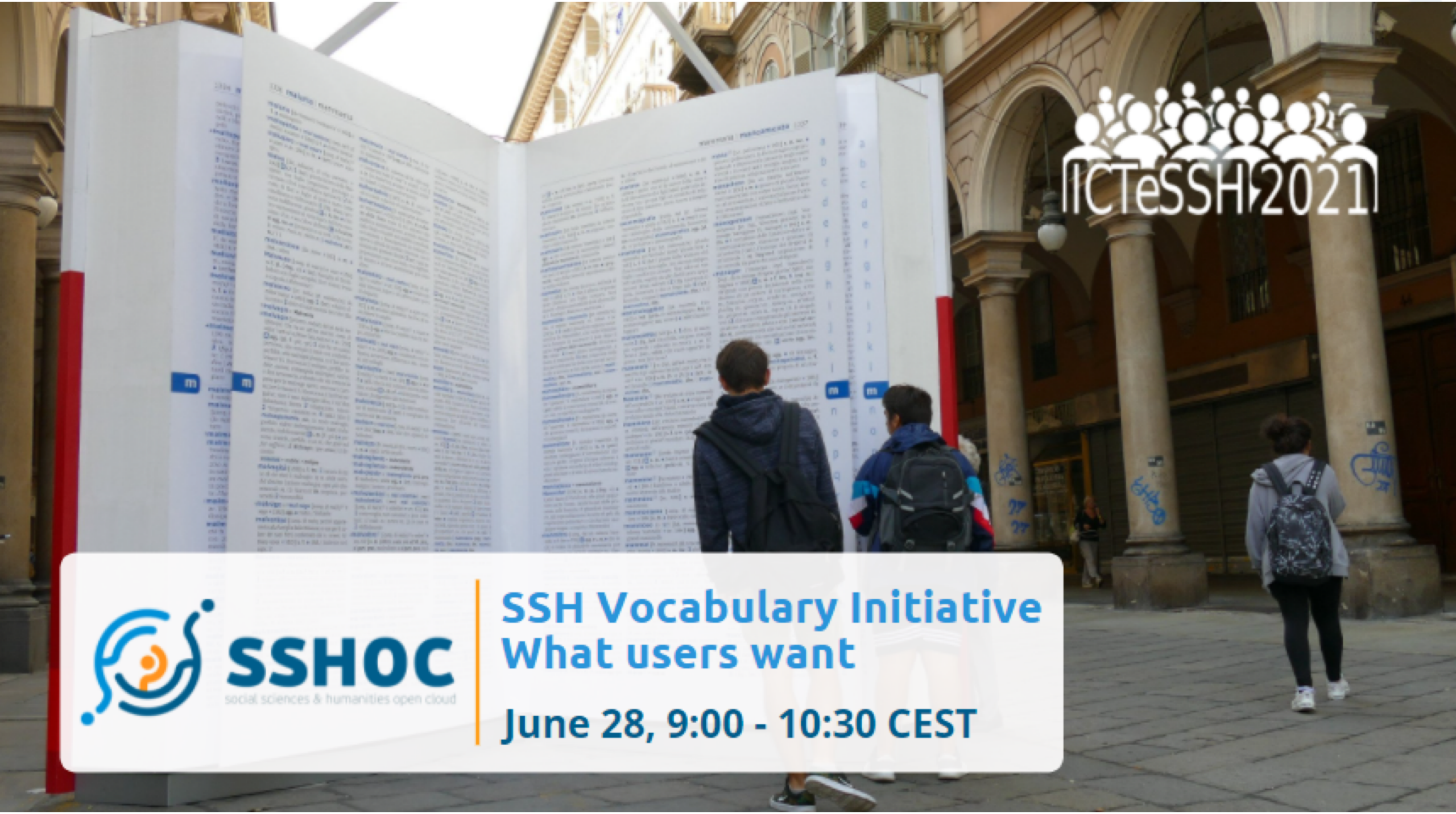

ICTeSSH2021
SSHOC
social sciences & humanities open cloud
SSH Vocabulary Initiative
What users want
June 28, 9:00 - 10:30 CEST

Agenda

1. Welcome - Marieke Willems (Trust-IT)
2. SSH Vocabulary Initiative - Daan Broeder (CLARIN)
3. How can researchers use vocabulary in SSH tools & use cases
   a. USAGE: Using Vocabularies in the SSH community - Iulianna van der Lek (CLARIN)
   b. FINDABILITY: Current ways of locating suitable vocabularies and using the SSH Open Marketplace - Matej Ďurčo (ACDC-CH & DARIAH)
   c. INTEROPERABILITY: The Vocabulary Matching Tool - Holly Wright (Archaeology Data Service)
   d. INTEROPERABILITY: Using MT to create multilingual vocabularies - Monica Monachini (CNR-ILC & CLARIN-IT)
4. The SSH Vocabulary Initiative - What users want. Panel Discussion
5. Wrap up

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Monica Monachini
CLARIN

Holly Wright
Archaeology Data Service

Daan Broeder
CLARIN

#SSHOCaVocabulary
#SSHOCifyCLARIN

Taina Jääskeläinen
Finnish Social Science Data Archive

Matej Durco
ACDH-CH

Iuliana van der Lek
CLARIN

SSHOC
social sciences & humanities open cloud

CLARIN

We are interested in learning more about your experience with vocabularies. Help us out by filling in the following polls!

SSH VOCABULARY INITIATIVE

Daan Broeder, CLARIN ERIC

ICTeSSH, June 2021

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSHOC

social sciences & humanities open cloud

Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:

Research and Innovation action

(INFRAEOSC-04-2018)

Project budget:

€ 14,455,594.08

Project website:

www.SSHOpenCloud.eu

Partners: 48

(23 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Duration: 40 months

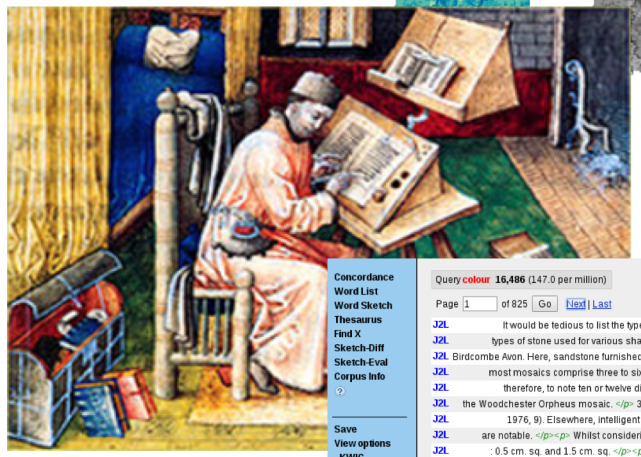
(January 2019 – 30 April 2022)

Objectives:

- creating the social sciences and humanities **(SSH)** part of European Open Science Cloud **(EOSC)**
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

Diversity in describing phenomena

- A considerable part of research concerns describing and analyzing phenomena using descriptive schemas and concepts
- Typical for the SSH is a high variety of such schema and concepts, caused by
  - wide variety of data types, sub-community specifics, schools of thought,…
  - divergent purposes and available effort
- Suitable well crafted vocabularies are essential for
  - accurate descriptions and classifications, countering interpretative vagueness (reduce ambiguity)
  - efficient retrieval and search

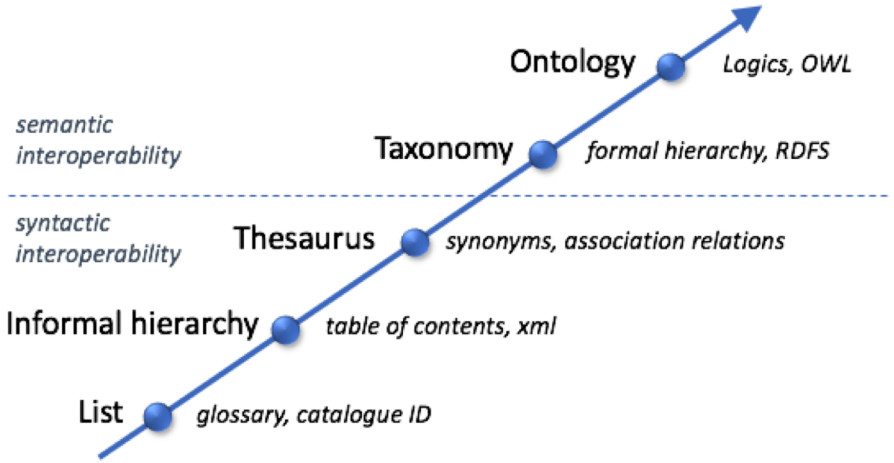


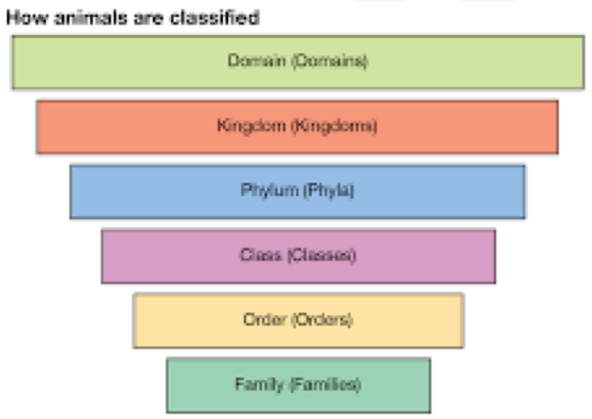
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

