



HAL
open science

Basic DARIAH Services and Demonstrators

Markus Matoni, Stefan Schmunk, Carsten Thiel

► **To cite this version:**

Markus Matoni, Stefan Schmunk, Carsten Thiel. Basic DARIAH Services and Demonstrators. [0]
Göttingen State and University Library; DARIAH. 2017. hal-01574981

HAL Id: hal-01574981

<https://hal.science/hal-01574981>

Submitted on 17 Aug 2017

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.



D6.2 Basic DARIAH Services and Demonstrators

HaS-DARIAH

INFRADEV-3-2015-Individual implementation and operation of ESFRI projects

Project ID: 675570

Date: 28-07-2017

Version: 1.0



Project financed under the Horizon 2020 Programme

Grant Agreement no.:	675570
Programme:	Horizon 2020
Project acronym:	HaS-DARIAH
Project full title:	Humanities at Scale: Evolving the DARIAH ERIC
Partners:	DIGITAL RESEARCH INFRASTRUCTURE FOR THE ARTS AND HUMANITIES CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN – KNAW GEORG-AUGUST-UNIVERSITAET GOETTINGEN STIFTUNG OEFFENTLICHEN RECHTS
Topic:	INFRADEV-3-2015
Project Start Date:	01-09-2015
Project Duration:	28 months
Title of the document:	D6.2 Basic DARIAH Services and Demonstrators
Work Package title:	Basic Services Cluster 2: Supporting the Lifecycle of Digital Humanities Basic Services
Estimated delivery date:	01.07.2017
Actual date of delivery:	31.07.2017
Lead Beneficiary:	UGOE-SUB
Author(s):	Markus Matoni (matoni@sub.uni-goettingen.de) Stefan Schmunk (schmunk@sub.uni-goettingen.de) Carsten Thiel (thiel@sub.uni-goettingen.de)
Quality Assessor(s):	Ulrike Wuttke (wuttke@fh-potsdam.de) Alexander Steckel (steckel@sub.uni-goettingen.de)
Keywords:	Framework, Service Integration, DARIAH

Revision History

Version	Date	Author	Partner or Linked Third Party	Description
0.1	April 4, 2017	Matoni, Markus Thiel, Carsten	University of Göttingen	First draft
0.2	June 22, 2017	Matoni, Markus Schmunk, Stefan Thiel, Carsten	University of Göttingen	Second draft
0.3	July 20, 2017	Matoni, Markus Schmunk, Stefan Thiel, Carsten	University of Göttingen	Final draft
1.0	July 28, 2017	Matoni, Markus Schmunk, Stefan Thiel, Carsten	University of Göttingen	Final version

Table of Contents

TABLE OF CONTENTS	3
EXECUTIVE SUMMARY	4
1. INTRODUCTION AND SCOPE OF THE REPORT.....	5
2. DARIAH INFRASTRUCTURE	6
2.1. Contributing a Service to DARIAH	7
2.2. Integrating a Service into DARIAH	7
2.2.1. API Integration	7
2.2.2. Visual Integration	8
2.2.3. Service Maintenance	8
2.2.4. Prerequisites for Integration.....	9
3. DEMONSTRATORS	10
3.1. The Demonstrators.....	10
3.1.1. Redmine Project Management.....	10
3.1.2. Collabora Cloudsuite.....	11
4. CONCLUSION	13
5. REFERENCES	13

Executive Summary

This report discusses challenges the DARIAH community faces in implementing new services on a sustainable basis and illustrates its argument through the implementation of two demonstrators (Redmine project management and Collabora Cloudsuite). The two demonstrators were identified as part of the improvement of the DARIAH infrastructure by previous work in the form of the “Report on researchers’ service needs” (D6.1, Oltersdorf et al. 2016)¹.

