

Dataverse: Switchboard plugin

Centre Meeting 2021
Cesare Concordia
ISTI - CNR

The Language Resource Switchboard (LRS)

- [Language Resource Switchboard](#)
- Development started within the CLARIN-PLUS project.
- The LRS can be seen as a Virtual Tool Registry.
- For a given resource uploaded on LRS:
 - Identifies all tools that can process the resource
 - Allows a user to select and invoke one tool
 - Transfers to the tool all relevant information about the resource
- Tools are sorted in terms of the tasks they perform

The Harvard Dataverse

- The Harvard Dataverse is an open source web application to share, preserve, cite, explore research data.
- Dataverse is developed and maintained by the Institute for Quantitative Social Science (IQSS) at Harvard University since 2006
- A large open Dataverse instance is hosted Harvard University Library, 69 known installations of Dataverse exist throughout the world (as of today, <https://dataverse.org/installations>)
- Users can
 - deposit and describe data files using DDI and Dublin Core-compliant standards, as well as certain discipline-specific metadata standards,
 - generate PIDs and data citations,
 - assign access permissions and licence terms
 - link the dataset to related publications and datasets
 - ...

Switchboard plugin: roadmap

1. **Switchboard integration.** Integrating the Switchboard in Dataverse such that a user that browses the Dataverse catalogue UI can select a specific data object and invoke the Switchboard. **Done**
2. **Text processing and dictionary lookup.** In Dataverse it is possible to pre-view text and pdf by invoking specific viewers. We can enable the text processing and dictionary lookup functionality offered by services integrated in the Switchboard. **Done**
3. **Data resubmission by Switchboard invoked services.** For some services it may make sense to enable resubmission of a service processing result into Dataverse. This has to be enabled in the specific service itself. **In Progress**

Dataverse External Tool Integration

- The Dataverse system enables developers to integrate external tools that can provide additional features that are not part of Dataverse itself, such as data file previews, visualization, and curation.
 - <https://guides.dataverse.org/en/latest/admin/external-tools.html>
- An integration module has been developed to enable the Resource Switchboard to be integrated in Dataverse as an external tool.
 - previewers are based on:
<https://github.com/GlobalDataverseCommunityConsortium/dataverse-previewers>
- Three possible integrations

File sent to LRS from the Dataverse Dataset View

The screenshot shows the Dataverse interface for a dataset named "Test". At the top, there is a navigation bar with "Dataverse" logo, search, user guide, support, sign up, and log in options. Below the navigation bar, the dataset title "Test" is displayed, along with its version "Version 12.0". The dataset description includes the text: "Concordia, Cesare, 2020, 'Test', https://doi.org/10.5072/FK2/47RQOO, Root, V12, UNF:6p2j3walC+5WhDR3rJGRB8w== [fileUNF]". There are buttons for "Access Dataset", "Contact Owner", and "Share". Below this, there are links for "Dataset Metrics" and "388 Downloads". The "Description" and "Subject" sections are visible, with "Description" containing "This is a test" and "Subject" containing "Other". A tabbed interface shows "Files", "Metadata", "Terms", and "Versions". A search bar is present with the text "Search this dataset...". Below the search bar, there are filters for "File Type: All" and "Access: All". A list of 6 files is shown, each with a download icon and a share icon. A tooltip for the "d31_abstract.txt" file reads "SSHOC preview Process text file with LRS".

The screenshot shows the Language Resource Switchboard interface. At the top, there is a navigation bar with "Language Resource Switchboard", "Upload", "Tool Inventory", and "Help" options. Below the navigation bar, there is a section for "Resources". The "Resources" section displays a table with columns for "Resource", "Mediatype", and "Language". The table shows a resource named "d31_abstract.txt" with a size of "5.03 KiB", a mediatype of "text/plain", and a language of "English". Below the "Resources" section, there is a section for "Matching Tools". The "Matching Tools" section displays a list of tools, each with a "Group by task" button and a "Search for tool" input field. The tools listed are: "Constituency Parsing" (WebLicht Const Parsing EN), "Dependency Parsing" (Spacy (hosted by D4Science) - EN, UDPipe, WebLicht Dep Parsing EN), "Distant Reading" (Voyant Tools), "Lemmatization" (CSTLemma (hosted by D4Science), WebLicht Lemmas EN), "Machine Translation" (LINDAT Translation), "Morpho-syntactic tagger" (Tagger NLTK), and "Morphological Analysis" (WebLicht Morphology EN).

