

# Choice of Vocabulary publication platforms for SSHOC

Monica Monachini

SSHOCWP3.1 leader

CNR-ILC

CLARIN - IT National Coordinator



A poster for the SSHOC Workshop. It features a large blue circle on the right side containing the CLARIN and SSHOC logos. The text on the poster reads: 'SSHOC Considerations for the Vocabulary Platforms', 'SSHOC Workshop', '06 November 2020', '09:00 to 13:15', and 'Online'. The background of the poster shows a document with text from the SSHOC glossary, including definitions for 'Training', 'Heritage', 'Cloud', 'Data', 'SSHOC', and 'Open Science'.

Humanities

Training *noun* • the process of  
ring the skills you need to do a  
lar job or activity.

Heritage *noun* • features belong  
to the culture of a particular  
e, such as traditions, lan-  
guages, buildings, that were cre-  
ated and still have histo-

Cloud (Comput-  
ing) *adjective*  
relating to a computer  
network that allows  
users to access and  
store data and pro-  
grams over the inter-  
net.

Data *noun*  
y facts or infor-  
mation that is col-  
lected and stored  
electronically and  
can be used for  
analysis.

SSHOC  
SSHOC is a  
project of the  
European Com-  
mission, funded  
by the Horizon  
2020 research  
and innovation  
programme.

Open Science  
Open Science  
is a movement  
to make sci-  
entific research  
publications, data,  
code, and software  
openly accessible  
to all, including  
the research com-  
munity, amateur  
researchers, and  
the public.

Research *noun* • a detailed  
study of a subject,  
especially in order to  
discover (new) infor-  
mation or reach a  
deeper understand-  
ing.

Humanities *noun* •

CLARIN

SSHOC

**SSHOC Considerations for the  
Vocabulary Platforms**

**SSHOC Workshop**

**06 November 2020**

**09:00 to 13:15**

**Online**



# SSHOC WP3

What

Lifting Technologies and Services  
into the SSH Cloud

How

- Contribute to the SSH Cloud development wrt both
  - infrastructure components and content, and
  - policies and governance
- Make CLARIN technologies also useful for the other SSH infrastructures



# SSHOC T3.1

## What

Provide background resources and tools to support better discovery of SSH research data in order to ensure better access and reusability.

## How

- Data and services accessible and usable in SSH is a matter of providing
  - translated metadata in different languages and
  - a service to publish the vocabularies as a catalogue

Multilinguality

Searching, Browsing

# SSHOC T3.1.3

## What:

Milestone MS08:  
*Provide motivated recommendations for a vocabulary publication platform for SSHOC*

## How:

- Survey; interviews; main technological features to provide a SSHOC solution for the publication platform
- <https://drive.google.com/drive/u/0/folders/1XvRSb7hevHYWBemPpKEqAQDb9C8tsY5p>



Research and Innovation Action

**Social Sciences & Humanities Open Cloud**

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Report on **Milestone MS8.0**

Choice of Vocabulary Publication platform for SSHOC

Dissemination Level	PU
Due Date of Milestone	30/06/2019, M8
Actual Achievement Date	30/06/2019
Lead Beneficiary/LTP	16. CNR
Work Package	WP3 Lifting Technologies and Services into the SSH Cloud
Task	3.1. Multilingual Terminology
Version	V1.1
Number of Pages	p. 1 – p. 18

**Abstract:** The goal of the activities in WP3.1.3 is to permit metadata-based discovery in different languages and one of

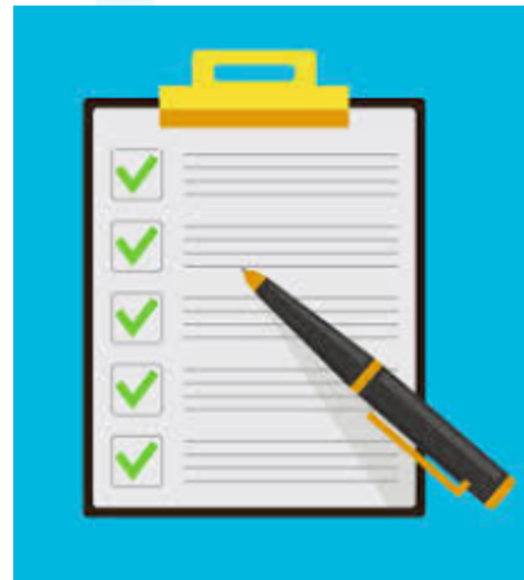


## T3.1.3 The Survey of vocabulary management tools

Survey of existing systems for managing vocabularies to describe and allow discovery of research data in SSH infrastructures:

[https://docs.google.com/spreadsheets/d/1cqMujo0SwQHRogkuv2Xw6QV\\_PXIEGH-z74CwfWLT08/edit#gid=0](https://docs.google.com/spreadsheets/d/1cqMujo0SwQHRogkuv2Xw6QV_PXIEGH-z74CwfWLT08/edit#gid=0)

