualising objects in the Linked Data cloud [8]. The technical platform provided by task 3.2 of WP3 in DM2E is now fully operational – see tutorial [9] and support such functionalities. The two tools at the heart of the DM2E scholarly environment are **Pundit** and **Ask**. **Pundit** is a semantic annotation tool that enables researchers to create annotations in Linked Open Data. **Ask** is a web service that enables people to create and query notebooks based on annotations made in Pundit.

AGGREGATED JEWISH CONTENT

Judaica Europeana sought to achieve four main goals: provide substantial digitised resources for Europeana; increase the reuse of such resources; disseminate the use of standards assuring the compatibility of digital contents to Europeana requirements and demonstrate semantic interoperability; and involve the relevant community of knowledge to demonstrate the added value of digital access in the thematic domain for scholarship and curatorial purposes. The project originally focused on the theme of "Cities": Jewish urban expressions may be outlined graphically from a community core to individual expressions: the ways Jewish communities managed their internal affairs (mutual help, education, politics, theatre, music, newspapers); through the Jewish expression in the urban landscape, occupations and enterprises seen by their neighbors as characteristic of Jews; and finally, in their fully individual expression as celebrities. These plentiful expressions of cultural creations are well documented in thousands of photos, films, books, pictures, documents, texts, works of art, monuments, archaeological excavations, buildings, and cemeteries from all over Europe. Judaica Europeana aggregated so far more than five million cultural heritage objects from institutions in Europe, North America and Israel. Judaica Europeana focus now is in developing means for the enrichment of its metadata enabling the intelligent contextualization of the digitized objects grounded on the extraordinary potential of the new Europeana Data Model (EDM).

Europeana is pioneering the application of a new concept of the Web that enables the seamless integration of heterogeneous databases in one searchable space of Linked Data. It extends the web of documents to a web of data. This technology applies web-based standards for encoding datasets and linking them to other published datasets, so that applications can exploit data from many different sources.

The Europeana Data Model (EDM) [10] is a schema for structuring the data that Europeana is ingesting, managing and publishing. The EDM not only supports the full richness of the content providers' metadata, but also enables data enrichment from a range of third party sources. For example, a digital object from Provider A may be contextually enriched by metadata from Provider B. It may also be enriched by the addition of data from authority files held by Provider C, and a web-based thesaurus offered by Publisher D.

Judaica Europeana developed a detailed work program - Winer [11, 12] - for expressing relevant vocabularies in the required format and employing them in the Europeana context. These vocabularies have the potential for enriching the metadata descriptions of the

digitized resources providing contextual meanings for objects that otherwise may lack any substantial significance for the user. Vocabularies concerning Names, Places, Periods, General interest and Regional interest vocabularies are described and the tasks to be carried outlined. Some parts of the Thesaurus of the Israel Museum Jerusalem are already operational and there is ongoing work by Dr. Kai Eckert [13] from the University of Mannheim in publishing the YIVO Encyclopedia of Jews in Eastern Europe as RDF/Linked Data and using it to enrich the metadata already uploaded to Europeana.

VIRTUAL RESEARCH ENVIRONMENTS (VRE)

Christopher Brown [14], programme manager of Digital Infrastructure at JISC defines VRE as comprising a set of online tools and other network resources and technologies interoperating with each other to facilitate or enhance the processes of research practitioners within and across 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.

Michael Fraser [15] defines Virtual research environments (VREs) as comprising digital infrastructure and services which enable research to take place. The VRE helps to broaden the definition of e-science from grid-based distributed computing for scientists with huge amounts of data to the development of online tools, content, and middleware within a coherent framework for all disciplines and all types of research. A VRE is best viewed as a framework into which tools, services and resources can be plugged.

Guus van den Brekel [16] key recommendations for VRE development suggest that they should be built Bottom-up and with a focus on researcher's **needs** and specific research communities

Carusi and Reimer (2010) [17] argue that fundamentally, the most important point to have emerged from their comprehensive study on VREs is that they need to be conceptualised as **community building projects rather than technology projects**. Another of their conclusions is that VREs have the potential to benefit research in all disciplines at all stages of research. The access to data, tools, computational resources and collaborators that VREs facilitate leads to **faster research results** and **novel research directions**.

DARIAH [18] the Digital Research Infrastructure for the Arts and Humanities is now seeking to develop the DARIAH VRE. They state that since nearly every research project requires an amalgamation of specific types of scholarly data in combination with generic or specific tools for data querying, enrichment, and exploration, it is of utmost importance to provide facilities for the definition and implementation of a virtual research environment (VRE). Such environments are constructed upon the technical means provided by the e.infrastructure and based on a closed collaboration between scientists.

These diverse experiments and probings of VREs allocate a critical role to the collaborative and social elements as the main factors in a VRE eventual success or failure. We argue then that the development of a VRE should be oriented by an appropriate conceptualization of Communities of Practice (CoP) and the way they evolve and are maintained.