LRS in the Dataverse file view 1

d31_abstract.txt

This file is part of "Test".

Version 12.0

File Citation

Concordia, Cesare, 2020, "d31_abstract.txt", *Test*, Root, V12

[Cite Data File](#) ▾

Learn about [Data Citation Standards](#).

Access File ▾

Contact Owner

Share

File Metrics [?]

165 Downloads [?]

Dataset Citation

Concordia, Cesare, 2020, "Test", <https://doi.org/10.5072/FK2/J7RQOO>, Root, V12, UNF:6:p2j3waLC+5WnDR3rJGRB8w== [fileUNF]


[Cite Dataset](#) ▾

Learn about [Data Citation Standards](#).

Preview

Metadata

Versions

 Explore on SSHOC previewer

 [Process with Language Resource Switchboard](#)

SSHOC D3.1 Report on SSHOC (meta)data interoperability problems

Broeder, Daan; Trippel, Thorsten; Degl'Innocenti, Emiliano; Giacomi, Roberta; Sanesi, Maurizio; Kleemola, M

The SSHOC project aims to build a Cross-lingual and Cross-domain Semantic Graph (CSHG) component of the European Co

LRS in the Dataverse file view 2

The screenshot shows the Dataverse interface for a file named 'd31_abstract.txt'. The browser address bar shows the URL: `dataverse.philae.isti.cnr.it/file.xhtml?fileId=7&version=12.0`. The page header includes the Dataverse logo, a search bar, and navigation links for 'User Guide', 'Support', 'Sign Up', and 'Log In'. The main content area displays the file name 'd31_abstract.txt' and a note that it is part of a dataset named 'Test'. Below this, there are sections for 'File Citation' and 'Dataset Citation', each with a 'Cite' button and a link to learn about Data Citation Standards. On the right side, there is an 'Access File' button with sub-options for 'Contact Owner' and 'Share', and a 'File Metrics' section showing '162 Downloads'. At the bottom of the main content area, there are tabs for 'Preview', 'Metadata', and 'Versions', and a button to 'Explore on SSHOC previewer'. A footer bar contains a 'Language Resource Switchboard' with links for 'Upload', 'Tool Inventory', and 'Help', along with the CLARIN logo.

d31_abstract.txt
This file is part of "Test".
Version 12.0

File Citation
Concordia, Cesare, 2020, "d31_abstract.txt", Test, Root, V12
[Cite Data File](#) - [Learn about Data Citation Standards.](#)

Dataset Citation
Concordia, Cesare, 2020, "Test", <https://doi.org/10.5072/FK2/J7RQOO>, Root, V12, UNF:6:p2j3waLC+5WnDR3rJGRB8w== [fileUNF]
[Cite Dataset](#) - [Learn about Data Citation Standards.](#)

[Preview](#) [Metadata](#) [Versions](#)

[Explore on SSHOC previewer](#)

[Language Resource Switchboard](#) [Upload](#) [Tool Inventory](#) [Help](#)

Resources

[d31_abstract.txt](#) 5.63 KIB

Mediatype: Language:

Matching Tools Group by task

Constituency Parsing

- [Open](#) WebLicht Const Parsing EN

Dependency Parsing

- [Open](#) Spacy (hosted by D4Science) - EN
- [Open](#) UDPipe
- [Open](#) WebLicht Dep Parsing EN

LRS in the Dataverse file view: *popup* menu

The screenshot displays the Dataverse interface. At the top left is the Dataverse logo. The top right navigation bar includes links for Search, User Guide, Support, Sign Up, and Log In. A secondary navigation bar contains a button labeled 'Explore on SSHOC previewer'. The main content area features a document titled 'SSHOC D3.1 Report on SSHOC (meta)data interoperability problems' by Broeder, Daan; Trippel, Thorsten; Degl'Innocenti, Emiliano; Giacomi, Roberta; Sanesi, Maurizio; Kleemola, M. The document text discusses the SSHOC project's goals and the diversity of SSHOC communities. A 'Resource Switchboard' popup menu is overlaid on the document, showing the selected resource 'Social Sciences' and a language selection dropdown. Below the popup, a 'Matching Tools' section is visible, including a 'Group by task' checkbox and a 'Search for tool' input field. Under the 'Lookup Tools' section, two tools are listed: 'Content Search' and 'Getty Art and Architecture Thesaurus', each with an 'Open' button and an external link icon.