Matrix of systems/features



# T3.1.3 The Matrix

Data model	Notes	Concepts	Synonyms	Hierarchy	Definitions or use notes	Concept schemes	Multilinguality translation	Import Export	Browser	Editor	User management	Documentation	Linking Mapping Alignment	Examples	Versioning and history
------------	-------	----------	----------	-----------	--------------------------	-----------------	-----------------------------	---------------	---------	--------	-----------------	---------------	---------------------------	----------	------------------------

<b>ACDH Vocabulary Service</b>	<a href="https://vocabs.dariah.eu/en/about">https://vocabs.dariah.eu/en/about</a> collaborative creation, maintenance and publication of vocabularies and taxonomies of any kind														
<b>THEMAS</b>	<a href="https://github.com/isl/THEMAS/">https://github.com/isl/THEMAS/</a> (Thesaurus Management System) is a Web based system for creating and managing multi - faceted multilingual thesauri														
<b>CESSDA Vocabulary Service</b>	<a href="https://vocabularies.cessda.eu/">https://vocabularies.cessda.eu/</a> discover, browse, edit and download controlled vocabularies in a variety of languages.														
<b>OpenSKOS</b>	<a href="https://github.com/OpenSKOS/OpenSKOS">https://github.com/OpenSKOS/OpenSKOS</a> web service based approach to publication, management and use of vocabulary data														
<b>Skosmos vocabulary browser</b>	<a href="http://skosmos.org/">http://skosmos.org/</a> Open source web-based SKOS browser and publishing tool														
<b>iQvoc</b>	<a href="https://iqvoc.net/">https://iqvoc.net/</a> A Vocabulary Management System for the Semantic Web.														
<b>TemaTres</b>	<a href="https://www.vocabularyserver.com/index.html">https://www.vocabularyserver.com/index.html</a> open source vocabulary web application to manage and exploit vocabularies, thesauri, taxonomies														
<b>VocBench3</b>	<a href="http://vocbench.uniroma2.it/">http://vocbench.uniroma2.it/</a> management of publication workflows, history, validation														

# T3.1.3 The Interviews

Interviews with experts, concerning the characteristics of the most used vocabularies (and vocabulary platforms): what they use, what they wish that doesn't exist or what they would need from the platform:

<https://docs.google.com/spreadsheets/d/1FSDIEPBfYdRU6TQgPXoWofpibkAGGv3YzZrMy0FBslo/edit#gid=0>



Link to the vocabulary	<a href="http://tadrah.dariah.eu/vocab/index.php">http://tadrah.dariah.eu/vocab/index.php</a>	<a href="https://vocab.dariah.eu/backbone-the-europe/en/">https://vocab.dariah.eu/backbone-the-europe/en/</a>	<a href="https://vocabularies.cesda.eu/#/discover">https://vocabularies.cesda.eu/#/discover</a>	<a href="https://www.clarin.eu/cor">https://www.clarin.eu/cor</a>
Does the vocabulary need a) an Editor platform for maintaining it, b) a browsing and publication platform to provide access to it, or is the current enough?	Editing is done on GitHub <a href="https://github.com/dtaxonomy/TaO-RAH/">https://github.com/dtaxonomy/TaO-RAH/</a>	The vocabularies are maintained and published in CESSDA Vocabulary Service (link above). DOI publishes their vocabularies on their own website as well. So CESSDA does not need for new vocabulary/ontology platforms.	Ideally both, but right now data input is done in batch mode based on tabular data imports	
Name of organisation maintaining it	DARIAH external users	CESSDA	CLARIN	
Name of the person filling in the information	Younn Moraville	Taina Jaaskeläinen	Dieter Van Uytvanck	
What is the vocabulary used for?	For research activities, research objects and research techniques in DH	Used in metadata for describing data, and in CESSDA data catalogues to allow filtering results.	Metadata disambiguation	
Who uses it?	openmethods.clarin-dariah.eu teresah.dariah.eu dirdirectory.org and more (see github page)	CESSDA Data Catalogue and CESSDA Service Providers.	<a href="http://www.clarin.eu">http://www.clarin.eu</a>	
Who can browse/see it?	Anyone	Anyone	Anyone	
What is important in browsing, how would the user need to see the vocabulary, i.e. with all language versions of a term together, or each language separately as a whole?	Can't understand the question, sorry	At present, each language version of the vocabulary presented separately.	all language versions of a term together	
Approximate number of terms/concepts	121 terms ( <a href="http://tadrah.dariah.eu/vocab/sobre.php">http://tadrah.dariah.eu/vocab/sobre.php</a> )	From a few terms up to 100, depending on vocabulary	3163	
Who can edit the vocabulary, i.e. all interested parties or should editing rights be restricted to certain persons/roles?	Rights are restricted to maintainers	Only people with a relevant role can log in and access the Editor. Anyone can browse published vocabularies, logged-in users see also the draft versions.	restricted to editor role	
Does it have hierarchy, that is, broader and narrower terms? (Yes/No)	Yes	Yes	yes	
Does it have related terms? (Yes/No)	No	No	do not know	
Do terms have synonyms/alternative labels?	No	No	yes	
Does it have definitions?	Partly: Activities are, Techniques and Objects are not	Yes	yes	
Does the vocabulary exist in different languages/are terms translated into different languages? Please specify how many languages approximately.	Yes, 4: English, French, German, Spanish	Yes, presently about 7 language versions, more coming in.	I believe so. If yes, nr of languages depends on subcollection (many translations stem from the earlier ISOcat incarnation)	
If translated, specify what content/elements are		CV name, CV definition, descriptive term of the code,		

