

Geslovniki, ontologije in taksonomije v družboslovju in (digitalni) humanistiki

Andrej PANČUR, Inštitut za novejšo zgodovino

Irena VIPAVC BRVAR, Arhiv družboslovnih podatkov, Fakulteta za družbene
vede Univerze v Ljubljani

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,
Usposabljanje podatkovnih strokovnjakov, 14. – 17. 10. 2024



Geslovniki in taksonomije

- Lahko zelo projektno specifični
- Težnja (ne pa pravilo) po uporabi splošno sprejetih
- V TEI skupnosti razširjena uporaba taksonomij, vključenih v `teiHeader`:
`<classDecl>` vsebuje eno ali več taksonomij, ki opredeljujejo morebitne klasifikacijske kode, uporabljene drugje v besedilu. Lahko sama navedba taksonomije ali pa natančno razdelana

 TomazErjavec Fix spacing and newlines in taxonomies. ✓

Name



..

 ParlaMint-taxonomy-CHES.xml

 ParlaMint-taxonomy-NER.ana.template.xml

 ParlaMint-taxonomy-NER.ana.xml

 ParlaMint-taxonomy-UD-SYN.ana.xml

 ParlaMint-taxonomy-USAS.ana.xml

 ParlaMint-taxonomy-merge.log

 ParlaMint-taxonomy-parla.legislature.template.xml

 ParlaMint-taxonomy-parla.legislature.xml

 ParlaMint-taxonomy-politicalOrientation.template.xml

 ParlaMint-taxonomy-politicalOrientation.xml

 ParlaMint-taxonomy-speaker_types.template.xml

 ParlaMint-taxonomy-speaker_types.xml

 ParlaMint-taxonomy-subcorpus.template.xml

 ParlaMint-taxonomy-subcorpus.xml

 taxonomy-translation-include.tsv

 taxonomy-translation-responsibility.tsv



Ontologije in semantični splet

- Ontologije pridobivajo na pomenu
- semantični (RDF) trojček: subjekt (osebek) → predikat (povedek) → objekt (predmet)
- `http://primer.ime#JanezNovak1083` →
`http://xmlns.com/foaf/0.1/pisati` → <http://primer.objekt#knjiga>
- semantični (RDF) grafi
- Da bo delovalo, je namesto projektno specifičnih oznak je nujno potrebno uporabljati splošno sprejete ontologije

Splošne (osebe, kraji, dogodki, čas, dokumenti)

- CIDOC Conceptual Reference Model (CRM) – zagotavlja definicije in formalno strukturo za opisovanje implicitnih in eksplicitnih konceptov in razmerij, ki se uporabljajo v dokumentaciji kulturne dediščine
- CRMinf Argumentation Model – razširitev CIDOC-CRM za podporo argumentaciji
- FOAF - povezovanje ljudi in informacij
- Functional Requirements for Bibliographic Record (FRBR) – za opis dokumentov in njihovega razvoja
- Geonames – pokriva vse države in vsebuje več kot enajst milijonov imen krajev, ki so na voljo za brezplačen prenos
- GeoSPARQL – prostorske ontologije, plus izbirna razširitev SPARQL.
- IDS - The Intermediate Data Structure (IDS) je standardni format
- podatkov, ki ga je sprejelo več velikih longitudinalnih podatkovnih baz o zgodovinskih populacijah.
- INTRO - ontologija medbesedilnih odnosov za literarne vede



Splošne (osebe, kraji, dogodki, čas, dokumenti)

- PROV ontologija - niz razredov, lastnosti in omejitev, ki se lahko uporabljajo za predstavitev in izmenjavo informacij o provenienci
- LIO - lahka ontologija za opisovanje slik
- OWL Time - časovna ontologija v OWL
- Perio.do - zgodovinska obdobja
- Simple Event Model (SEM) - ontologija za modeliranje dogodkov
- Simple Knowledge Organization System (SKOS) – standardi za podporo uporabe sistemov organizacije znanja (KOS), kot so tezavri, klasifikacijske sheme, sezname predmetnih oznak in taksonomije v okviru semantičnega spleta
- schema.org - sheme za strukturirane podatke na internetu, na spletnih straneh, v e-poštnih sporočilih in drugod
- Virtual International Authority File (VIAF) – združuje več avtoritetnih datotek za imena v eno samo avtoritetno storitev za imena
- Normativna datoteka osebnih in korporativnih imen (CONOR-SI)

Kulturna dediščina (objekti kulturne dediščine, muzeji, arheologija itd.)