Nature of the deliverable		
	R	Document, report
✓	DEM	Demonstrator, pilot, prototype
	DEC	Websites, patent fillings, videos, etc.
	OTHER	
Dissemination level		
✓	P	Public
	CO	Confidential only for members of the consortium (including the Commission Services)
	EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

Disclaimer

The Humanities at Scale is project funded by the European Commission under the Horizon 2020 programme. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

¹ <https://hal.archives-ouvertes.fr/hal-01351267>

1. Introduction and Scope of the Report

Within the Humanities at Scale (HaS) project², DARIAH³ tries to improve and disseminate the DARIAH infrastructure. The work package “Implementation of basic tools and services in the DARIAH infrastructure” (WP6) aims to enhance the DARIAH infrastructure by focussing on services that are easy to use and able to support future research projects in the humanities. This report describes the results of the task “Implementation of basic tools and services in the DARIAH infrastructure” (T6.2) that implemented two basic service demonstrators identified through the survey described in deliverable D6.1 “Report on researchers’ service needs” (Oltersdorf et al. 2016).

The main result is the successful implementation of the two demonstrators, see section “Demonstrators” below.

Additionally, the work package also discussed possibilities for a framework of basic services and set out definitions and guidelines for integrating services into and with the DARIAH infrastructure. This was achieved by informing a general discussion on service development, integration, and operation within DARIAH in close collaboration with DARIAH’s VCC “e-Infrastructure”⁴. The present document thus also summarises the outcomes of this discussion, which will be picked up by the DESIR⁵ project, in the Section DARIAH Infrastructure below. In particular, a solution to turn one of the demonstrators into a sustained DARIAH service has already been found.

The DARIAH-EU digital research infrastructure is a distributed and collaboratively supported European infrastructure hosted and developed by several partners and institutions. One of its core goals is to offer permanent digital services within the European Research Area and to establish a framework focussed on interfaces and data exchange between those services. It does not have a static and monolithic structure, but rather consists of a network of existing tools and services.

The work of HaS WP6, which resulted in the implementation of the two demonstrators, has the goal to identify new basic services required by scholars of the Digital Humanities – building on the previous requirements analysis done in D6.1. The demonstrators have been integrated into the existing infrastructure in order to improve the overall offers of DARIAH-EU. The efforts undertaken here form the basis for a framework that can be expanded and used in the future to implement newly identified services and – ultimately – their sustained availability.

² <http://has.dariah.eu/>

³ <http://www.dariah.eu>

⁴ <http://www.dariah.eu/activities/e-infrastructure.html>

⁵ DARIAH ERIC Sustainability Refined, Horizon 2020, Project ID 731081

2. DARIAH Infrastructure

Within DARIAH, the term *infrastructure* does not only refer to technological infrastructure but also includes the entire community of Digital Humanities in terms of the actual researchers and students as well as all materials they use and produce.

The focus of this document, however, is on the technical aspects in terms of DARIAH services. By *service* we mean a hosted software solution that can be accessed via web protocols, i.e. commonly by direct access with a browser by the user or via a programmatic Application Programming Interface (API) from within other services or tools.

The technical DARIAH infrastructure itself is distributed. This means that the services are offered by the individual national or local institutions that are involved with DARIAH, herein referred to as *providing institution*. They do this through national or otherwise pre-existing funding and in fulfillment of the national obligation towards the DARIAH ERIC. Conversely, they assume responsibility for availability and sustainability of the services they offer, generally independent from DARIAH. At the same time, they serve as first contact for support or any form of cooperation with regards to this service.

For DARIAH this approach has the advantage that there are no single points of failure and no vendor or operator lock-ins, where the entire infrastructure is and can only be managed by a small group of people. However, the absence of a single point of contact and authority also reduces the general overview and oversight. It can impact visibility and discoverability of services and simultaneously enable a culture where parallel developments towards identical goals occur without coordinated consolidation. These are also two complaints with the DARIAH infrastructure identified by previous work of this work package (D6.1), both of them now being addressed outside of WP6. Discoverability is a key focus of WP8 while an improvement in coordinating parallel efforts is addressed with increased priority by the DARIAH-EU Board of Directors.

