## Collaborative European Digital Archive Infrastructure

**Project Acronym:** CENDARI  
**Project Grant No.:** 284432  
**Theme:** FP7-INFRASTRUCTURES-2011-1  
**Project Start Date:** 01 February 2012  
**Project End Date:** 31 January 2016

<table>
<thead>
<tr>
<th>Deliverable No.</th>
<th>9.2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title of Deliverable:</strong></td>
<td>Full Release Portal/VRE</td>
</tr>
<tr>
<td><strong>Date of Deliverable:</strong></td>
<td>October 2015</td>
</tr>
<tr>
<td><strong>Revision No.:</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>WP No.:</strong></td>
<td>9</td>
</tr>
<tr>
<td><strong>Lead Beneficiary:</strong></td>
<td>INRIA</td>
</tr>
<tr>
<td><strong>Author (Name and email address):</strong></td>
<td></td>
</tr>
</tbody>
</table>
  - Jean-Daniel Fekete Jean-Daniel.Fekete@inria.fr  
  - Nadia Boukhelifa Nadia.Boukhelifa@inria.fr  
  - Emmanouil Giannisakis Emmanouil.Giannisakis@inria.fr  
  - Carsten Thiel thieli@sub.uni-goettingen.de  
  - Hanna Meiners meiners@sub.uni-goettingen.de  
  - Natasa Bulatovic bulatovic@mpdl.mpg.de  
  - David Stuart david.stuart@kcl.ac.uk |
| **Dissemination Level:** | PP = restricted to other programme participants |
| **Nature of Deliverable:** | O = other |
| **Abstract (150 words):** | Full Release Portal/VRE: This release brings together all elements of the CENDARI development, adding full connectivity through the integration layer developed by WP |
This report provides an overview of the portal used to access all the services and data of the CENDARI research environment. The portal is based on “Liferay” (http://www.liferay.com/), an open-source portal that integrates a large number of social-networking services. The CENDARI portal is available at: https://portal.cendari.dariah.eu/home
Table of Contents

Contents
Introduction ................................................................................................................ 4
Integrating Services into the CENDARI Virtual Research Environment ............... 5
Organization of the Portal........................................................................................... 7
  Research ................................................................................................................ 7
  Standardize .......................................................................................................... 17
  Lab........................................................................................................................ 18
  Collaborate ........................................................................................................... 18
  Extensions of the Portal........................................................................................ 19
**Introduction**

This report provides an overview of the portal used to access all the services and data of the CENDARI project.

As stated in the CENDARI Description of Works, the objectives of the W9 work package are:

1. To provide the user interface and human-oriented services to weave together all the technical resources of CENDARI to allow historians to conduct their research with transnational digital support.
2. To develop a portal that allows integrated discovery of and access to resources held in the participating archives and other web-accessible resources, in coherence with the architecture and functional specifications defined in WP9.
3. To develop functionalities that allow integrated discovery of and access to resources held in the participating archives and other web-accessible resources.
4. To develop an enquiry environment that (i) provides shared access to tools for integrating and processing data; (ii) provides a collaborative space for sharing and discussion; (iii) supports the research lifecycle of the targeted communities.
5. To develop visualisation tools and services to ensure that available information can be maximally exploited by researchers according to his/her research interests.

The portal is therefore an important component of the whole user experience. However, technically, it boils down to being a collector and access point to other technologies developed or used by the CENDARI project. As such, its deployment is essentially a technical issue, and involves less complications than the rest of the CENDARI tools. This is why the CENDARI project has decided to focus more on finalizing the tools themselves before the portal. The tools being under experimentations, the portal can then be deployed.

The portal is based on “Liferay” ([http://www.liferay.com/](http://www.liferay.com/)), an open-source portal that integrates a large number of social-networking services. Liferay has been chosen because several CENDARI partners were familiar with it; in particular, multiple Dariah projects are hosted on a Liferay platform, making it easier to install, maintain, and possibly adapt for CENDARI. The CENDARI portal is available at: [https://portal.cendari.dariah.eu/home](https://portal.cendari.dariah.eu/home)
Integrating Services into the CENDARI Virtual Research Environment

As a web portal, Liferay provides standard services such as a calendar, forums, a wiki, and page management. In addition, in can link to other web-based application with three levels of integrations:

**Level 1: Link-based Integration:** The simplest mechanism relies on Web links pointing to other applications listed on a Portal page. This mechanism has the advantage of simplicity but when accessing the linked application, users see a brutal change in their user interface, and cannot access the services of the portal unless they leave the portal open in another window.

**Level 2: Frame-based Integration:** another level of integration consists in inserting a CENDARI service inside a portal page using an HTML Internal Frame (iframe). We have experimented with this option with the CENDARI Note-Taking Environment (NTE) and found it unsatisfactory because the NTE uses the whole screen to display multiple components. Integrating it inside a sub-window of the portal reduces its available screen real-estate for no clear benefit.