- Linked Art Data Model– CIDOC-CRM profil aplikacije, ki se lahko uporablja za opisovanje virov kulturne dediščine, s poudarkom na umetniških delih in muzejsko usmerjenih dejavnostih.
- CRMArcheo Excavation Model – ontologija za kodiranje metapodatkov o procesu arheološkega izkopavanja (razširitev CIDOC-CRM)
- Europeana Data Model (EDM) - za Europeano.
- Getty AAT Art and Architecture Thesaurus - razvijajoč se besednjak za umetnost in arhitekturo, ki raste in se spreminja
- JUSO - Ontologija Juso je spletni besednjak za opis geografskih naslovov in značilnosti
- ROAR - Ontologija za opis oseb, lokacij itd. opazovanj v arhivskih virih
- ICONCLASS - večjezični sistem klasifikacije kulturnih vsebin



Muzikologija (partiture, glasbeni metapodatki, simbolni zapisi)

- DOing REusable MUSical data (DOREMUS) datamodel - dogodki, dela, izrazi okoli klasičnih glasbenih podatkov
- MIDI ontologija - za objavo MIDI (Musical Instrument Digital Interface) skladb, dogodkov in njihovih atributov
- MusicOWL - ontologija za kodiranje partitur zahodne glasbe
- Music ontology (MO) - ponuja glavne koncepte in lastnosti za opisovanje glasbe (tj. izvajalcev, albumov in skladb) na semantičnem spletu



Jezik (leksikologija, prepoznavanje imenskih entitet itd.)

- Lemon – model za modeliranje leksikonov in strojno berljivih slovarjev, v povezavi s semantičnim spletom
- Lexvo – prinaša informacije o jezikih, besedah, znakih in drugih entitetah, povezanih s človeškim jezikom, v povezani s semantičnim spletom



Publikacije (literatura, časopisi, revije)

- Enumeration and Chronology of Periodicals Ontology (ECPO) opredeljuje skupne bibliografske izraze za opis naštevanja in kronologije, ki se običajno uporabljajo za serijske publikacije.
- Drammar - celovita ontologija dramskih besedil
- Literary Theme Ontology (LTO) – taksonomija definiranih literarnih tem za tematsko označevanje leposlovnih del



Zakaj

Usklajenost

Uporabnost v raziskovalnem procesu

Interoperabilnost / najdljivost in filtriranje po ključnikih



Možnosti kreiranja - izzivi

Avtomatsko

Obdelave naravnega jezika

Večjezičnost

Podatkovni modeli, ontologije in taksonomije povezani z repozitoriji

- *Dublin Core* (**DC**, Dublin, Ohio); skupnostna metapodatkovna shema, danes osnovna shema za opis objektov [<http://purl.org/dc>]
- *Portland Common Data Model* (**PCDM**, Portland); model repozitorijskih objektov, ki omogoča interoperabilnost [<https://github.com/duraspace/pcdm/>]
- *Confederation of open access repositories* (**COAR**): access rights, resource types, version types [<https://vocabularies.coar-repositories.org/>]
- *International Image Interoperability Framework* (**IIIF**); sklop odprtih standardov in protokolov, ki institucijam omogočajo objavljanje visokokakovostnih digitalnih slik, povezanih metapodatkov in naprednih funkcionalnosti na spletu [<https://iiif.io/>]

Včasih se uporabljajo tudi ontologije višjega reda, najpogosteje:

- *Simple knowledge organization system* (**SKOS**, W3 konzorcij); standard za konceptualiziranje ontologij in taksonomij [<https://www.w3.org/TR/skos-reference/>]
- *Linked data platform vocabulary* (**LDP**, W3 konzorcij); ontologija konceptov in izrazov za področje povezanih podatkov [<https://www.w3.org/ns/ldp>]
- *Friend of a friend* (**FOAF**); [<http://xmlns.com/foaf/spec/>]

Ontologije, nadzorovani besednjaki, normativni imeniki

Univerzalni ontologiji

- *Wikipodatki (WD, Wikidata)*; hišna Wiki ontologija in graf znanja; prednost/slabost, da lahko vsakdo dodaja vrednosti; postaja vozlišče različnih ontologij
- *Schema.org*; univerzalni nadzorovani besednjaki s primarno nalogo strukturiranja vsebin na spletu

Osebe & organizacije

- *Virtual International Authority File (VIAF)* [<https://viaf.org/>]
- *Normativna datoteka osebnih in korporativnih imen (CONOR-SI)* [<https://plus.cobiss.net/cobiss/si/sl/conor>]
- **GND ID** (*Katalog der Deutschen Nationalbibliothek*); **BNF ID** (*Bibliothèque nationale de France ID*); ...
- *Open Researcher and Contributor ID (ORCID)* [<https://orcid.org/>]
- *Research Organization Registry (ROR)* [<https://ror.org/>]

Zemljepisna imena:

- *Geonames* [<http://www.geonames.org/ontology>]
- *Geographic technology standard models & schemas (ISO/TC 211)* [<https://www.isotc211.org/>]
- *EIONET data dictionary: LAU2, NUTS, INSPIRE ...* [<https://dd.eionet.europa.eu/>]

Domenske taksonomije, nadzorovani besednjaki

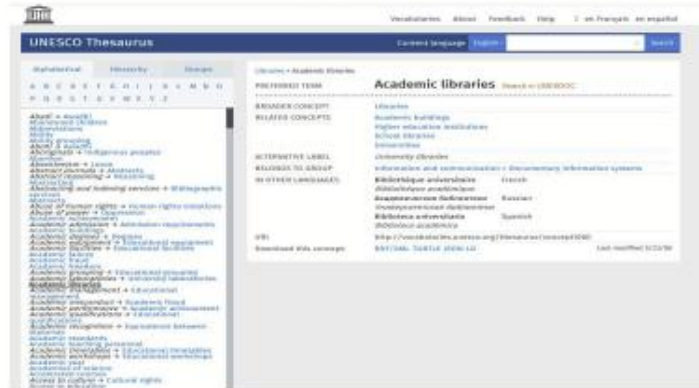
- *Language code agency (LCA)*; ISO 639.3 [<http://www.lexvo.org/>]
- *Getty vocabularies*: Art & Architecture Thesaurus, Cultural Objects Name Authority, Union List of Artist Names [<https://www.getty.edu/research/tools/vocabularies/>]
- *Fields of Science and Technology (FOS)*; [Frascati manual OECD, Aris: <https://www.aris-rs.si/sl/gradivo/sifranti/sif-frascati.asp>]
- *EU vocabularies*; med drugim gostijo nekaj besednjakov povezanih z repozitoriji [<https://op.europa.eu/en/web/eu-vocabularies/data-catalogue>]
- *Relators (LOC, Library of Congress)*; [<http://id.loc.gov/vocabulary/relators/>]

- Spletni brskalnik in orodje za objavlanje besedišča SKOS
 - SKOS je standard povezanih podatkov za predstavitev tezavrov, naslovov predmetov, taksonomij, klasifikacij, lahkih ontologij ...
- Olajša predmetno indeksiranje, iskanje informacij in razvoj besedišča
- Razvila ga je nacionalna knjižnica Finske
 - s prispevki mnogih drugih
 - Odprta koda (licenca MIT)