In general, an integrated infrastructure for the Arts and Humanities aims to enlarge and enrich the possibilities its components provide in a way single solutions cannot. To this end, the individual parts are designed based on specific researchers' needs and driven by their requirements. Together they form a larger infrastructure with more usable tools. Optimally, this leads to more services accessible by everyone, a better level of interoperability among them and ultimately feeds back into the growing community.

For the following discussion, we make a distinction between *contributing* and *integrating* a service. These two concepts are orthogonal. Contributions are mostly an administrative concept used for registration and visibility to the community, while integration refers to technological aspects and design. In particular, contributions are made by DARIAH partner institutions on behalf of their country as DARIAH member, while anyone may integrate a service with DARIAH to add functionality to their own services or those offered by DARIAH.

2.1. Contributing a Service to DARIAH

As part of their obligation within the DARIAH ERIC, the national DARIAH institutions offer their existing services to the entire DARIAH network for use by scholars and researchers across Europe. These are the so called *inkind contributions*⁶ that form the technological part of the distributed DARIAH infrastructure. In this sense, any service offered by a partner to the DARIAH network is seen as a *contribution*.

In parallel, work package 5 focuses on the overall goal to establish a common DARIAH ecosystem, in order to make available, assess, and evaluate *inkind contributions* to the DARIAH infrastructure. To be accepted as *inkind contribution*, it has to – at the very least – be usable and accessible by all DARIAH users. This may involve a (manual) registration and verification process specifically for this service or with the providing partner.

2.2. Integrating a Service into DARIAH

Generally, integration refers to the underlying design of the service. This means the integration works on a technical or functional level with other DARIAH infrastructure components as well as a visual integration. The aim is to establish connectivity and interoperability of the given service. In our case of a distributed infrastructure, this can be done in one or both of the following aspects of integration:

1. Connecting the service to other DARIAH services via interfaces (API), i.e. the integration on the underlying technological level.
2. Adopting the overall design for a higher identification with DARIAH by branding the service, i.e. integration on the visual level and the user interface through the use of icons and logos, color schemes, or textual attribution.

2.2.1. API Integration

The connection to other DARIAH services via one or more APIs is a set of defined methods of communication between various software components. This can improve the quality of the service by extending its functionality and decreasing the amount of – possibly duplicate – developments. By focusing on the relevant and distinguishing features, the service's functionality and, possibly, also its security is increased.

One of the key problems in Digital Humanities research is the discoverability of research data. Data can be archived and published through various institutional and national repositories that take care of persistence and sustainable usability of the data. Through the use of persistent identifiers and established standard protocols for data exchange such as

⁶ See Article 18 of Statutes of the DARIAH ERIC,

http://www.dariah.eu/fileadmin/Documents/statutes/170405_DARIAH_ERIC_Statutes.pdf

However, the current contribution system for the DARIAH ERIC is currently under review and will be informed by the work conducted in WP5 of the HaS project.

OAI-PMH⁷, the data can be reused. Therefore, any service within a digital research infrastructure, not only for the Arts and Humanities, must put a particular focus and emphasis on providing and using the APIs required to enable sharing and reuse as well as forming a standard means of service integration.

From the technical perspective of infrastructures, the source code of the software and applications providing the services is a particularly relevant kind of research data. DARIAH develops its services in an open source approach and uses OSI⁸ approved licenses. Source code is also being archived in institutional repositories, as defined also in Raciti et al. 2017.

One central service within the DARIAH infrastructure is the Authentication and Authorization Infrastructure (AAI)⁹ which allows service creators and operators to outsource user management and all associated tasks and risks. This also addresses security concerns as there are no passwords handled by the services.

At the same time, providing these Single Sign On capabilities reduces the need for individual sign-up procedures to different services and thus improves the user experience as there is only one account and password to remember for all DARIAH services.

Overall, the connection with other DARIAH services is an advantage for both the users and the creators of the service. These connections between DARIAH services lead to a richer and improved infrastructure.