# T3.1.3 The Interviews

Does the vocabulary need a) an Editor platform for maintaining it, b) a browsing and publication platform to provide access to it?

**What is the vocabulary used for?**

Who uses it?

Who can browse/see it?

What is important in browsing, how would the user need to see the vocabulary?

**Approximate number of terms/concepts**

Who can edit the vocabulary, i.e. all interested parties or should editing rights be restricted to certain persons/roles?

Does it have hierarchy, that is, **broader and narrower terms**? (Yes/No). Does it have **related terms**? (Yes/No)

Do terms have **synonyms**/alternative labels? Does it have **definitions**?

Does the vocabulary exist in **different languages**? Please specify how many languages approximately.

Does the vocabulary have a **license**?

If translated, specify what content/elements are translated (terms, definitions, etc.). Who does the translation?

Is import of current vocabularies needed? If yes, what is the format of the vocabulary at present.

What export formats are needed?

Is API needed?

Is versioning needed, i.e. does the vocabulary have version number and publication date, or is it updated continuously?

Do users need to know what has been changed from last publication?



## T3.1.3 Technical requirements

Standalone installation	Interoperability Protocols	Linked Data	API	Source Code	License
-------------------------	----------------------------	-------------	-----	-------------	---------

List of features that the SSHOC publishing platform should comply with as concerns interoperability protocols, provision of source code, API, the type of format and accessing the vocabularies.

[https://docs.google.com/spreadsheets/d/1s5\\_StggMB1AburKRG4U6kK6oj5Dsc4weaGtXw0w1YNA/edit#gid=0](https://docs.google.com/spreadsheets/d/1s5_StggMB1AburKRG4U6kK6oj5Dsc4weaGtXw0w1YNA/edit#gid=0)

# Vocabulary platform: SSHOC considerations

For the vocabulary publication platform in SSHOC

- not only a publishing tool, but a specialized tool for searching and browsing SKOS vocabularies
- not only a point of access, but also *editing* functionalities
- *linking/mapping/alignment* functionalities to reconcile vocabularies
  - Semi-automated alignment (suggestions, etc...)
  - Synchronization
  - Import/export
  - Role management

# SSHOC Requirements for a vocabulary hosting platform

<b>Data model</b>	- SKOS I/O t(he most used data model in all vocabularies)
<b>Structure</b>	- support hierarchies of concepts may be an added value
<b>Multilinguality</b>	- Multilingual support is required to improve discovery and maximize the accessibility
<b>Concept schemes:</b>	- CS support is important in view of facilitating faceted search
<b>APIs support</b>	- important feature for vocabulary access

# SSHOC Requirements for a vocabulary hosting platform

<b>Browsing interface</b>	<ul style="list-style-type: none"><li>- browsing is regarded as an important added value if accompanied by specialized features</li></ul>
<b>Advanced user management roles</b>	<ul style="list-style-type: none"><li>- system of roles for the maintenance and the incrementation of vocabularies</li><li>- given the necessity for various partners to collaborate in different capacities</li><li>- producing translations would optimally be user-friendly:</li></ul>
<b>Versioning and status</b>	<ul style="list-style-type: none"><li>- Strict versioning with possibility to go back to previous states of the vocabulary and mappings and reconstruct the history of modifications</li><li>- Status and contributions should be monitored; states. e.g., candidate, approved (wrt. the role of the contributors); or draft, initial review, final review, published.</li></ul>

# SSHOC Requirements for a vocabulary hosting platform

- Import and export of thesauri in SKOS / RDF format;
- Unified access to all vocabularies
- Editing with collaborative functionalities
- Alignment functions between vocabularies and external resource
- Management of terms (status, language, URI, hierarchical structure --narrower, broader--, association (related terms), semantic relationships (synonyms), translations, facets; set of concepts (micro-thesauri);
- Management of different roles and workflows
- Management of versioning
- API services
- Friendly and intuitive interface, suitable for use by non-expert users
- Flexible to be adapted to new needs and standards

# What's next



Discuss a more agreed-on solution for the vocabulary platform with experts

- further discuss hosting vocabulary platforms also investigating other possible solutions -- [Sept.2020](#)
- discuss/integrate the features -- [Nov..2020](#)