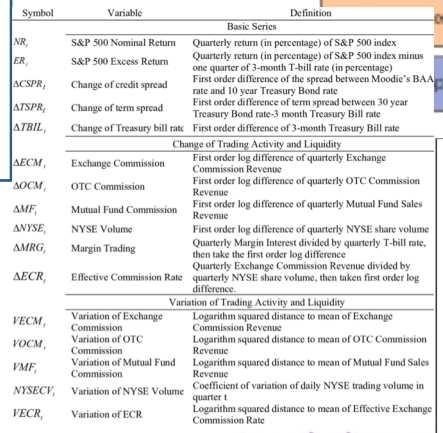
Controlled Vocabularies

- Lists
- Glossaries
- *Thesauri*
- *Taxonomies*
- *Ontologies*

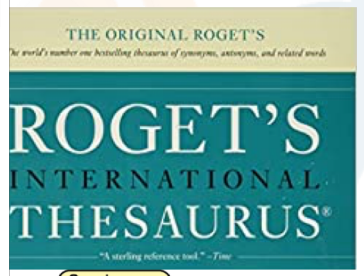
W3C std. RDF, SKOS, OWL: widely accepted and supported formal way to represent vocabularies

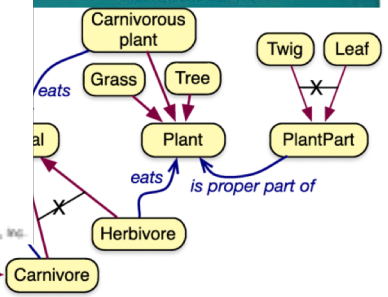


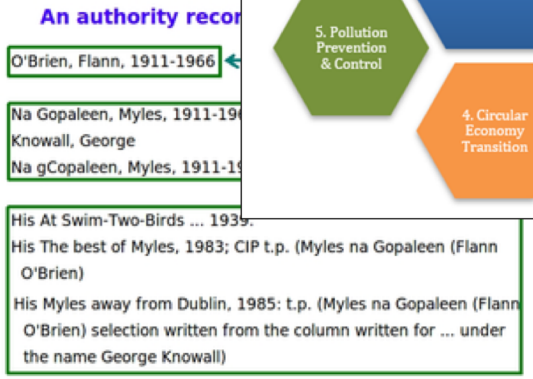




Symbol	Variable	Definition
	Basic Series	
NR_t	S&P 500 Nominal Return	Quarterly return (in percentage) of S&P 500 index
ER_t	S&P 500 Excess Return	Quarterly return (in percentage) of S&P 500 index minus one quarter of 3-month T-bill rate (in percentage)
$\Delta CSPR_t$	Change of credit spread	First order difference of the spread between Moodie's BAA rate and 10 year Treasury Bond rate
$\Delta TSPR_t$	Change of term spread	First order difference of term spread between 30 year Treasury Bond rate-3 month Treasury Bill rate
$\Delta TBIL_t$	Change of Treasury bill rate	First order difference of 3-month Treasury Bill rate
	Change of Trading Activity and Liquidity	
ΔECM_t	Exchange Commission	First order log difference of quarterly Exchange Commission Revenue
ΔOCM_t	OTC Commission	First order log difference of quarterly OTC Commission Revenue
ΔMF_t	Mutual Fund Commission	First order log difference of quarterly Mutual Fund Sales Revenue
$\Delta NYSE_t$	NYSE Volume	First order log difference of quarterly NYSE share volume
ΔMRG_t	Margin Trading	Quarterly Margin Interest divided by quarterly T-bill rate, then take the first order log difference
ΔECR_t	Effective Commission Rate	Quarterly Exchange Commission Revenue divided by quarterly NYSE share volume, then taken first order log difference.
	Variation of Trading Activity and Liquidity	
$VECM_t$	Variation of Exchange Commission	Logarithm squared distance to mean of Exchange Commission Revenue
$VOCM_t$	Variation of OTC Commission	Logarithm squared distance to mean of OTC Commission Revenue
VMF_t	Variation of Mutual Fund Commission	Logarithm squared distance to mean of Mutual Fund Sales Revenue
$NYSECV_t$	Variation of NYSE Volume	Coefficient of variation of daily NYSE trading volume in quarter t
$VECR_t$	Variation of ECR	Logarithm squared distance to mean of Effective Exchange Commission Rate







This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabularies in the SSHOC project

- Coordination wrt. vocabularies: originally a limited effort
  - Investigation of a common recommended platform for publishing and sharing vocabularies
  - Testing machine translation for vocabularies
  - Flexible integration of vocabularies in tools: e.g. SSHOC Dataverse
  - Identifying & creating proper vocabularies for SSH Marketplace and others
- Identified more opportunities during the project
  - Inventory and registration of relevant SSH vocabularies
  - Recommendations for further common approaches e.g. CV authoring tools
  - Opportunity & need to represent SSH interests with other stakeholders e.g. software & service providers

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabulary visibility and discovery

- Vocabularies not always FAIR yet; they should be properly registered and published, researchers & infrastructure providers should be able to find and reuse -> **FAIR semantic artefacts**
- SSH Vocabulary registry or a general one that supports sufficient discipline specificity e.g. Bartoc (3300 entries whereof 1200 SSH)
- Vocabulary search facility, that searches in vocabulary metadata **but also** the vocabulary terms themselves.
- *Note that providing optimal recommendations for researchers can be complicated e.g. also aspects of context and user profile play a role*

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabularies & Interoperability

- **Technical / Format interoperability.** SKOS and OWL are broadly accepted
  - but many projects use spreadsheets and tables and are locked in silos using highly specific software to manage and use these
  - Specific recommendations for vocabulary versioning are needed
- **Semantic interoperability**. Coming from different traditions different organizations and projects have developed different vocabularies to describe similar data. Normalization or conversion needed; the vocabularies involved can be huge and expertise expensive (see the VMT presentation).
- **Cultural & Human interoperability** aspects. Multi-lingual vocabularies, localization aspects.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSH Vocabulary Initiative

We Identified the following topics for working towards common recommendations and implementations

- Vocabulary versioning (also in relation with recommendation for vocabulary authoring tools)
- SSH Vocabulary Registry; board a general registry or run our own
- Vocabulary recommendations; discovery via minimal metadata, term/token search, context via the SSH open marketplace?
- Recommended management platforms: SKOSMOS, VocBench, …, and the APIs

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Presentations

- USAGE:
  - Use of Vocabularies in the SSH community - Iulianna van der Lek
- FINDABILITY
  - Current ways of locating suitable vocabularies - Matej Ďurčo
  - What can the SSH Open Marketplace add to that?
- INTEROPERABILITY
  - Vocabulary Matching Tool - Holly Wright
  - Use of MT for creating multilingual vocabularies - Monica Monachini

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Using vocabularies in the SSH community

Iulianna van der Lek

CLARIN ERIC

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSH Vocabulary survey

- Timeline: Jan – June 2020
- Authors: CNRS/ HUMA-NUM
- Goal:
  - To learn about vocabulary usage and practices in SSH
  - To improve discoverability in SSHOC Open Marketplace

Clara Petitfils, Suzanne Dumouchel, Nicolas Larrousse, Laure Barbot, Klaus Illmayer, Matej Ďurčo and Tomasz Parkola. (2021). *SSHOC D7.6 Resources for Marketplace content description.* Zenodo. doi:10.5281/zenodo.4558339

https://zenodo.org/record/4558339#.YNi4nugzZPY

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

SSH Vocabulary survey

52 out of 72 complete responses - >used vocabularies

- **SSH disciplines**: Linguistics, Archaeology & Prehistory; Sociology, Communication Sciences
- **Organizations**: Universities and research organisations (62,%), researchers, research libraries & archives, policy-making organisations
- **Countries**: France (55%), Spain, the Netherlands and Germany
- **Languages**: English, French, German and Spanish