2.2.2. Visual Integration

Adapting the overall design for a higher identification (branding) with DARIAH offers the possibility to reach the whole DARIAH community, which leads to general recognition of the service and, consequently, to a higher number of possible users.

On the one hand, the visual integration may include a reference to DARIAH on the service's webpage, either in the form of a textual note or logo attribution, or more generally in following DARIAH style conventions. One example is the DARIAH-DE style guide template¹⁰, which can inform a future DARIAH-EU branding. On the other hand, it can relate to explicitly labelled interfaces with other DARIAH services and existing repositories or processing tools.

2.2.3. Service Maintenance

While the DARIAH infrastructure is distributed in the sense that all services are operated by the respective partner institutions, the community as a whole can and must create and

⁷ <https://www.openarchives.org/pmh/>

⁸ <https://opensource.org/licenses>

⁹ <https://wiki.de.dariah.eu/display/publicde/DARIAH+AAI+Documentation>

¹⁰ <https://github.com/DARIAH-DE/StyleGuideTemplate>

maintain a set of guides, recommendations, and best practices for creating, maintaining, and operating tools and services. A basis for a future official DARIAH-approved set can be the Software Quality Guidelines by CLARIAH-NL¹¹. For a further discussion of quality assessment see also Buddenbohm et al. 2017.

There are tools and services in the Digital Humanities which disappear once development stops at the end of the initial project's funding phase because they do not have a sustainability plan. DARIAH should and must address this problem and provide ways to take over these "orphaned" services. This must be a continuous process and could be modelled similar to the DARIAH-DE Service Life Cycle¹².

2.2.4. Prerequisites for Integration

Integrating a service into DARIAH is not currently subject to any formalised procedures and requirements. However, a baseline of recommended preconditions has been identified by the work done in this WP. The following preconditions ensure – among other advantages – a good usability and they work towards improving sustainability. They address language specifications, use of certain licenses, specific technological requirements, and visual integration.

Firstly, to provide services internationally within the European Digital Humanities community and beyond, providing the service and its documentation in English is a fundamental requirement. In aiming to build a research infrastructure for Open Science and Open Data, the technological foundation should be aligned with this goal and thus all software should be developed under appropriate Open Source Licences such as Apache-2.0 or EUPL-1.2. See also Raciti et al. 2017 where this policy is described.

On top of that, for all services provided and operated, contact details and support information as well as legal disclaimers are required. Furthermore, documentation for users, developers, and operators must be provided. Generally, the use of service monitoring and usage statistics is highly encouraged.

Finally, in terms of the concrete technological realisation concerning operating system compatibility and portability as well as packaging and frameworks to be used, no general recommendations can be defined. Each institution has a set of technologies they employ and can support. When creating new services, all design decisions should be aligned with these to ensure maximal sustainability.

¹¹ <https://github.com/CLARIAH/software-quality-guidelines>

¹² <https://wiki.de.dariah.eu/x/nyirAg>

3. Demonstrators

The selection of the two demonstrators produced by this work package was directly influenced by the result of task 6.1, in particular the deliverable D6.1 (Oltersdorf et al. 2016).

The survey conducted by Humanities at Scale and summarized in D6.1 identifies a number of important basic services:

- Research data repository
- Collaborative editing
- Project management
- Code hosting and issue tracking
- File syncing and file sharing

The selection of demonstrators for HaS aims to address these needs. However, due to the long list of existing and established data repository solutions and the limited scope of the project, implementing such a service as a demonstrator without long-term commitment and sufficient resources was not a viable option.

Therefore, the work focussed on addressing the remaining four aspects with the two demonstrators. This decision though, does not prohibit supporting the integration of already existing repositories into the DARIAH infrastructure.

3.1. The Demonstrators

3.1.1. Redmine Project Management

The open source software Redmine¹³ offers functionality for *project management*, *code hosting* and *issue tracking*. DARIAH-AT partner ACDH (Austrian Center for Digital Humanities) has been successfully using Redmine for all three purposes for several years. DARIAH-DE partner GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung) has been using (and offering for all of DARIAH) Chili project¹⁴, a now defunct redmine fork offering the same feature set, and is currently preparing to migrate to a compatible solution.