Dataverse

Search User Guide Support Sign Up Log In

Explore on SSHOC previewer

Process with Language Resource Switchboard

SSHOC D3.1 Report on SSHOC (meta)data interoperability problems

Broeder, Daan; Trippel, Thorsten; Degl'Innocenti, Emiliano; Giacomi, Roberta; Sanesi, Maurizio; Kleemola, M

The SSHOC project aims to build a Social Sciences and Humanities Open Cloud (SSHOC) as part of the European Open Science and metadata standards that are current and older's infrastructures were interviewed. In line with these goals, this report provides a large selection of different media types. In the context of this deliverable, the SSHOC communities and the diversity of SSHOC communities and the diversity of SSHOC communities and the diversity of SSHOC communities. The interviews and the desk research were decided based on the interviews and the diversity of SSHOC communities. The diversity of SSHOC communities and the diversity of SSHOC communities. The priorities for providing content were decided based on the interviews and the diversity of SSHOC communities. Beyond WP3, this document is relevant for the discussion between the work packages and the diversity of SSHOC communities. The authors of this report wish to share their time and valuable contribution.

Resource Switchboard

" Social Sciences"

Language Select the language of the resource

Matching Tools

Group by task Search for tool

Lookup Tools

Open Content Search

Open Getty Art and Architecture Thesaurus

The SSHOC Previewer, developed by Cesare Concordia, is available at <https://www.globaldataversecommunityconsortium.com/globaldataversecommunityconsortium/dataverse-previewers>.

LRS & Dataverse file previewer

The screenshot shows the Dataverse web interface. At the top, the URL is `dataverse.philae.isti.cnr.it/dataset.xhtml?persistentId=doi:10.5072/FK2/J7RQO0&version=12.0`. The page title is "Test" (Version 12.0). The dataset description is "Concordia, Cesare, 2020, 'Test', https://doi.org/10.5072/FK2/J7RQO0, Root, V12, UNF:6p2jhwsl.Cv-SWwDR3JGR9Bw== [fileUNF]". The file list contains six items:

File Name	Format	Size	Date	Downloads
gratorngofrompub.pdf	Adobe PDF	186.7 KB	Jan 4, 2021	88
d31_abstract.txt	Plain Text	5.6 KB	Dec 22, 2020	163
dftotg.tab	Tabular Data	1.1 KB	Jan 4, 2021	11
file_example_MP3_1MG.mp3	MP3 Audio	1.0 MB	Jan 4, 2021	8
login.png	PNG Image	26.1 KB	Dec 22, 2020	0
gh.geogson	application/geogson	1.0 MB	Feb 10, 2021	43

A tooltip for the "SSHOC Text Preview" button is shown, with the text: "Process text file with LRS".



SSHOC Text Preview

Filename: d31_abstract.txt

In *Test* (version 12.0), by Concordia, Cesare

Download File

Process with Language Resource Switchboard

Close Preview

File uploaded on 2020-12-22

SSHOC D3.1 Report on SSHOC (meta)data interoperability problems

Broeder, Daan; Trippel, Thorsten; Degl'Innocenti, Emiliano; Giacomi, Roberta; Sanesi, Maurizio; Kleemola, Mari; Moilanen, Katja; Ala-

The SSHOC project aims to build a Social Sciences and Humanities Open Cloud (SSHOC) as part of the European Open Science Cloud (EOSC) by In line with these goals, this report provides an inventory of data formats and metadata standards that are currently used and relevant for the context of this deliverable, selected experts from the SSHOC stakeholder's infrastructures were interviewed about research data and metadata. The interviews and the desk research indicated both the diversity of the SSHOC communities and the diversity of metadata standards used in the SSHOC project. The diversity of SSHOC communities is also shown in their data: there is a large selection of different media types and an enormous selection of metadata standards. The priorities for providing conversion services and planning solutions were decided based on the interviews and on the needs from other work packages. Beyond WP3, this document is relevant to WP4, WP5, WP7 and WP9. Ongoing discussion between the work packages and tasks is needed. The authors of this report wish to thank the interviewed informants for their time and valuable contribution.

Documentation, source code and test installation

- Technical documentation and source code: <https://gitea-s2i2s.isti.cnr.it/concordia/dataverse-lrs>
- User documentation: <https://docs.google.com/document/d/1N-TXWzinCUjsFF7kl5ev5QQheR3FEmhhZwNI22wHYr4/edit#>
- Test installation and demo: <https://dataverse.philae.isti.cnr.it/>

Questions?

- Thank you.