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSH Vocabularies Survey

Vocabularies are used for:

- 58,49% “Concepts: Disciplines”
- 54,72% “Concepts: General concepts”
- 52,83% “Named entities : geographical entities”
- 45,28% “Named entities : persons”
- 45,28% “Concepts: Scholarly activities”
- 28,3% “Named entities: Institutions”
- 30,19% Other

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSH Vocabulary Survey

Linguistics
Philosophy
Archaeology & Prehistory
History & Sociology
Art & Art History

Type of used vocabularies:

- Controlled vocabulary (DDI, CESSDA)
- Metadata schemas (Dublin Core)
- Generic SSH thesaurus (ELSST)
- Specialised thesaurus in SSH fields (Pactols)
- Domain-specific dictionaries and glossaries
- Ontologies representing and structuring data (SKOS)
- Gazetteers, i.e. authority files of particulars, persons, places (Geonames, Getty TGN)
- Research identifiers (ORCID)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabulary inventory

Inventory of vocabularies and thesauri for SSHOC

File Edit View Insert Format Data Tools Add-ons Help Last edit was on 17 May

Title	Homepage	Description	Creator	Publisher	URI	Domain	Subjects	Type of data set	Type (registry, repository, service))	Link to repository	Size	Format	User group	How it is used	Languages	License
AGROVOC - Food and Agriculture Organization of UN	http://www.fao.org/agrovoc/search	The largest Linked Open Data set about agriculture available for public use and facilitates	FAO	FAO	http://aims.fao.org/aos/agrovoc	Food and Agriculture		Thesaurus	Repository	http://www.fao.org/agrovoc	38442 concepts	RDF,SKOS-XL	Researchers, librarians, information managers	AGROVOC is used by researchers, librarians and	Multilingual	CC-BY
BackBone Thesaurus DARIAH	https://vocabs.acdh.oeaw.ac.at/backbone_thesaurus/	It is a thesaurus for the humanities developed and maintained by the Thesaurus Maintenance	BBT maintenance WG	Austrian Centre for Digital Humanities	http://bartoc.org/en/node/18706	Arts & Humanities	Computer science, Information and general	Thesaurus	Repository		30 concepts, 10 collections	RDF, SKOS, XML	Researchers in Humanities	It can be used to navigate the BBT hierarchies and	French German	https://c
BARTOC Register of Thesauri, Ontologies and Classification	http://bartoc-skosmos.unibas.ch/en/	The Basel Register of Thesauri, Ontologies & Classifications (BARTOC) is a database of	Andreas Ledl as the Basel Register of Thesauri, Ontologies &	Common Library Network (GBV) in Gottingen		Computer Science, Social Sciences, Science,		KOS	KOS Registry		5000 terminologies and 100 terminology registries		Researchers in Social Sciences and	Skosmos is a web based open source ontology	Any	BARTO available
BARTOC Software inventory for controlled vocabularies	BARTOC.org: Software (gbv.de)	A wiki page to collect basic info about software for controlled vocabularies.	Common Library Network (GBV) in Gottingen	Common Library Network (GBV) in Gottingen				Controlled vocabulary	GitHub registry	https://github.com/gbv/bartoc.org/wiki/Software-for-c			Researchers in Social Sciences and			
BARTOC Vocabularies	https://bartoc.org/vocabularies	The Basel Register of Thesauri, Ontologies & Classifications (BARTOC) is a database of	Different creators	Different publishers	Each entry has its own URI	SSH	DDC, EuroVoc, ILC	Vocabulary	Registry		3289 vocabularies	KOS type	Researchers in Social Sciences and		Multilingual	Various
BBTalk	https://www.backbonethesaurus.eu/BBTalk/Login	It is a submission and correction management tool that facilitates communication between the	BBT maintenance WG	DARIAH		Arts & Humanities		Thesaurus management platform	Online service		n.a		Researchers in Humanities	Users can submit requests for possible changes	French German	
Bibliographic Ontology (BIBO)	http://bibliographic-ontology.org/	The Bibliographic Ontology Specification provides main concepts and properties for	Bruce D'Arcus Frédérick Giasson Structured	Structured Dynamics LLC	http://purl.org/ontology/bibo/	Semantic Web		Ontology	Repository	https://github.com/structureddynamics/Bibliographic		RDF, XML	Researchers in Social Sciences and	Can be used as a citation ontology, as a document	EN	The Bib Copyrig
Biblissima, the Observatory for Medieval and Renaissance Written Cultural Heritage	https://biblissima.fr/	A virtual library of libraries to discover the history of various texts and books that were written,		Biblissima		Cultural Heritage		Observatory	Repository				Researchers in Social Sciences and			CC0 1.0
BIO A vocabulary for biographical information	https://vocab.org/bio/	This document describes a vocabulary for describing biographical information about	Ian Davis Davie Galbraith		http://purl.org/vocab/bio/0.1/	Arts & Humanities	Biographical information (terms are grouped by theme:	Ontology	Repository			OWL	Researchers in Social Sciences and		EN	No info
CESSDA Vocabulary service	https://vocabularies.cessda.eu/#/discover	CESSDA Vocabulary Service enables users to discover, browse, and download controlled		CESSDA ERIC				Vocabulary	Service		28 vocabularies		Researchers in Social Sciences and			
COAR Controlled Vocabularies	https://www.coar-repositories.org/news-updates/what-we-do/controlled-vocabularies/	A collection of controlled vocabularies developed and maintained by the Editorial Board	COAR	COAR	http://purl.org/coar/	Repositories	Resource type vocabulary, Access Rights vocabulary,	Controlled vocabulary	Repository	https://www.coar-repositories.org/news-updates/wha		SKOS	Researchers in Social Sciences and	To enhance the interoperability across	Multilingual	
CREM Instruments		A terminology database of musical instruments		CREM Centre de Recherche en Ethnomusicologie	https://archives.crem-cnrs.fr/instruments/	Cultural Heritage	Instruments	Thesaurus	Repository			Dublin core	Researchers in Social Sciences and		EN	
Cyborgeo - EU Journal of Geography	https://journals.openedition.org/cyborgeo/33	A catalogue of 552 journals that allows search by keywords.		Open Edition Journal - a journals platform for the humanities and				Catalogue	Catalogue		552 journals		Researchers in Social Sciences and		Multilingual	
DBPedia	https://wiki.dbpedia.org/	A project aiming to extract structured content from the information created in the Wikipedia		Leipzig University				Open Knowledge Graph	Repository	https://github.com/dbpedia/			Researchers in Social Sciences and	Allows users to semantically query relationships and		GNU G
dcterms	https://standaarden.overheid.nl/owms/4.0/doc/eigenschappen	Metadata standards based on Dublin Core.		Dutch government.				Controlled vocabulary	Repository			XML, OWMS 4.0	Researchers in Social Sciences and		NL	
DDAlliance Controlled Vocabularies	https://ddalliance.org/controlled-vocabularies	A set of controlled vocabularies that can be used with DDI and other applications.		DDAlliance		Social & Behavioral Sciences		Controlled vocabulary	Repository			The vocabularies are published in an XML	Researchers in Social Sciences and			
Dewey Decimal Classification (DDC)	https://www.oclc.org/en/dewey/webdewey.html	The Dewey Decimal Classification—conceived by Melvil Dewey in 1873 and first	Melvil Dewey in 1873	OCLC				Classification system	Service				Libraries	The DDC is the most widely used classification	EN	Since it than 15
Dictionnaire de sociologie clinique	https://www.cairn.info/dictionnaire-de-sociologie-clinique--9782745257648.htm	The Dictionary of Clinical Sociology describes the central methods and questions of research	Agnès Vandevelde-Rougale, Pascal Fugier, Vincent de Gaulejac	CAIRN INFO		Social Sciences	Clinical sociology	Dictionary	?		245 terms	Book	Researchers in Social Sciences and		FR	
Dublin Core categories DCMI Metadata Terms	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/	An up-to-date specification of all metadata terms maintained by the Dublin Core Metadata			http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/			Controlled vocabulary	Repository		15 terms of the Dublin Core™ Metadata Element	RDF (creators of non-RDF metadata can	Researchers in Social Sciences and	Those terms are intended to be used in combination	EN	http://cr
DYAS Humanities Thesaurus	https://humanitiesthesaurus.academyofathens.gr/	The Thesaurus is the intellectual property of the Academy of Athens (AA), the National and	Academy of Athens (AA), the National and Kapodistrian University	BBT Backbone Thesaurus	https://humanitiesthesaurus.academyofathens.gr		Ancient history, Anthropology and ethnology	Thesaurus	Repository		1539 concepts	RDF, XML, X-Turtle	Humanities researchers		English Greek	https://c
EIONET Data Dictionary - Vocabularies	http://dd.eionet.europa.eu/vocabularies	A vocabulary repository with search functionalities.		European Environment Agency		Environment	Air quality directives, Birds directives, Habitats directives,	Vocabulary	Repository				Researchers in Social Sciences and			
Encoded Archival Description - tag library	https://www.loc.gov/ead/tag/eli/	Encoded Archival Description (EAD) is the international metadata transmission standard for		Developed by Society of American Archivists, published by Library of	https://www.loc.gov/ead/			Schemas	Repository	https://github.com/SAA-SDLA-EAD/ead-v1.1.1			Researchers in Social Sciences and			Creative Internat
EU Education Thesaurus	https://vocabularyserver.com/tee/en/index.php	Thesaurus of EU educational terms	Unknown	TermTree 3.1, vocabulary server	https://vocabularyserver.com/tee/en/	Education		Thesaurus	Repository		7710 terms	SKOS, ISO30.other	Educational		Multilingual	No info

