



HAL
open science

Open Data in the Humanities Platform

Stefan Buddenbohm, Maaïke de Jong, Mike Priddy, Yoann Moranville, Paulin Ribbe

► **To cite this version:**

Stefan Buddenbohm, Maaïke de Jong, Mike Priddy, Yoann Moranville, Paulin Ribbe. Open Data in the Humanities Platform: Humanities at Scale: Evolving the DARIAH ERIC. [Technical Report] DARIAH; DANS-KNAW; TGIR Huma-Num (UMS 3598). 2017. hal-01686320

HAL Id: hal-01686320

<https://hal.science/hal-01686320>

Submitted on 17 Jan 2018

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.



D7.3

Open data in the humanities platform

HaS-DARIAH

INFRADEV-3-2015-Individual implementation and operation of ESFRI projects
Grant Agreement no.: 675570

Date: 18-12-2017

Version: 1.0



Project funded 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:	Open data in the humanities platform
Work Package title:	Open Data Infrastructure
Estimated delivery date:	31-12-2017
Lead Beneficiary:	DARIAH
Author(s):	Stefan Buddenbohm (UGOE-SUB) Maaïke de Jong (DANS-KNAW) Mike Priddy (DANS-KNAW) Yoann Moranville (DARIAH) Paulin Ribbe (CNRS)
Quality Assessor(s):	Raisa Barthauer (UGOE-SUB) Francesca Morselli (DANS-KNAW)
Keywords:	open data, open humanities data platform, research data sustainability, repository registry, recommender, data sharing.

HaS-DARIAH

Horizon 2020 – Individual Implementation and operation of ESFRI projects
Grant Agreement no.: 675570

Revision History

Version	Date	Author	Beneficiary	Description
0.1	14-11-2017	Paulin Ribbe	CNRS	First draft
0.2	01-12-2017	Maaïke de Jong, Paulin Ribbe, Stefan Buddenbohm, Yoann Moranville	all	Implementation of additional information
0.3	15-12-2017	Maaïke de Jong, Paulin Ribbe, Stefan Buddenbohm, Yoann Moranville, Francesca Morselli, Raisa Barthauer	all	Last review and quality check
1.0	18-12-2017	Marco Raciti	DARIAH	Submission of the final document

Table of Content

Executive Summary	4
1. Background and context	5
2. Description of the service	5
3. Summary of design and sustainability considerations	6
4. Current state and future perspectives	7

Executive Summary

This document presents the third and final deliverable for the Humanities at Scale (HaS) work package 7: ‘Open Data Infrastructure’.

The Data Deposit Recommendation Service (DDRS) facilitates identifying suitable research data repositories for humanities researchers and other users, depending on case-specific requirements. As an added value service, the DDRS offers the initiation of the ingest and communication process between user and repository by forwarding a deposit request along with a structured description of the research data to the appropriate point of contact.

While the deliverable is the DDRS service itself, this document gives a short overview of the service and its development background during the funding phase, and refers to the technical documentation of the service. The documentation intends to encourage re-use and further development of the DDRS, particularly within the DARIAH context.

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 author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

1. Background and context

As part of the H2020 Humanities At Scale (HaS) project, work package (WP) 7 investigates the sustainability of open data services within the Humanities, with the goal to design and deliver a sustainable open data service for humanities researchers. HaS WP7 consists of the following deliverables:

- Deliverable 7.1 ‘State of the Art Report on Open Access Research Data for the Humanities’ provided an overview and analysis of the current landscape of open access publishing and reuse of research data in the humanities.
- Deliverable 7.2 ‘Design and Sustainability Plan for an Open Humanities Data Platform - Concept for a Data Deposit Recommendation Service’ researched different functional possibilities for an open data service, taking into account existing services and sustainability considerations. It presents the Data Deposit Recommendation Service (DDRS) as the chosen platform and details its design, use cases, technical implementation aspects, and possible future extensions.
- Deliverable 7.3, described in the HaS Description of Work as an ‘Open Data in the Humanities Platform’, consists of the Data Deposit Recommendation Service (DDRS) as a fully functional demonstrator. In this accompanying document we briefly outline a description of the service, the design and sustainability considerations leading up to the service (which are described in detail in D7.2), and perspectives for future developments, particularly within the DARIAH context.

2. Description of the service

A major aim of DARIAH-EU and the HaS project is to promote open science by providing services to humanities researchers. A key part of open science is the sharing and reuse of research data. The Data Deposit Recommendation Service (DDRS) is aimed at facilitating the process of sharing data for researchers.

The DDRS is a user-friendly online service for humanities researchers in the EU to identify suitable research data repositories, using the Re3data.org (Registry of Research Data Repositories) database. Re3data.org is one of the main existing registries for research data repositories, currently containing nearly 2000 records, initially funded by the German Research Foundation (DFG). However, the selection of relevant and reliable repositories is a challenge for all stakeholders involved. How can researchers find the repository that is offering the services they need, relevant for their domain, institution or location, or is compliant with their funders’ policies? Furthermore, research funders, journals, and universities, all would like to point to trusted repositories as citing of research data becomes an increasingly common research practice.