**Level 3: Content-based Integration:** Liferay provides a rich Application Programming Interface (API), in particular to create web components for Liferay. The main protocol/mechanism for creating these components is called a “Portlet” and we have tried to see if some of our tools could be integrated that way. After multiple tries, we decided not to use Portlets because they did not provide the right level of integration. It is meant to fit tools in the Liferay ecosystem, but not to fit Liferay into an ecosystem of tools. We would have to specialise the CENDARI tools too deeply towards Liferay and decided not to.

In the end, we have decided to simply link our tools from portal pages using static images of the interfaces. This integration is very light and flexible, allowing using clickable static images to present the reachable applications to users; this impact of a brutal change will be reduced through the adoption of a consistent user-interface on linked applications. This flexibility is also important because several services are actually provided by the NTE, such as the faceted browsing, the note-taking application itself, and the Archival Research Guides. The portal provides a nice illustration for each of them, but they actually all link to the NTE with different internal web pages.
Figure 1: The CENDARI Portal

The portal integrates multiple tools and offers multiple services that we list in the next section.
Organisation of the Portal

The portal organises the CENDARI tools and services into three main groups according to their primary audience: tools and services for the historian conducting research ('Research' on portal menu); metadata schemas and ontologies for the information professional wishing to build upon the CENDARI knowledge framework ('Standardise' on the portal menu); and resources for developers wishing to build upon the code developed by CENDARI ('Lab' on the portal menu). The Research page acts as the homepage of the portal as Standardise and Lab consist of more specialised or experimental tools aimed at a smaller subset of users.

In addition, the portal menu includes:

- **Collaborate** - a dropdown menu of Liferay modules that have been implemented to encourage collaboration between researchers.
- **Docs & Guides** - the documentation and a set of guides on how to use CENDARI tools for researchers and users, developers and administrators
- **What's new** - the news feed for the CENDARI blog.
- **Who we are** - information about the CENDARI project.
- **Sign in** - a login window for registered users.

Research, Standardise, Lab and Collaborate and their associated integrated tools are discussed in more detail below.

**Research**

**Research** currently consists of links to six tools.

- “Archival Directory”: pointing to the list of archives in Europe
- “Notes”: pointing to the Notes Taking Environment
- “Search”: faceted search over all the documents indexed by the CENDARI project.
- “Medieval Sources”: allowing search over the medieval documents via TRAME.
- “Archival Research Guides”: pointing to the ARGs project in the NTE
“Ask Questions”: pointing to the list of parameterised queries using semantic-web technologies on the CENDARI platform.

**Archival Directory**: The CENDARI Archival Directory is a database of archival descriptions and collections, with a strong transnational focus. It aims to include archives, libraries and other institutions containing resources on the Medieval era and World War One. Special focus is given to “hidden” institutions i.e. those who are not well known by historians or do not have sufficient online presence. Directory is implemented by the AtoM tool, a web-based archival description software that is based on the International Council on Archives (ICA, see https://www.accesstomemory.org) standards. AtoM is the abbreviation of "Access to Memory". This software was chosen because it adheres to international standards and comes with a user friendly tool to produce archival descriptions (see Fig. 2). Informational pages of the Archival Directory are available in several languages: English, German, French, Italian, Russian, Polish and Dutch.
Notes: the Note Taking Environment is the main component of the VRE. It is a collaborative online platform that unifies a set of functionalities to support historical research workflow from data collection and analysis to synthesis and publishing. In particular, the NTE allows for organising, authoring, sense-making, searching and sharing of resources related to historical research. Organising & Authoring: Similarly to how historians organise their material on their computers, the NTE allows for creating projects within which resources are organised as notes, documents and transcripts. These resources can be created within the environment using a rich set of text editing tools (using TinyMCE editor), or by uploading existing material such as for high resolution images. Sense-making: An important feature of the NTE is the semantic enrichment of resources, either manually or automatically, via a tagging mechanism where strings of text are recognised as named entities (person, place, event etc.) and resolved by associating an entity with a reference URI (e.g. to a DBpedia repository entry). By default, user created files (i.e. notes, documents, transcripts and uploaded images) are private, while the semantic information in the
form of RDF data contained inside these resources (i.e. the entities) is public and accessible to the CENDARI project and others. This semantic information is also available via dashboard type visualisations within the platform. The NTE supports frequency visualisations for person, place and event entities and an interactive map. Such visualisations provide context for projects and facilitates sense-making of the collected data, for example in terms of time spans, geographical distributions and frequency of events. **Searching & Publishing**: all material created within the NTE is accessible via a faceted search facility and navigation.