91 entries

A few potential orphan vocabularies

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Vocabularies in SSH: **Challenges**

Domain coverage

Semantics

Creation & Data model

Link to existing vocabularies

Controlled Vocabularies

Taxonomies

Thesauri

Harmonization

Maintenance

Import into translation and authoring environments

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Locating suitable vocabularies

How the SSH Open Marketplace can help?

Matej Ďurčo, ACDH-CH

ICTeSSH Conference
SSHOC Vocabulary Initiative – What Users Want
28 June 2021

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabularies use & reuse

Vocabularies serve:

- Formalize/conceptualize a specific dimension/aspect of an application domain
- Make semantics explicit
  - Through verbose descriptions/definitions
  - Through relations between concepts
- Semantic interoperability across project and dataset boundaries
  - However only if they are being reused

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Where to look for useful/ already used (SSH) vocabularies?

- BARTOC - Basic Register of Thesauri, Ontologies & Classifications
  - > 3300 Vocabularies
  - No (partial) concept-based search/index
- LOV - Linked Open Vocabularies
  - Broad meaning of “vocabularies”!
- EU vocabularies
  - Good authority
  - multilingual!
- D3.1 Report on SSHOC (meta)data interoperability problems
  - Most used vocabularies in SSH: CESSDA Topic Classification, CLARIN Concept Registry, ISO 639-1 language list or TADIRaH