Wenger (1998) [19] defines Communities of Practice as groups of people who share a concern or a passion for something they do and who interact regularly to learn how to do it better. He stresses that a community is a Community of Practice only if it has the following three dimensions. First, *mutual engagement* defines the community of practice with a certain shared domain for the common engagement. It enables engagement of diverse

members of the community who possess mutual relationships among themselves. Second, *joint enterprise* suggests that it be important to set the common regulations in the community. Members of the CoP should cooperate and help each other, and share information. The underlying rules can be represented or realized in various templates and their operations to support such dimension. Third, a *shared repertoire* is provided to and created by the community members. The practice is an important factor in the CoP and its results in form of resources, experiences, multimedia, stories, and tools etc. can be shared among the community members.

Klamma et al. (2005) [20] wrote, based on Wenger [19] that Communities of Practice (CoP) of scholars are characterized by common conventions, language, tool usages, values and standards. The development of a common practice which defines the community comprises the negotiation of meaning among the participants as well as the mutual engagement in joint enterprises and a shared repertoire of activities, symbols and artifacts. A CoP is inseparable from issues of (individual and social) identity. Identity is mainly determined by negotiated experience of one's self in terms of participation in a community and the learning process concerning one's membership in a CoP.

Klamma [20] adds further that systems supporting CoP should aim at providing scholars with a flexible (online) environment to create, annotate and share media-rich documents for their discourses by relying on metadata standards. Those standards allow scholars to create exchange and collaborate on multimedia artifacts and collections between communities across disciplines and distances. For the purpose of comparing different approaches in developing systems that support CoP, Klamma details the requirements for such systems. These include: (1) **Collective hypermedia artifact repository** (2) **Transcription and semantic enrichment of data**. Collaborative learning is encouraged by annotations accessible to and possibly transcribed by other scholars. (3) **Search and retrieval** – all the processes of retrieval, manipulation and management should be accessible as objects in the repository. (4) **Community management** – flexible, providing with access rights on different levels and for different roles. (5) **Personal and group collections**. The need for fluid archives on personal and group level, which can be navigated, sorted and annotated by community members. (6) **Hypermedia and interrelation graphs.** The expression of the full context and complexity of objects requires the visual representation of knowledge by hypermedia graphs. (7) **Ontologies**. These are applied for information brokering and provide users with content deemed the most suitable in a particular context.

Ralf Klamma [21] lists as features of CoP: (1) Situated Learning: Learning is a function of the activity, context and culture in which it occurs (i.e. it is situated) (2) Informal and Co-located: The gradual acquisition of knowledge and skills learned from experts in the context of their everyday activities. (3) Group Knowledge: Knowledge is mediated through social interaction and collaboration in the group (4) Legitimate Peripheral Participation (LPP): Novices move from peripheral to full participation as they gain legitimacy in the group. He then defines a CoP in the following terms: (1) What it is about: The activity/body of knowledge that the community has organized itself around - a joint enterprise. (2) How it functions: How people are linked through their involvement in common activities - mutual engagement. (3) What it produces: The set of resources the members of a CoP build up over time - their shared repertoire. He refers to Participation, Reification and Community Awareness as critical elements in CoP; Knowledge Management (KM) in form of CoP implies in (1) Turning utterances and data into needs and knowledge (2) Finding out what is really happening in your organization.

Such awareness that organizational development and other social technologies are critical in achieving well functioning Communities of Practice has lead to the development of guides e.g. one offered by EDUCAUSE [22] that integrates the experiences of several higher education institutions. Gilbert Probst and Stefano Borzillo [23] offer a analyse why communities of practice succeed and why they fail.

CONCLUSION

The Europeana projects Judaica Europeana and Digitised Manuscripts for Europeana have dramatically advanced the availability of two critical resources able to support Digital Humanities in Jewish Studies: relevant digitised contents and cutting edge semantic web tools based on the concept of scholarly primitives. The enrichment of such resources enabling their best use in the DM2E environment has been defined with the identification or relevant vocabularies and encyclopedias and their ongoing publication as RDF/Linked Data.

The main challenge ahead is to engage the relevant communities of practice that deal in Jewish Studies and should be the primary target population for using these twinned resources. This challenge should be informed by the social technologies identified as crucial for the proper functioning of Virtual Research Environments within a Communities of Practice approach.

ACKNOWLEDGEMENTS

References

[1] J. Unsworth, "Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?" part of a symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London, May 13, 2000. Retrieved on March 20th, 2014 from
http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

[2] S. Mazover, Editor, "Project Bamboo Scholarly Practice Report" (2008-2013). Retrieved on March 20, 2014 from:
https://wikihub.berkeley.edu/display/pbamboo/Project+Bamboo+Scholarly+Practice+Report

[3] C. L. Palmer, L.C. Teffeau, C. M. Pirmannn, " OCLC: Scholarly Information Practices in the Online Environment", OCLC, 2009. Retrieved on March 2[0th],2014 from:
http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf?urlm=162919

[4] T.Blanke and M. Hedges, "Scholarly primitives: Building institutional infrastructure for humanities e-Science", *Future Generation Computer Systems*, vol. 29, pp. 654-661, 2013.

