

The CENDARI Project: A user-centered 'enquiry environment' for modern and medieval historians
[Poster]

Jakub Beneš, Alexander O'Connor, Evanthia Dimara

► **To cite this version:**

Jakub Beneš, Alexander O'Connor, Evanthia Dimara. The CENDARI Project: A user-centered 'enquiry environment' for modern and medieval historians [Poster]. Digital Humanities, Jul 2014, Lausanne, Switzerland. pp.434-436. hal-01061909

HAL Id: hal-01061909

<https://hal.inria.fr/hal-01061909>

Submitted on 8 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DH 2014 Conference Abstract

Poster title: “The CENDARI Project: A user-centered ‘enquiry environment’ for modern and medieval historians”

Presenters: Jakub Beneš (University of Birmingham), Alexander O’Connor (Trinity College Dublin), Evanthia Dimara (INRIA-Institut national de recherche en informatique et en automatique)

Poster abstract:

Introduction

This poster showcases the *Collaborative European Digital Archive Infrastructure* (CENDARI) project, a European-Union Seventh Framework project, which takes an innovative approach to digital historical research as well as to data integration and curation. The presenters, representing diverse disciplinary perspectives, explain their converging paths toward the creation of a research infrastructure that will further historical inquiry in the digital age.

The Digital Humanities remains an exotic garden to many historians. While software developers have focused on sophisticated analytical tools that require large datasets and pointed research questions, historians often consider themselves unready to use such tools or regard them as superfluous once they have gathered and organized sufficient research data. To many, digitization projects seem too narrowly conceived to represent disciplinary breakthroughs, in part because they typically neglect archival sources. Moreover, immense national and institutional asymmetries exist in efforts to further digital history.¹

The CENDARI project overcomes some of these constraints. On the most basic level, it is integrating data and metadata from archives, libraries, and museums across Europe relevant to the project’s two historical domain test cases: Medieval culture and World War I. In order to further transnational and comparative research, and to overcome entrenched historiographical and digital asymmetries, the project includes eastern and southern European repositories (‘hidden archives’ to many historians) along with the more visible western European institutions.

From a computer science perspective, the relevant data is dizzyingly heterogeneous in terms of languages, formats, level of granularity,

¹ In this, digitization initiatives have generally reflected supply and opportunity rather than demand. See Melissa Terras, “Digitization and digital resources in the humanities” in Claire Warwick, Melissa Terras, and Julianne Nyhan, eds., *Digital Humanities in Practice* (London, 2012).

completeness, encoding standards, annotation schemes, etc. Therefore CENDARI has implemented a capacious approach to data integration and curation based on the concepts of 'data space' and 'blackboard'. This will produce a flexible and interactive digital ecosystem, underpinned by various ontologies, that enables collaborative research using a variety of digital tools. Cooperation with the European digital humanities infrastructure DARIAH will ensure the ecosystem's sustainability.

Historians will be able to access data by pursuing their own research projects through a dynamic user interface. While the enquiry environment is focused on the initial, exploratory phases of research, it will go beyond "search and retrieval." Historians will be able to analyze data with the help of sophisticated data mining and visualization tools; they will be able to upload their own research to a personal research space, and they will be able to curate and exchange data with other researchers through annotations, tags, semantic links, and other tools. Project partners have developed this enquiry environment based on interactive participatory design sessions, domain specific "use cases", and two domain-specific "prototype projects," all designed to integrate the user's perspective while the research infrastructure is built.²

CENDARI incorporates archival data, and creates a research space where users can see projects through from finding and organizing sources to analyzing and sharing data with sophisticated tools. The project overcomes the national 'siloes' of digitization efforts and historical inquiry. Perhaps above all, it may help open digital history to the majority of professional historians, representing a major breakthrough in *digital cultural empowerment*.

Approach to data

How can CENDARI help users answer questions they did not know they wanted to ask, and how can these users then be helped to record and share the process and results of those questions? The CENDARI project offers a unique opportunity to demonstrate "serendipity through heterogeneity". There is already an enormous number of web-based tools and projects which offer the web browser digital access to archives and collections. CENDARI will not attempt to become a "big data" repository for all of them. Instead, the project should recognize that the value of scholarship is in the interlinking of different concepts, objects, collections and content to highlight insights that are not otherwise obvious. This is among the primary goals of the project: to foster serendipity in research processes as well as to support auditable, traceable research trails.

² This reflects a growing consensus in recent digital humanities scholarship that more user-centered approaches are needed in tool development. See Gibbs, Fred, and Trevor Owens, "Building Better Digital Humanities Tools: Toward broader audiences and user-centered design" *Digital Humanities Quarterly*, 2012, Vol. 6, No. 12.]; Claire Warwick, "Studying users in digital humanities" in Claire Warwick, Melissa Terras, and Julianne Nyhan, eds., *Digital Humanities in Practice* (London, 2012).