BARTOC.org
Basic Register of Thesauri, Ontologies & Classifications

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

BARTOC.org

BARTOC.org Vocabularies Registries Software About Contact

Vocabularies

Search Filter

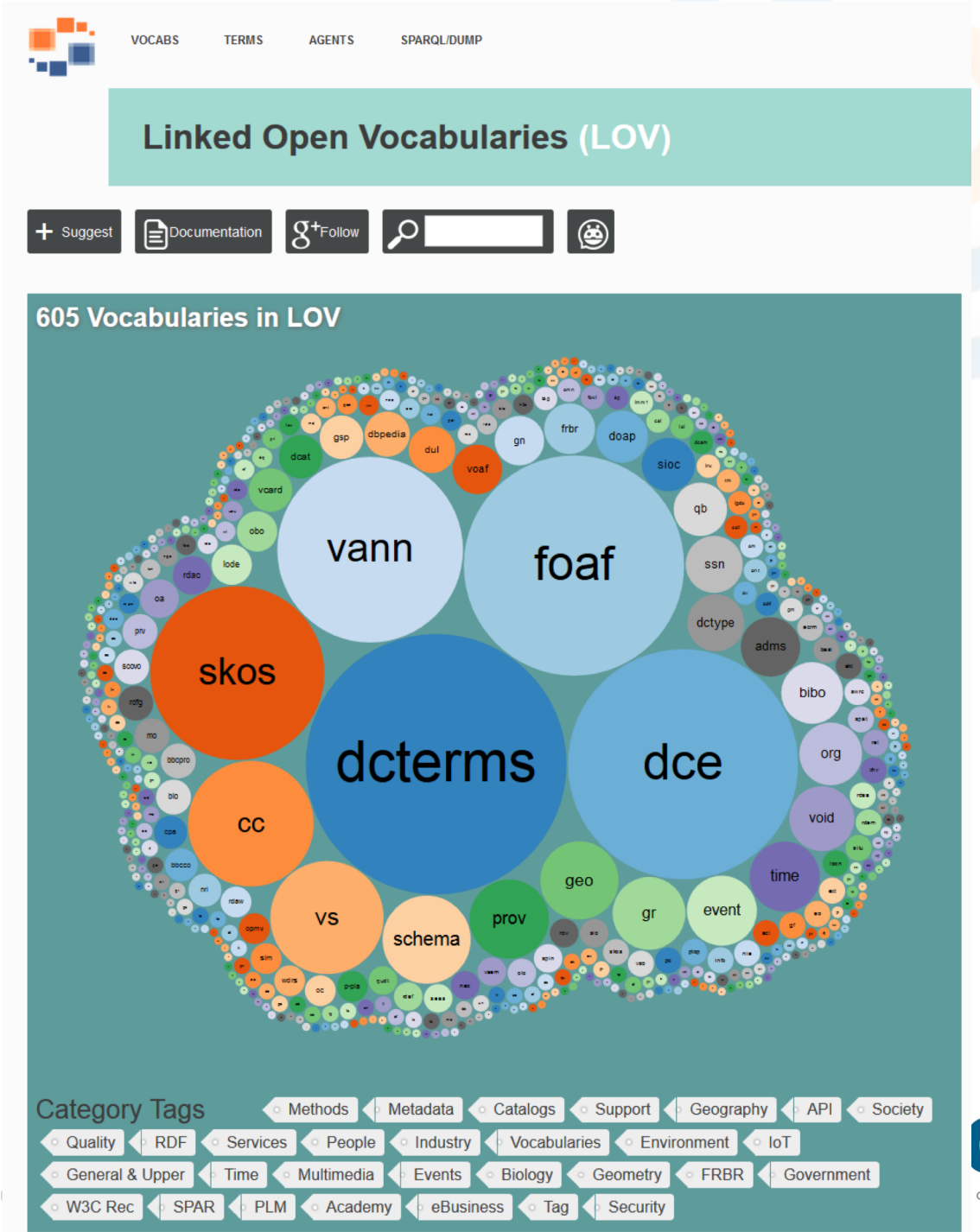
Search search

Full-text search across vocabulary description, ranked by relevance

1...500 501...1000 1001...1500 1501...2000 2001...2500 2501...3000 3001...3306

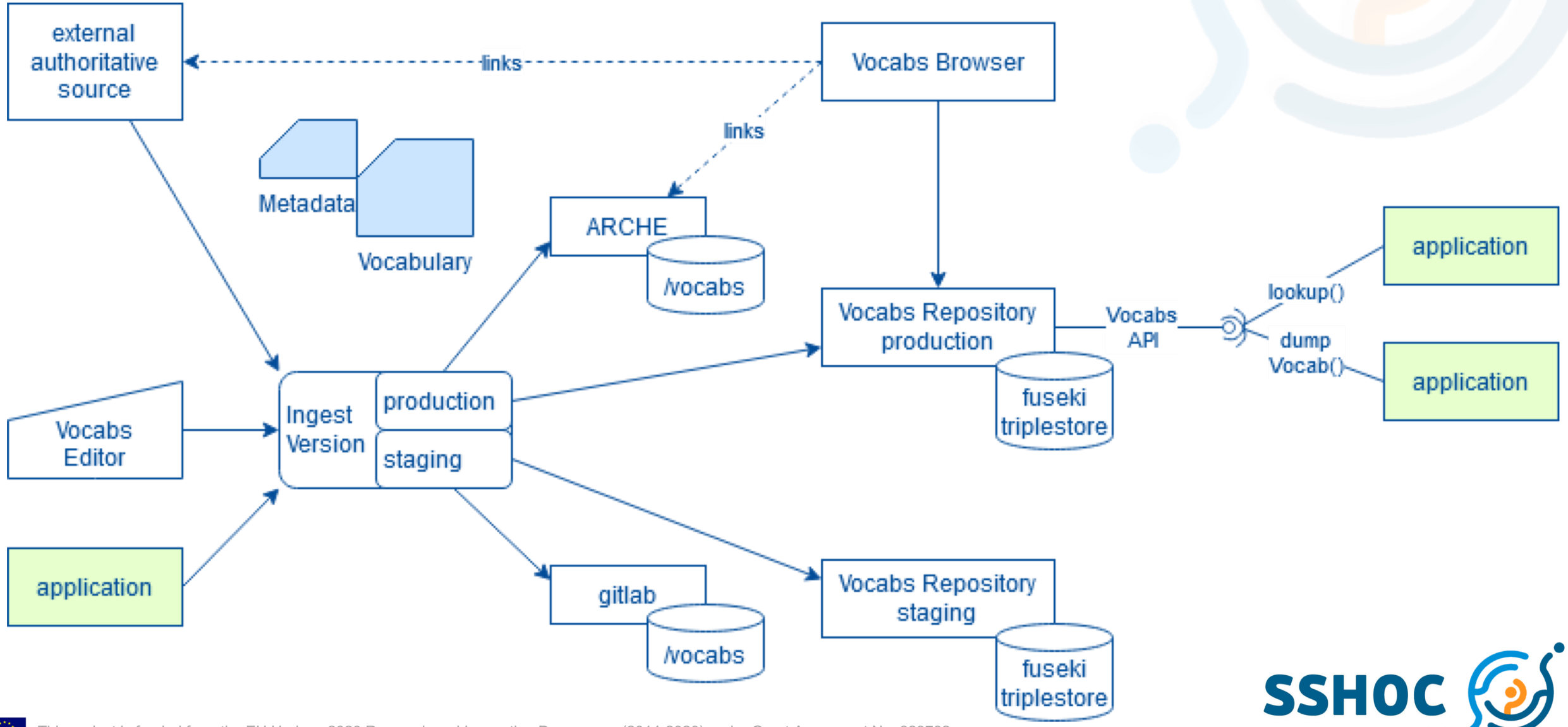
Name	Links
BARTOC data formats (bartoc-formats) Data formats for Knowledge Organization Systems as listed in BARTOC.org...	API
BARTOC access modes (bartoc-access) This vocabulary is used to specify the access modes of a knowledge organization system listed in BARTOC.org: free (the vocabulary is freely available), registered (access requires registration), or licensed (access requires a personalized license)....	API
Dewey Decimal Classification (DDC) "The Dewey Decimal Classification (DDC) system, devised by library pioneer Melvil Dewey in the 1870s and owned by OCLC since 1988, provides a dynamic structure for the organization of library collections...	API
ISO Language Codes (639-1 and 693-2) and IETF Language Types (ISO639) Comprehensive language code information, consisting of ISO 639-1, ISO 639-2 and IETF language types...	API
KOS Types Vocabulary "Related to this DC Application Profile is a KOS Type vocabulary...	API

Linked Open Vocabularies



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020)

Vocabulary management workflow (at ACDH-CH)



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

“Suitable”?

- Covering the “right” dimension
- Comprehensive
- Stable availability
- Stable reference to concepts
- Well-established
- Maintained (as opposed to orphaned)
- Never perfect match

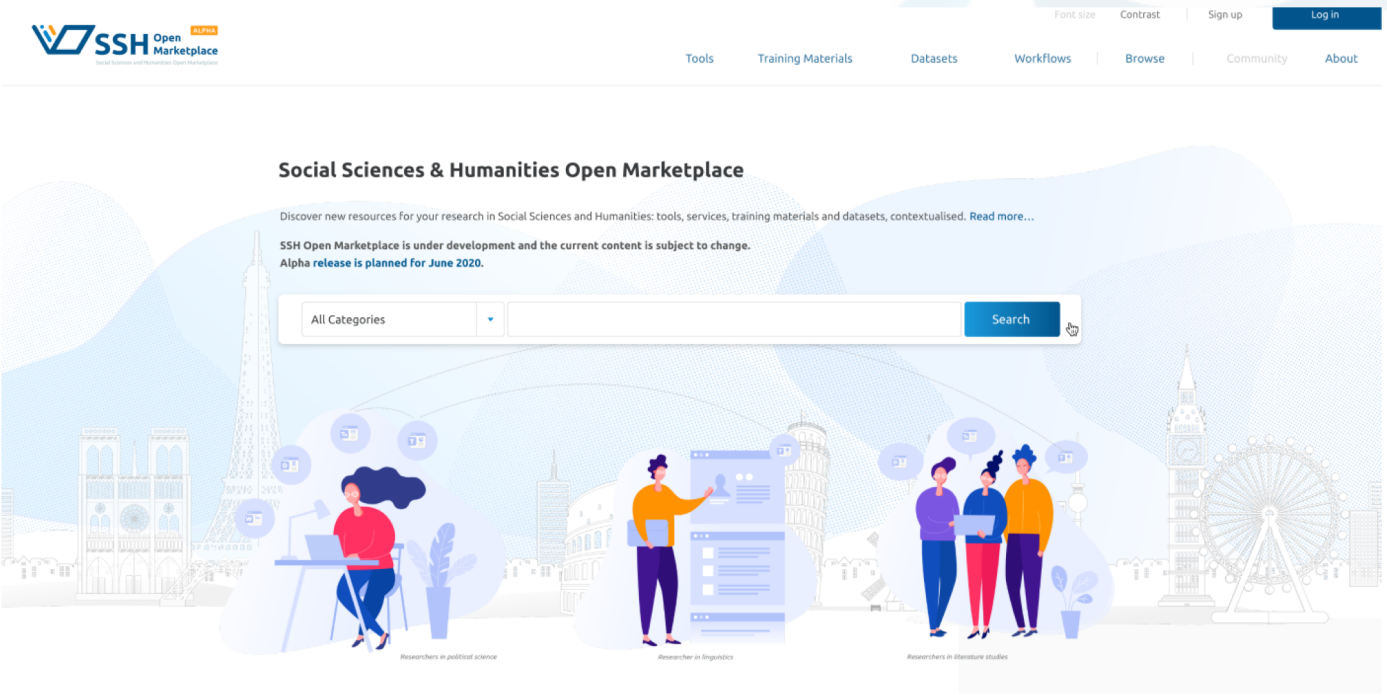
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSH Open Marketplace as one of the SSHOC services

- Discovery portal for SSH resources
  - Tools & services
  - Training materials
  - Workflows
  - Datasets
  - Publications

- 3 guiding principles
  - Contextualisation
  - Curation
  - Community



Beta version: marketplace.sshopencloud.eu

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Vocabulary entries in the SSH Open Marketplace

- Vocabularies as first-class citizens (represented as tools)

Contextualized (by related items):

- Training materials to support vocabulary use Controlled Vocabularies and SKOS
- Linked publications
- Examples of projects reusing the vocabularies
- ….

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

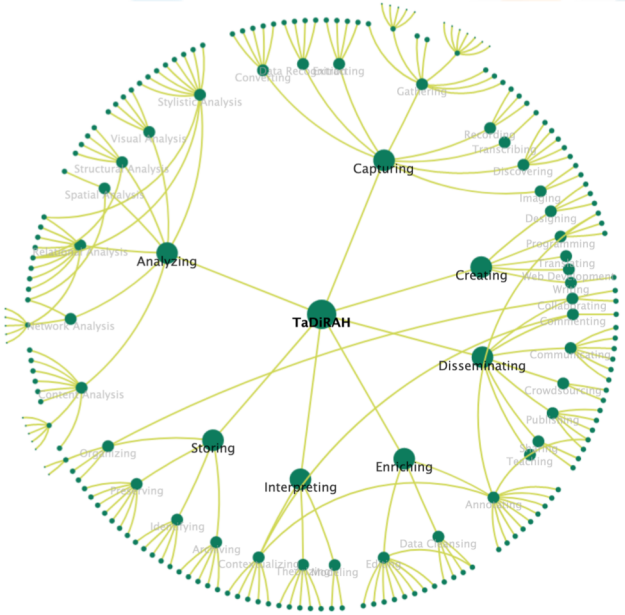
SSHOC
social sciences & humanities open cloud