Building on the experience of the two institutions who jointly chair VCC-1, the decision fell on Redmine due to its ongoing maintenance by a large existing development community. Building on the existing documentation and standard frameworks used, the present work package focussed on integrating Redmine with the DARIAH AAI. Following the successful

¹³ <http://www.redmine.org/>

¹⁴ <https://www.chiliproject.org/>

implementation of an authentication plugin, ACDH started offering the service for DARIAH through eduGAIN and simultaneously to CLARIN. As a DARIAH partner institution and Austrian coordinating institution, ACDH will be offering Redmine sustainably as a DARIAH service for project management and development.¹⁵

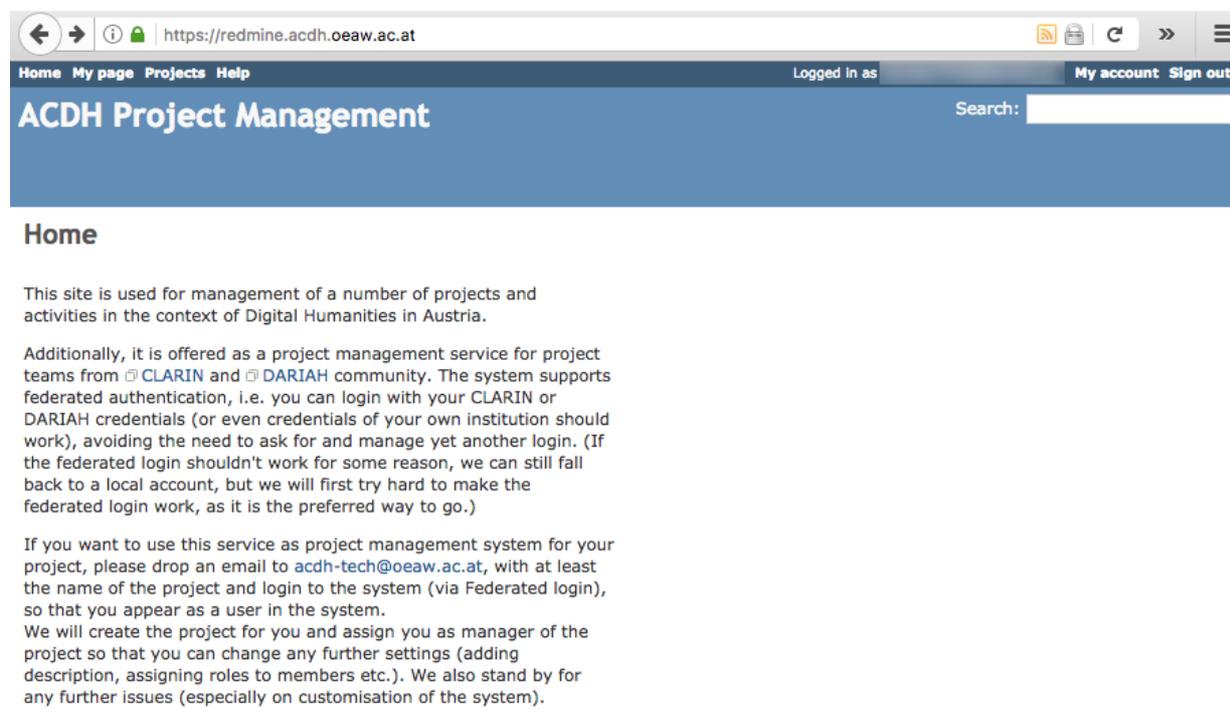


Figure 1: Screenshot of the ACDH Redmine Service (<https://redmine.acdh.oeaw.ac.at/>)

3.1.2. Collabora Cloudsuite

Collabora Cloudsuite¹⁶ is an implementation of LibreOffice¹⁷ as a web based collaborative office suite for real time collaboration, similar to Google Docs¹⁸, and is also available as an Open Source version. The request for an alternative to Google Docs that is hosted by a DARIAH partner was voiced clearly in the survey conducted by HaS and is still regularly discussed within the DARIAH community on various occasions. This need stems from a general uneasiness with storing possibly sensitive data outside of the jurisdiction of European data protection laws, which can even be found as strict prohibitions in institutional ICT regulations.