CENDARI data are heterogeneous in the origin of their sources, formats, metadata profiles, type of content they hold, methods of acquisition or creation and distribution rights pertaining to them. In some cases, data will be stored within CENDARI, such as data produced within the context of the CENDARI Archival Directory (as metadata manually edited or coming from a particular repository with links to the original sources); in other cases these will have a more transient character, e.g. if based on a search results retrieved from external system.

A design goal of the CENDARI data infrastructure is to build an interoperable data platform, overcoming various data siloes and leveraging the potential of already existing platforms and their existing data services “below the level of work”³. Additionally, CENDARI aims to reach a more detailed level of data granularity as the basis for real scholarly work and employ services that support knowledge discovery, organization and sharing.

We address the aspect of infrastructure development that embraces data diversity, i.e. the “data soup,” and takes an incremental approach to the data integration, based on the concept of Dataspaces⁴, SOA⁵ and an adapted “Blackboard” model approach⁶, while employing information extraction, NLP tools and statistical methods in order to build infrastructure components for historical research.

Approach to the Virtual Research Environment

Researcher involvement was seen as a key element in all aspects of the technical development. The partners in charge of defining the system architecture and designing the User Interface (UI) employed several methods, such as video brainstorming sessions for the creation of mockups, for understanding the user requirements and methodological needs of the target users: World War I historians and medievalists. Project historians also analyzed their own research methods, and began communicating them to technical specialists, by creating a number of scenarios drawing on concrete research

³ P. Edwards, S. Jackson, G Bowker and C Knobel, *Understanding Infrastructure: Dynamics, Tensions and Design* <http://hdl.handle.net/2027.42/49353>, last accessed 21 May 2013.

⁴ Norbert Antunes, Tatiana Malyuta, and Suzanne Yoakum Stover, “A Data Integration Framework with Full Spectrum Fusion Capabilities,” Presented at the Sensor and Information Fusion Symposium, Las Vegas, NV, August 2009, 2-3.

⁵ Krafzig, Dirk; Karl Banke, Dirk Slama: [Enterprise SOA: Service-Oriented Architecture Best Practices](#), (New Jersey, 2004), ISBN 978-0-13-146575-6.

⁶ Lee D. Erman, Frederick Hayes-Roth, Victor R. Lesser, and D. Raj Reddy. The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys*, 12(2):213–253, June 1980. See also <http://www.thecepblog.com/2008/07/20/a-brief-introduction-to-blackboard-architectures/>, 20.07.2008

inquiries. The two most detailed of these were selected to serve as “prototype projects” that constituted both real research endeavors and a means of defining the technical functionalities of the enquiry environment.

The iterative design process revealed strong user interest in a VRE centered on an advanced note-taking environment with links to the CENDARI data space, continuously enriched by historians’ notes. This result came from the conjunction of interesting findings: all the historians take notes, either on paper, in digital form, or both. From their notes, they try to resolve people (who is that person?), places, dates, artifacts, events, and organizations, among other entities. This resolution leads them to search for related entities (e.g. the family of that person, the archive holding information related to that event), until they reach a point where they have a clearer picture of a situation, or they give up for lack of information. Relating entities is a complex task not well supported by existing digital environments. Historians would like to search in their colleagues’ notes for hints, but are opposed to sharing their own notes by fear of being “scooped”. To avoid the problem, the VRE allows searching in entities contained in notes without disclosing the contents of the notes in their entirety. Brainstorming with historians revealed that they would accept sharing the entities only (with some control). Therefore, note-taking from multiple historians weaves a network of entities, creates a resource that facilitates connecting information, and allows asking appropriate colleague historians for help.

Our primary design goal is a technology that does not interrupt historians’ workflow. We propose a smooth and on-demand integration of intelligent tools, like the entity recognizer, so the researcher has full control of his project. In order to make the VRE easy to learn, our design mimics the traditional historian’s physical workspace. Based on the participatory design insights, the VRE aims to interpret the affordances of the historian’s personal library, note taking, entity highlighting, annotations or work organization to digital tools. The notion of “affordance” here implies that the appearance of the tool reveals a part of its functionality to the user. Once the researcher is able to accelerate his working rate in VRE, we enrich the workflow with individualized visualizations based on the user scenario’s queries. Our design approach is based on the researcher’s daily routine. We use an agile software development methodology to allow quick adaptation of the system to historians’ needs.

In an era in which the digital can drive much scholarly innovation, this note-taking environment meets and serves the needs of historians, who generally keep a traditional research diary or notebook. At the same time, it seems to foster new research approaches and new attitudes towards the organization and use of archival sources. Seen from a user/researcher’s perspective, the note-taking environment could therefore be an interesting platform for both organizing existing data and notes, and for envisaging new research directions. The concept of selective sharing represents a new opportunity for experiencing research work in a selected and collaborative environment which, when properly understood

and used, might boost the potential of archival work accomplished across different countries.