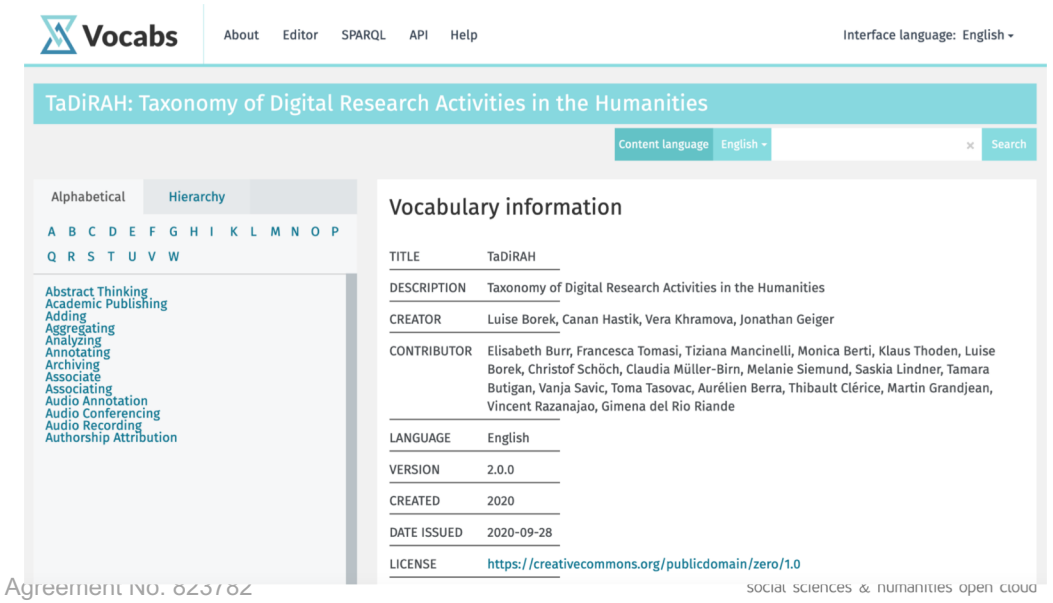
Referencing and contextualizing the most used SSH vocabularies - TaDIRAH example

TaDiRAH - Taxonomy of Digital Research Activities in the Humanities

[Github repo]
[Editors' Interview]
[Twitter]
[Linked publication]
Related services using TADIRaH

....



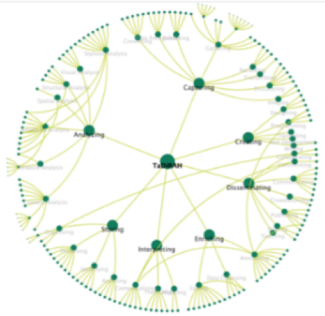


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

What it looks like in the SSH Open Marketplace

TaDiRAH - Taxonomy of Digital Research Activities in the Humanities

The taxonomy of digital research activities in the humanities has been developed for use by community-driven sites and projects that aim to structure information relevant to digital humanities and make it more easily discoverable. The taxonomy is expected to be particularly useful to endeavors aiming to collect information on digital humanities tools, methods, projects, or readings.



Go to Tool or service

Details

ACCESS

License: Creative Commons Zero v1.0 Universal

CATEGORISATION

Keyword: vocabulary

CONTEXT

See also: https://openmethods.dariah.eu/2021/02/10/openmethods-spotlights-2-interview-with-luise-borek-and-canan-hastik-about-tadirah/, https://tadirah.info/, https://twitter.com/tadirah_dh, https://epub.uni-regensburg.de/44951/1/isi_borek_et_al.pdf

TECHNICAL

Version: 2

EDITOR

Canan Hastik
Website

Jonathan Geiger
jonathan.geiger@adwmainz.de

Luise Borek

Vera Khramova
vera.khramova@stud.h-da.de

GitHub: https://github.com/dhtaxonomy/TaDiRAH

DOI: 10.5281/zenodo.32492

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Vocabulary mapping tool for archaeology in ARIADNEplus

Holly Wright, Archaeology Data Service

Ceri Binding & Douglas Tudhope

University of South Wales, Trefforest

tgn:7029392 World
- tgn:1000003 Europe
  - tgn:7008591 United Kingdom
    - tgn:7002443 Wales
      - tgn:7018963 Rhondda Cynon Taf
        - tgn:7441565 Trefforest

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

- ARIADNE
  - 24 partners, 13 countries, 9 languages, 27 subject vocabularies
  - 1.9 million data records aggregated/integrated
  - Subject vocabularies coordinated via mapping to Getty AAT – total 6416 mappings produced

- ARIADNE*plus*
  - 41 partners, 29 countries, 22 languages, ?? subject vocabularies
  - Data aggregation/integration work currently in progress
  - Reusing, revising and supplementing previous mappings
  - Adding vocabulary mappings from new data partners
  - Adding Wikidata mappings (multilingual entry vocabulary)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Why do we need vocabulary matching in ARIADNE?

- Source datasets not necessarily produced with aggregation, consolidation, cross-search and reuse in mind
- I say "*potato*", you say "*pomme de terre*", she says "*maris piper*" – multiple barriers to cross-searching subject metadata: language, punctuation, spelling, homonyms, synonyms, level of specificity
- Text-based search is limited by any/all of these
- Need to establish mutually agreed meaning…

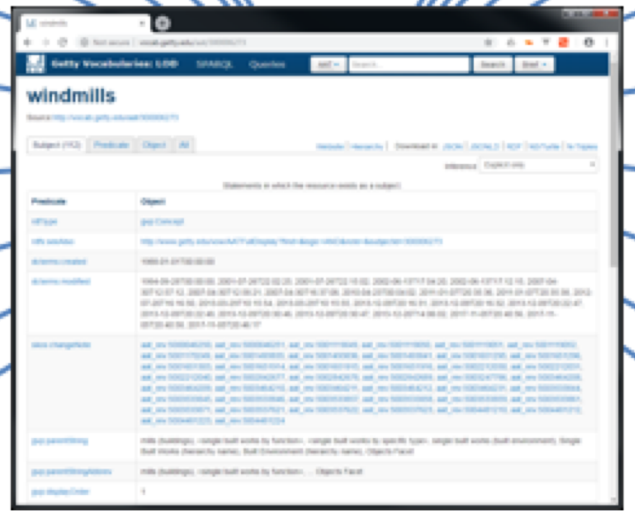
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Map local terms to a central concept

"windmill"@en
"windmolen"@nl
"Moulin à vent"@fr
„Windmühle"@de
"mulino a vento"@it
"melin wynt"@cy
"väderkvarn"@sv
"molino de viento"@es
"風車"@ja
"15. vuosisadan mainos"@fi
"vindmølle"@da
"větrný mlýn"@cs

The ***words*** may be different, but the ***concept*** is (more or less) the same…



"szélmalom"@hr
"moara de vant"@ro
"αιολικό μύλο"@el
"vindmylla"@is
"טחנת רוח"@he
"vindmølle"@no
"moinho de vento"@pt
"veterný mlyn"@sk
"mlin na veter"@sl
"вятърна мелница"@bg
"szélmalom"@hu
"muileann gaoithe"@ga