Therefore, the implementation of Collabora Cloudsuite as a demonstrator focused around the question of secure data access and reducing the risk that third parties could misuse or lose data. It is not intended to replace existing solutions such as the commercial Google

¹⁵ The service is available at <https://redmine.acdh.oeaw.ac.at/>

¹⁶ <https://www.collaboraoffice.com/code/>

¹⁷ <https://www.libreoffice.org/>

¹⁸ <https://www.google.com/docs/about/>

Docs service. The demonstrator has been set up as a test instance with limited computational resources, which can be tested for suitability until the end of 2017, at which point DARIAH will need to evaluate whether there is sufficient demand to finance the service as a permanent offer.

The realisation of AAI integration was solved by DARIAH-DE partner DAASI International in coordination with HaS, while the work within the work package focused on the realisation as a permanent DARIAH service. This proved to be non-trivial, as the requirements for operating the service are considerable and the technology is still under heavy development. Due to the nature of file synchronisation and online editing, both the traffic and storage as well as the processing power required to provide the service are considerable. Therefore, and in contrast to Redmine, each additional user creates a measurable increase in hardware requirements. This proved to be a blocking concern when discussing the option of hosting the service with national institutions. There appears to be a general hesitance to commit unclear and possibly significant resources without having long-term financial security.

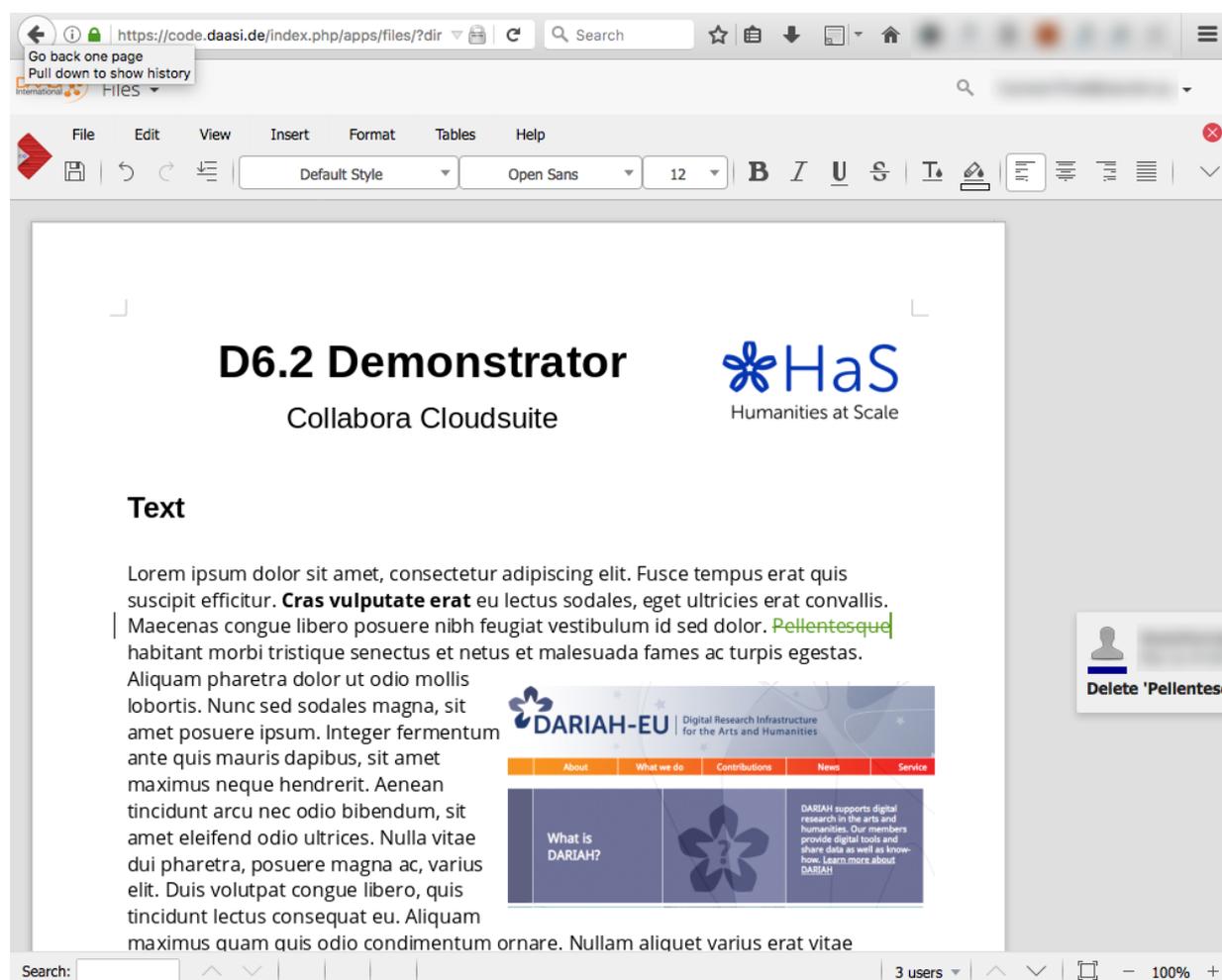


Figure 2: Screenshot of the Collabora Demonstrator (available at <https://code.daasi.de/> until the end of 2017)

4. Conclusion

Implementing the demonstrators as DARIAH services led to the identification of two major concerns relating to service operation. First, there is no central authority managing and deciding about the DARIAH portfolio. Second, recruiting volunteer institutions to operate services without a business model can be challenging.