**Search**: the Faceted search interface of CENDARI allows searching in all the resources maintained by the CENDARI project, whether created inside the project, or gathered by the project. The interface (Fig. 3) allows full-text search and shows many facets to facilitate the search and provide context to users. Users can use the facets to either get information about the results such as the number of events associated with the query, of geographical locations. The user can also use the facets to further filter the search results, selecting particular values, date ranges, or geographical bounds. Currently, the faceted search allows accessing all the documents entered in the NTE (notes, documents, entities), and all the documents indexed/ingested by the Data API. It also provides several datasets for the convenience of the researchers: all DBpedia (3.6 millions documents), and soon several encyclopaedias of World War One. The faceted search can also forward search query to TRAME, but unfortunately, the results are not faceted yet and appear in another window below the standard search panel.
**Medieval Sources:** TRAME (Texts and Manuscript Transmission of the Middle Ages in Europe) is a web-based metasearch engine which includes a variety of medieval sources, which can be searched by end user. TRAME application provides a layer of interoperability among different digital resources in the Medieval Culture domain.
Within CENDARI project the list of TRAME resources has been extended with resources from the World War One period for testing purposes. This has demonstrated its general-purpose functionality and applicability to domains other than Medieval.

**Archival Research Guides (ARGs):** Besides entities which are publically available, a special type of notes called ARGs are also accessible to the public from within the
NTE. An Archival Research Guide (ARG) is an introductory guide to archival material, structured around either one or many archival collections. Their main purpose is to offer a research framework for researchers, historians, scholars etc. who are starting a new research project. In contrast to the traditional archival guides which usually focus on one collection, institution or country, CENDARI uses a transnational approach. Depending on the author, each ARG can be divided into multiple notes which are linked to each other and they can differ in structure. ARGs support multimedia content such as images (and in the future video and sound) and a certain level of interactivity. In order for the NTE to support the ARGs, rich editing functionalities were added so that the ARG authors could format their document depending on their needs. In addition, the NTE now supports embedded images and a "Read Mode" view. By offering the ARGs through the NTE authors can tag/resolve entities, and make use of the collaborative environment and the visualisation tools.

Figure 5: The User Interface of the CENDARI Note-Taking Environment (NTE)
Although the ARGs can be achieved through the NTE, they are a valuable project on their own providing advice for historians and archivists who need to search in multiple archives spread all over Europe and sometimes the world.

**Ask Questions:** Pineapple is a semantic web service, allowing to access and search for various resources from the CENDARI Knowledge Base. While a document might be searchable via the NTE and faceted search, Pineapple understands the document as a “resource” and provides additional information about this resource such as: own properties (i.e. triples extracted from the document), which entities (Persons, Places, Events) are mentioned in a document, which other documents are mentioning this particular resource. This information is populated into the CENDARI Knowledge Base via text-mining and Indexing components (provided by WP7) such as: EAG2RDF Indexer, EAD2RDF Indexer, Named Entity Recognition Service, and NER2RDF Indexer.
Documentation and User Guides: The Documentation consists of two parts, whereas one part serves as a documentation space for the source code of the
individual CENDARI components. In this area developers are able to document their work and their code so other technicians may be better able to understand what was done and why.

The other part of the Documentation consists of the User Guides and serves as a manual for the end users of the CENDARI components. In this part there will be detailed descriptions of how to use the single tools, which possibilities were offered in the NTE and how to get to the different tools. Also, there will be FAQs for every single tool which arose from the testing workshops with the Trusted Users Group (TUG). For these workshops, training material was prepared for every single tool and component which was reused for the documentation.

The User Guides shall be accessible via a link on the portal itself.
The ‘Standardise’ tab will bring together the outcomes of the Knowledge Organisation Framework (WP6) section of CENDARI in one place, including both the work on the ontologies and the common metadata schema. This includes: visualisations of the ontology elements sets in WebVowl; access to the associated
ontology instances; ontology documentation; and guidelines and schemas for the application of the metadata standard.

**Lab**

The CENDARI Lab is envisioned as a practical and a technical corner for all users who would like to learn about and try new methods for data discovery, gathering, processing or visualisation within CENDARI or related developments.

In particular, the Lab allows accessing the source code of all the CENDARI components, for inspection and reuse in other projects.

It is the location where the tacit knowledge about the use of tools, tools development, data acquisition, data transformations, successes, failures and challenges is gathered and shared with the public.

It lists services, scripts and “how-to’s” on particular tasks or problems, and optionally, links to experimental services and applications being developed or deployed for the CENDARI project.

It is our hope that the Lab will become a referenceable exchange and information platform for tools, methods and processes supporting “digital humanities” techniques to help historian researchers in their work.

**Collaborate**

In addition to tools linked to from the portal, the portal deploys modules to encourage researcher collaboration. Initially this consists of:

- Discussions
- Calendar
- Blogs

Collaborative modules will be experimented with users and adapted as necessary.
Extensions of the Portal

We will use the portal as it is until the end of November 2015 to collect feedback from the users. We will then revise the structure of the portal according to the feedback, and possibly add or integrate new services offered by Liferay if our users ask for it.

We hope that historians will use the available tools to change their practice that remains essentially personal to a more social one with the CENDARI tools.