Now we can include any/all of these variants in a single query

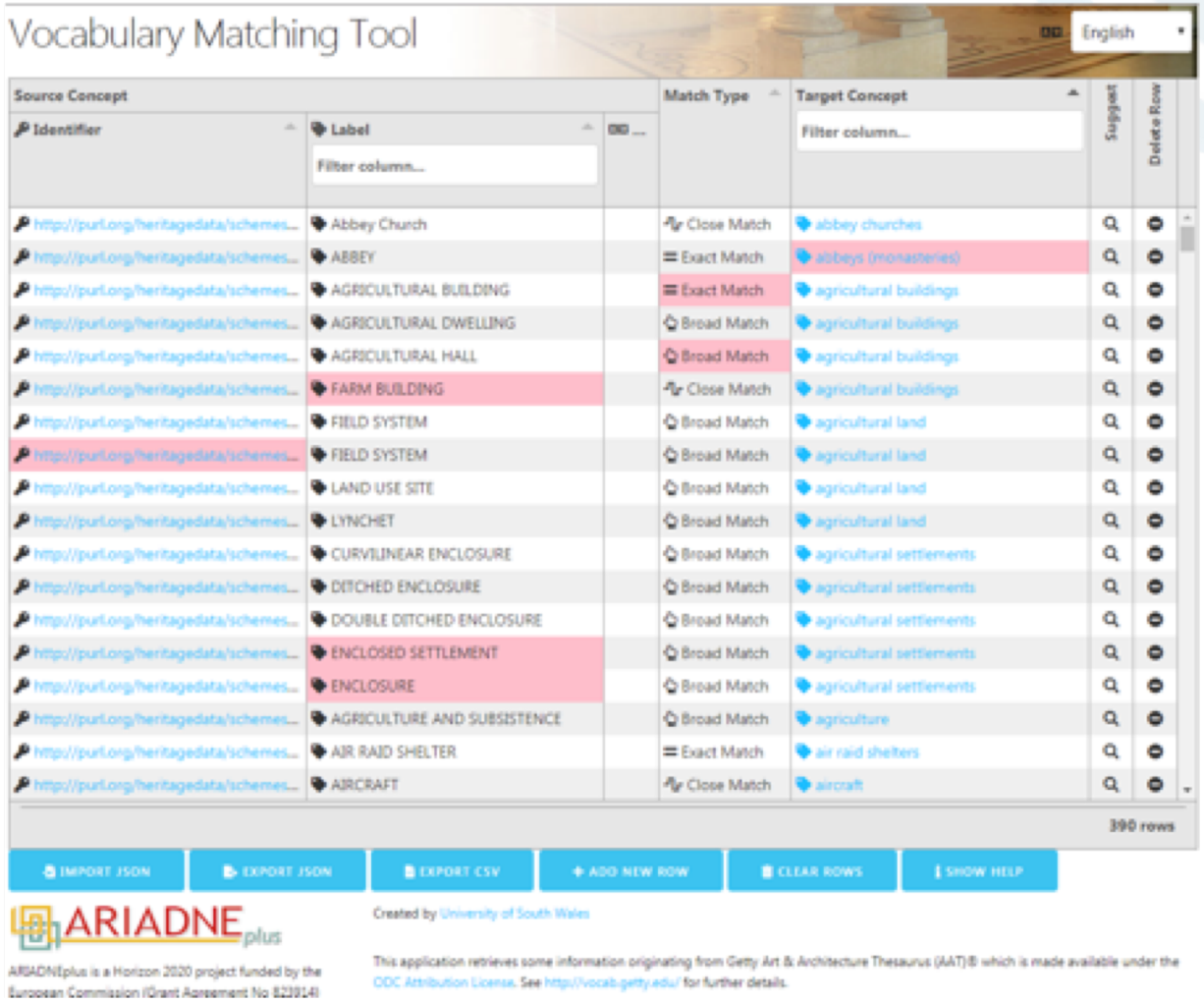
SSHOC
social sciences & humanities open cloud

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Vocabulary Matching Tool

https://vmt.ariadne.d4science.org/vmt/

- For matching local subject terms / concepts to Getty AAT concepts
- Search & browse Getty AAT
- No auto match: examine scope and context of source / target concepts



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Selected references and links

References

- Binding, C, Tudhope, D & Vlachidis, A 2018, 'A study of semantic integration across archaeological data and reports in different languages' Journal of Information Science, vol 45, no. 3, pp. 364-386. doi:10.1177/0165551518789874
- Binding, C & Tudhope, D 2016, 'Improving interoperability using vocabulary linked data' International Journal on Digital Libraries, vol 17, no. 1, pp. 5-21. doi:10.1007/s00799-015-0166-y

Links

- ARIADNEplus project: http://www.ariadne-infrastructure.eu/
- ARIADNE portal: https://ariadne-infrastructure.eu/portal/
- Vocabulary Matching Tool (VMT): https://vmt.ariadne.d4science.org/vmt/
- USW Hypermedia Research Group: https://hypermedia.research.southwales.ac.uk/

Contact

- ceri.binding@southwales.ac.uk ORCID: 0000-0002-6376-9613
- douglas.tudhope@southwales.ac.uk

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Use of MT for creating multilingual vocabularies

Monica Monachini, CNR-ILC & CLARIN IT

ICTeSSH, June 2021

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Topic: Why Do Multilingual Vocabularies matter

- Are essential for a proper description of resources and phenomena
- Help people find resources and determine their value
- Are critical in digital environments, where humans rely on computer processing for reliable and timely results
- Some use-cases in SSH:
  - users searching data in all languages

  - multi-country social surveys

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Challenges: Machine translations for multilingual vocabularies

- perform automatic translation of
  - occupation ontologies https://www.surveycodings.org/occupation-measurement
  - metadata and definitions (CLARIN CCR) https://concepts.clarin.eu/ccr/browser/
- using and testing different systems
  - “MT tool-suite” at UFAL https://lindat.mff.cuni.cz/services/translation/
  - “Google Translate” https://translate.google.com/
  - “DeepL” https://www.deepl.com/en/translator
  - “Reverso” www.reverso.net
- Languages: English French, Italian, German, Dutch, Greek, Czech, Slovene, Russian (evaluation running)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Challenges: Machine translations for multilingual vocabularies

		UFAL				DEEP-L (https://www.deepl.com/en/translator)				Google translate (https://translate.google.com				REVERSO		
inition	sou	term transl	te	definition translation (UFAL)	def	definition valid	ter	definition translat	def	term translation	ter	definition transla	de	term translatio	terr	c
esearch experiment i	(sou	magicien de	N	Expérience de recherche dans laquel	Y	magicien d'oz	N	Une expérience de r	Y	magiecien de l'oz	N	Une expérience de	Y	magicien de l'oz	N	L
visual representatio	(sou	systèmes d'é	Y	La représentation visuelle de la langu	Y	systèmes d'écritur	Y	La représentation vis	Y	systèmes d'écriture	Y	La représentation vi	Y	systèmes d'écritu	Y	L
cification of a persist	(sou	identifiant pe	M	Spécification d'un identificateur persi	M	identifiant persist	M	Spécification d'un id	Y	identifiant persistant	M	Spécification d'un ic	M	identifiant persist	M	S
access to the comm	(sou	public	Y	L'accès à l'événement de communica	Y	public	Y	L'accès à l'événemer	Y	publique	Y	L'accès à l'événeme	Y	public	Y	L
address of an organ	(sou	adresse	Y	L'adresse d'une organisation qui a pa	Y	adresse	Y	L'adresse d'une orga	Y	adresse	Y	L'adresse d'une org	Y	adresse	Y	T
number of years tha	(sou	âge	Y	Le nombre d'années de vie de quelqu	Y	âge	Y	Le nombre d'années	Y	âge	Y	Le nombre d'année	Y	âge	Y	L
estigator asks speake	(sou	élimé	N	L'enquêteur demande au ou aux locu	Y	a suscité	M	L'enquêteur demanc	Y	suscité	M	L'enquêteur deman	Y	obtenue	M	L
email address of a p	(sou	courriel	M	L'adresse électronique d'une personr	Y	email	Y	L'adresse électroniq	Y	email	Y	L'adresse e-mail d'u	Y	courriel	M	L
icates the structure c	(sou	structure évé	Y	Indique la structure de l'événement c	Y	structure des évér	M	Indique la structure	Y	structure de l'événer	M	Indique la structure	Y	structure d'événe	M	I
ntification of the loca	(sou	lieu d'exécut	Y	Identification de l'endroit où l'outil o	Y	lieu d'exécution	Y	Identification de l'en	Y	lieu d'exécution	Y	Identification de l'e	Y	emplacement d'e	M	I
ansmission of the co	(sou	cadre expéri	Y	Une transmission du contenu se déro	Y	milieu expériment	M	Une transmission du	Y	cadre expérimental	Y	Une transmission d	Y	cadre expériment	Y	T
transmission of the	(sou	face à face	Y	La transmission du message assure u	Y	face à face	Y	La transmission du n	Y	face à face	Y	La transmission du	Y	face à face	Y	L
trary to what is true	(sou	faux	Y	Contrairement à ce qui est vrai, error	Y	faux	Y	Contraire à ce qui es	Y	faux	Y	Contrairement à ce	Y	faux	Y	C
access to the comm	(sou	famille	Y	L'accès à l'événement de communica	Y	famille	Y	L'accès à l'événemer	Y	famille	Y	L'accès à l'événeme	Y	famille	Y	L
Fax number of a pe	(sou	numéro de te	M	Le numéro de télécopieur d'une pers	M	numéro de fax	Y	Le numéro de fax d'u	Y	numéro de fax	Y	Le numéro de fax d'	Y	numéro de téléco	M	N
nmunication event w	(sou	monologue	Y	Événement de communication avec u	Y	monologue	Y	Événement de comn	Y	monologue	Y	Événement de com	Y	monologue	Y	É
study of the structu	(sou	morphologie	Y	L'étude de la structure et de l'élector	N	morphologie	Y	L'étude de la structu	Y	morphologie	Y	L'étude de la structu	Y	morphologie	Y	L
speaker prepares in	(sou	prévu	N	L'intervenant prépare en détail la stru	Y	prévu	N	L'orateur prépare à l	Y	prévue	N	L'orateur prépare e	Y	prévue	N	L
icates in how far the	(sou	type de plani	Y	Indique dans quelle mesure le consul	Y	type de planificati	Y	Indique dans quelle	Y	type de planification	Y	Indique dans quelle	Y	type de planificat	Y	I
content is compose	(sou	poésie	Y	Le contenu est composé en vers ou e	Y	poésie	Y	Le contenu est compr	Y	poésie	Y	Le contenu est com	Y	poésie	Y	L