By its distributed nature, the DARIAH infrastructure does not have a central coordinating authority that decides on its portfolio and strategic plans to extend and improve its overall possibilities. Rather, the DARIAH member countries contribute the services their partner institutions operate to form the DARIAH infrastructure. Therefore, the decision to integrate a service is always up to random volunteers addressing national requirements.

As of July 2017, no final business model exists to finance a DARIAH service that is not part of a national roadmap, such as those implemented in EU-wide efforts like Humanities at Scale. Nevertheless, discussions have been started with the relevant DARIAH bodies, in particular with the Senior Management Team and the Board of Directors to address the question of how essential services can be sustained in ways other than a national partner taking on the responsibility without payment. This is also being actively investigated as part of the research into Service Level Agreements, where the question of compensation for assurances and increased effort plays a fundamental role. The DESIR¹⁹ project, which is also tasked with creating business plans, is already picking up this thread started by Humanities at Scale.

5. References

Stefan Buddenbohm, Markus Matoni, Stefan Schmunk, Carsten Thiel 2017. Quality Assessment for the Sustainable Provision of Software Components and Digital Research Infrastructures for the Arts and Humanities. *Bibliothek Forschung und Praxis*. 41(2): 231-241. [doi:10.1515/bfp-2017-0024](https://doi.org/10.1515/bfp-2017-0024)

Jenny Oltersdorf, Markus Matoni, Carsten Thiel, 2016. DARIAH Report on researchers' service needs. DARIAH. [hal-01351267](https://hal.archives-ouvertes.fr/hal-01351267)

Marco Raciti, Inês Queiroz, Carsten Thiel, Robert Jäschke. 2017. D1.2 Data Management Plan DESIR DARIAH ERIC Sustainability Refined. DARIAH. [hal-01563849](https://hal.archives-ouvertes.fr/hal-01563849)

¹⁹ DARIAH ERIC Sustainability Refined, Horizon 2020, Project ID 731081