The DDRS addresses the widespread deficiency of user-friendliness with research data deposit services, aims to simplify the identification process of suitable repositories, providing the user with a better understanding of the necessary efforts for depositing

their research data, and intends to lower the hurdle for initiating the communication with the research data repositories. The main characteristics are:

- The DDRS functions as a registry of suitable deposit services for humanities research data.
- It searches for repositories and research data centres and delivers standardised metadata on them, allowing the researcher to decide which repository may be the most qualified one for his case.
- The service accesses the re3data.org database for information retrieval supplementing this information by its own small database with default options (for instance if re3data.org retrieval returns zero results, see also ‘Current state and future perspectives’ section).
- The user experience of the DDRS has two stages: firstly, the user accesses a simple web interface, being recommended repositories after answering a small number of questions and secondly, offered to forward a deposit request to selected repositories.

3. Summary of design and sustainability considerations

In Deliverable 7.2, we reviewed the existing platforms and services in the open data landscape, in the field of humanities as well as general platforms. Furthermore, we researched the different sustainability scenarios and functionalities an open humanities data platform could have, such as a knowledge base, brokering function, access to research data, and registry of tools. Considering existing services and sustainability aspects, we came to the conclusion to focus on one or at least only a few functionalities, to base the service from the very beginning on existing and well-developed services so as not to reinvent the wheel, and to keep the service interoperable especially with regard to research data repositories and other data infrastructures.

We chose to build a service relying on the re3data.org database because it is one of the main registries for data repositories, but not intended to be used by users with little or no background knowledge of the data repository landscape. Because sharing and reusing data are essential elements of open science, it is important that researchers can easily find a suitable repository to deposit their data. The issue of sustainability has been a key factor of consideration in our design study for an open humanities data service. The chosen approach, which builds upon an existing and well-established service that is expected to grow in the future, requires relatively little future maintenance compared to most of the different possible architectures. It also includes the opportunity for further extensions due to the standardised character of the re3data.org infrastructure and the reliance upon a standardised and easy to use metadata format.

4. Current state and future perspectives

The DDRS was designed and developed in such a way that it allows for adaptability (e.g. change of questions, updated repository contact information, additional languages, affiliation/institute etc.) and service extensions for the changing requirements of the community. The service currently filters out repositories that do not provide data upload or persistent identifiers, but these filters can be easily changed, and other filters applied. For example, when the registry grows in the future and there will be more repositories that are certified (e.g. with Data Seal of Approval), this criterion can be applied as a filter. The dependency of the DDRS on re3data.org has clear benefits as outlined above and is described in detail in D7.2, however, there are disadvantages as well. Currently the records in re3data.org are biased in terms of disciplinary field and country, with certain disciplines (including the humanities) and countries underrepresented compared to others. Furthermore, any mistakes in the re3data.org records will be reflected by the DDRS as well. The last point is also related to the infrequent update cycles of the re3data.org database which is out of the hands of the DDRS. As re3data.org is expected to keep growing in both size, usage, and quality, these issues will become less relevant in the future. But again: the database used by re3data.org is a very broad, functional and up to data collection compared to other, similar services around.

One of the characteristics of the re3data.org website that makes it difficult to use, is that it filters out the more general repositories when the search criteria are more refined, while in many cases these general repositories are suitable for depositing data. For example, the EASY archive is the long-term archive for humanities and social sciences data in the Netherlands. But when a Dutch researcher searches with a discipline more specified than humanities or social sciences (e.g. philosophy), the EASY repository is not returned by re3data.org, while it should be a default archive. One of the benefits of the DDRS is that it works around this: the service returns the suitable general repositories at the national and EU level. A list of these default results for different combinations of country and discipline are given in the technical documentation ([link below](#)).

The DDRS was designed and developed with uptake by DARIAH in mind. The benefits for DARIAH to take up and sustain the service can be summarised in the following points:

- Future enhancements: the DDRS architecture is modular and was developed from the start with the possibility of future functional changes and extensions in mind.
- Entrance point for new users not yet familiar with DARIAH: to our understanding the service addresses a demand not only within the DARIAH community but in the humanities community in general. The service can such be seen as easy entrance point into the DARIAH service portfolio.
- Low costs: compared with other services (for instance collaborative services or services relying on a proprietary database), we expect that the costs of maintaining the service will be relatively low.

- Implementation example for an integrated service: a main idea behind the Ddrs was to base it on a well-established existing service, thereby avoiding developing a new database or infrastructure. By closely cooperating with re3data.org - a non-humanities specific service - we feel that the Ddrs could serve as role model for service developments within (and outside) DARIAH.

Source code and documentation

<https://github.com/DARIAH-ERIC/ddrs>

Data Deposit Recommendation Service

<https://ddrs-dev.dariah.eu/>