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Benefits for SSH Researchers

- Multilingual occupation ontologies
  - During the interviews, the respondent can self-select a job title from a list of occupations, and find the appropriate version (also the male vs. female form)
  - a multilingual occupational database where all titles are coded according to the International Standard Occupational Classification (ISCO)
- Multilingual controlled vocabulary
  - The user can perform a query in native language and retrieve data in all languages

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Final considerations

- At CNR, we are creating and experimenting with MT for creating multi-lingual vocabularies
- Platform where to host vocabularies: in general, the hosting of vocabularies (that do not have a safe home yet) is needed
- There is not yet a default solution for hosting and publishing such vocabularies in general, but we are looking into it.
- Vocabularies are also important data that should be FAIR

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Monica Monachini
CLARIN

Holly Wright
Archaeology Data Service

Daan Broeder
CLARIN

#SSHOCaVocabulary
#SSHOCifyCLARIN

Taina Jääskeläinen
Finnish Social Science Data Archive

Matej Durco
ACDH-CH

Iuliana van der Lek
CLARIN

SSHOC
social sciences & humanities open cloud

CLARIN

We are interested in learning more about your experience with vocabularies. Help us out by filling in the following polls!

ICTeSSH2021 SSH Vocabulary Initiative - What users want

17 - 29 Jun 2021

Poll results

slido

Table of contents

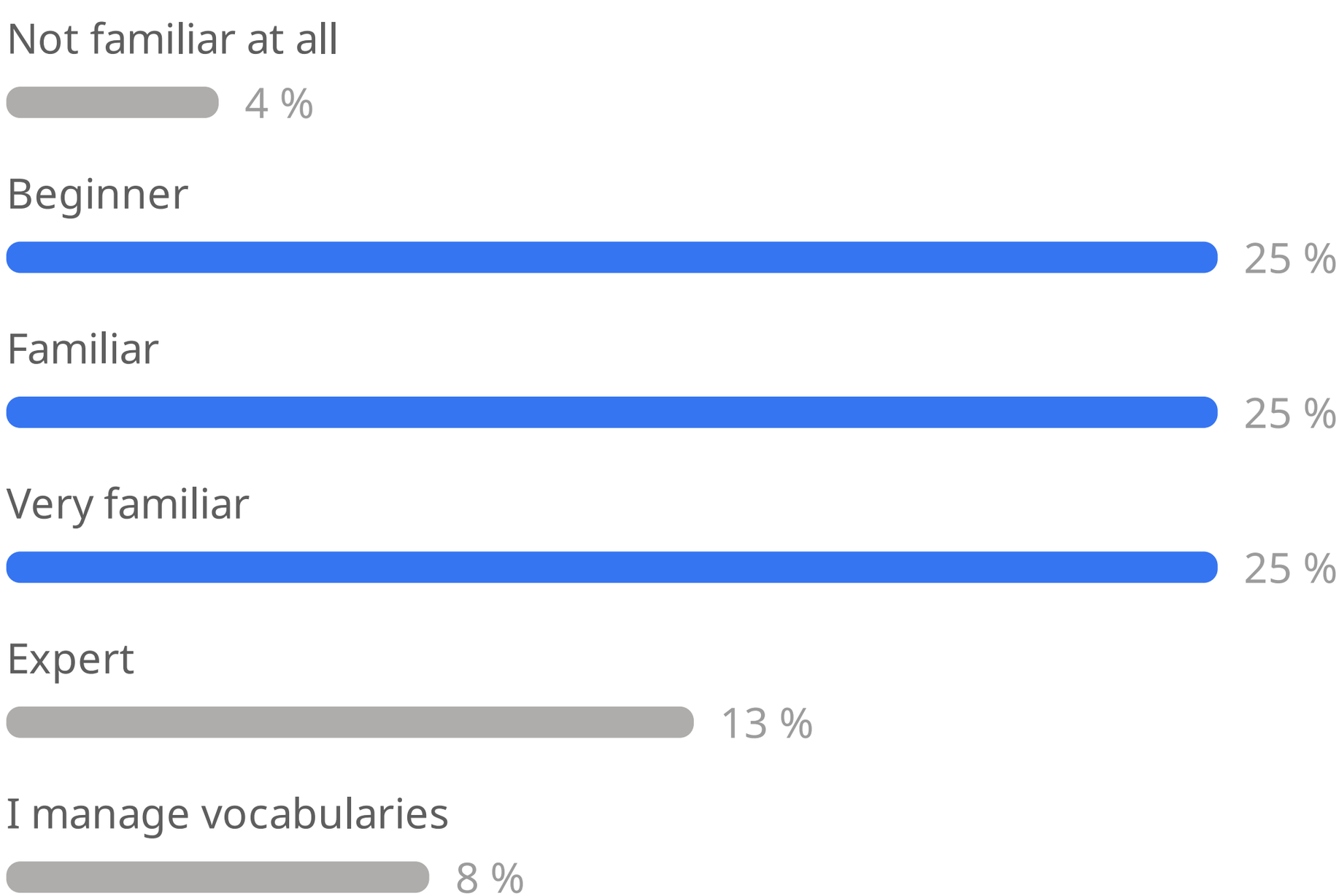
What is your role in your organisation?

0 2 2

Multiple-choice poll

Are you familiar with vocabularies?

0 2 4



Multiple-choice poll

How and where did you find the vocabulary/ies you decided to use?

0 2 5

Provided as part of the tool or workflow

8 %

Recommendations from peers

28 %

Recommendations from research infrastructures

32 %

Dedicated vocabulary registry

32 %

Other

0 %

Multiple-choice poll

How do you determine the quality of a vocabulary?

(1/2)

0 2 3

Intensive use

13 %

Provider

48 %

Recommendations by peers

4 %

Proscription

0 %

Standardisation

26 %

How do you determine the quality of a vocabulary?

(2/2)

0 2 3

Tradition

4 %

Other

4 %

Multiple-choice poll

How do you handle vocabulary that fails to cover your descriptive needs?

0 2 2

I add additional descriptions in a description field.

5 %

I propose new terms to the vocabulary manager.

32 %

I choose the closest alternative.

18 %

I don’t use that vocabulary.

27 %

Other

18 %

Mapping (conversion) between entries of different vocabularies, how do you manage that?

0 1 3

- not applicable
- In a shared spreadsheet.
- Sincerely, I don't convert. Reason is, I am not exposed yet to such tools in the part of the world I'm in.
- Cross walks
- I don't.
- I don't
- Ontology mapping tools (e.g., LOOM, AML, YAM++) and then the use of ontology repositories to share mappings.
- Start comparison on excel spreadsheet
- I don't
- With a lot of frustration ;)
- No good solution yet found, need to try it out with different tools and select the best for my needs
- Automated or manual?
- I don't.

Multiple-choice poll

Do you use vocabularies in your own language?

0 1 7

Yes

76 %

No

0 %

No, I use English versions

12 %

Would like to, if they would be available

12 %

What kind of support do you expect from research infrastructures/ support services for vocabularies?

0 1 1

- Guidance on which vocabulary to use for a particular purpose. Quality, authority, sustainability. Backed by the community. Fit for purpose. Domain relevance. Machine access.
- Ease of access, guidance on which vocabulary to use and support in terms of mapping
- Hosting a mapping tool
- to put a quality stamp to a certain vocabulary
- Support for as many vocabularies as feasible, crosswalks.
- Registries and vocabs services like mapping
- I expect them to hide all interoperability complexity
- advice how to create vocabs
- Ease of access.
- Suitable repos
- Technical, advisory support

SSHOC
social sciences & humanities open cloud

SSH Vocabulary Initiative
What users want

June 28, 9:00 - 10:30 CEST

ICTeSSH 2021

sshopencloud.eu/register
sshvocabularyinitiative@sshopencloud.eu

Thank you for your attention!

Join our community

sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/company/sshoc

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782