[5] P. Constantopoulos and M. Munson (VCC Convenors), "Scholarly methods ontology and DH resource portal", Presentation at the DARIAH General VCC Meeting, Copenhagen, 5-6/9/2013. Retrieved on March 20th, 2014 from:
http://dariah.eu/fileadmin/Documents/ontologyDay1.pdf

[6] S.Gradmann, "Digital Humanities Research, Europeana and the Embedded Semantic Research Library: A vision for KU Leuven's Library System (and for others?)", Presentation at the Cortona Digital Humanities Seminar, Scuola Normale Superiore di Pisa, January 18th, 2013. Retrieved on March 20th, 2014 from:

http://www.slideshare.net/netseven/digital-humanities-research-europeana-and-the-embedded-semantic-research-library

[7]   S.Gradmann and S.Schreibman et al., "Beyond Infrastructure: Further Modeling the Scholarly Domain", paper presented at the conference Digital Humanities 2013, Lincoln Nebraska 19.7.2013. Presentation retrieved on March 20th, 2014 from: http://www.slideshare.net/gradmans/20130719-dh2013-beyondinfrastructure

[8]   Digitised Manuscripts to Europeana, Annex I, Description of Work. Project Acronym: DM2E, Grant agreement for CIP-Best practice networks number: 297274, Date2012-01-22 (page 29).

[9]   DM2E Work package 3: Digital Humanities research and engineering. Pundit and Ask; Tutorials. Retrieved on March 20th, 2014 from: http://dm2e.eu/digital-humanities/

[10]   A. Isaac (Editor), "Europeana Data Model Primer", Europeana V1.0, 14.7.2013. Retrieved on March 20th, 2014 from: http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5 . For additional EDM documents: http://pro.europeana.eu/edm-documentation

[11]   D. Winer, "Judaica Europeana D2.5: Semantic Interoperability Report with representation of selected controlled vocabularies in RDF/SKOS." September 2011. Retrieved on March 20th, 2014 from: http://www.judaica-Europeana.eu/docs/D2-5_Semantic_interoperability_report.pdf

[12]   D. Winer, "Judaica Europeana: an infrastructure for aggregating Jewish content", *Judaica Librarianship,* vol. 18, March 2014. http://ajlpublishing.org/jl/

[13]   K. Eckert, "Judaica Links": http://judaicalink.org/about-us

[14]   C.Brown, "Guide: Implementing a virtual research environment (VRE)"m JISC, April 2, 2013. Retrieved on March 21st 2014 from: http://www.jiscinfonet.ac.uk/infokits/vre/ . See also presentation at: http://www.slideshare.net/chriscb/jisc-vreresearch-tools-presentation?from_search=2

[15]   M. Fraser, "Virtual research environments: overview and activity", *Ariadne*, vol.44, July 2005. Retrieved on March 19, 2014 from: http://ora.ox.ac.uk/objects/uuid%3A3f84ad77-9203-43cf-b308-a4f3b76817d3/datastreams/JOURNAL

[16]   G. Brekel, "Virtual Research Networks: towards Research 2.0", Central Medical Library, UMCG. Retrieved on March 19, 2014 from: http://www.slideshare.net/digicmb/virtual-research-networks-towards-research-20?from_search=1

[17]   A. Carusi and T. Reimer, "Virtual Research Environment Collaborative Landscape Study – a JISC funded project", Oxford e-Research Centre, University of Oxford and Centre for e-Research, King's College London, Jan. 2010. Retrieved on March 21, 2014 from: http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf

[18]   DARIAH Digital Research Infrastructure for the Arts and Humanities, Virtual Research Environments. Retrieved on March 21st 2014 from: http://www.dariah.eu/activities/research-and-education/virtual-research-environments.html

[19]   E. Wenger, "Communities of practice - Learning, meaning, and identity", Cambridge University Press, 1998.

[20]   R. Klamma, M. Spaniol, and M. Jarke, "MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies", *Informing Science Journal*, vol. 8, 2005. Retrieved on March 19th, 2014 from: http://www.inform.nu/Articles/Vol8/v8p003-038Klamma.pdf

[21]   R. Klamma, "Social Software in Knowledge Management of Organizations", Presentation at .CONTici, Tallinn, Estonia November 2008. Retrieved on March 19, 2013 from: http://beamtenherrschaft.blogspot.co.il/2008/11/social-software-in-knowledge-management.html

[22]   EDUCAUSE, "Community of Practice Design Guide A Step-by-Step Guide for Designing & Cultivating Communities of Practice in Higher Education". Retrieved on March 21st 2014 from: http://net.educause.edu/ir/library/pdf/nli0531.pdf

[23]   G. Probst and S. Borzillo, " Why communities of practice succeed and why they fail", *European Management Journal*, vo. 26, pp. 335–347, 2008.