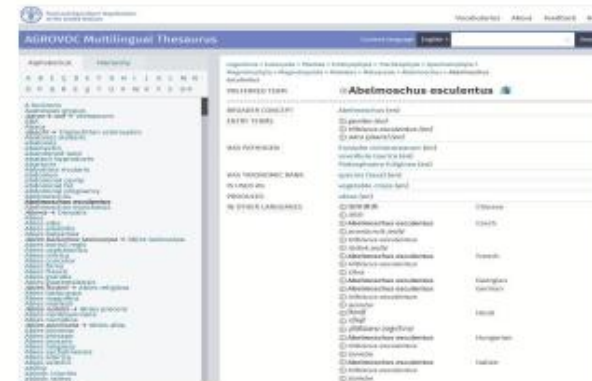
Skosmos

SKOSMOS po svetu

UNESCO Thesaurus



AGROVOC Thesaurus (FAO)



ChemSKOS



ELSST Thesaurus (CESSDA)



Loterre



CESSDA in ELSST



ELSST Thesaurus (Version 5 - 2024)

Content language

English ▾



Search

Alphabetical

Hierarchy

New and
Deprecated

A B C D E F G H I J K L M N O
P Q R S T U V W Y Z

ABILITY
ABILITY ASSESSMENT → ABILITY EVALUATION
ABILITY EVALUATION
ABILITY GROUPING
ABJURATION OF FAITH → RELIGIOUS EXPERIENCE
ABORIGINAL PEOPLE → INDIGENOUS PEOPLES
ABORTION
ABSENCE FROM SCHOOL → EDUCATIONAL ATTENDANCE
ABSENCE FROM WORK → ABSENTEEISM (WORK)
ABSENTEEISM (WORK)
ABSENT PARENT → PARENTAL DEPRIVATION
ABSENT PARENTS → PARENTAL DEPRIVATION
ABUSED CHILDREN → CHILD ABUSE
ABUSE OF THE ELDERLY → ELDER ABUSE
ACADEMIC ABILITY
ACADEMIC ACHIEVEMENT
ACADEMIC ADMISSION → EDUCATIONAL ADMISSION
ACADEMIC APTITUDE → ACADEMIC ABILITY
ACADEMIC DEPARTMENTS
ACADEMIC EVALUATION → EDUCATIONAL ASSESSMENT
ACADEMIC FACILITIES → EDUCATIONAL FACILITIES
ACADEMIC FREEDOM

Vocabulary information

TITLE	ELSST Thesaurus (Version 5 - 2024)
DESCRIPTION	The European Language Social Science Thesaurus (ELSST) is a broad-based, multilingual thesaurus for the social sciences. It is owned and published by the Consortium of European Social Science Data Archives (CESSDA) and its national Service Providers. The thesaurus consists of over 3,400 concepts and covers the core social science disciplines: politics, sociology, economics, education, law, crime, demography, health, employment, information and communication technology, and environmental science.
PUBLISHER	CESSDA ERIC
LICENSE	Creative Commons Attribution-ShareAlike 4.0 International License (CC-BY-SA 4.0) http://creativecommons.org/licenses/by-sa/4.0
VERSION	5
IDENTIFIER	urn:ddi:int.cessda.elsst:00000000-0000-0000-0000-000000000001:5
DC:ISVERSIONOF	https://elsst.cessda.eu/id/
DATE ISSUED	2024-09-23
RIGHTS HOLDER	Consortium of European Social Science Data Archives (CESSDA) and its national Service Providers
OWL:IMPORTS	http://purl.org/dc/terms/

CESSDA in ELSST, postopek dela

- Redna srečanja prevajalcev/arhivistov iz vključenih arhivov za pregled in identifikacijo potreb po spremembah ali izboljšavah že vključenih terminov
- Vsak arhiv lahko predlaga dodatne termine, ki so pomembni za nacionalni kontekst in ki bi služili izboljšanju metapodatkovnih opisov raziskav na nacionalni ravni ter posledično izboljšalo najdljivost v evropskem prostoru
- posebnost slovenskega jezika: prevodi so zaradi posebnosti slovenščine pri sklanjatvah vedno dodani v ednini, razen v primeru množinske rabe

CESSDA in ELSST, postopek dela

- Vnos in spremembe terminov potekajo preko vmesnika VocBench

Current project: ELSST_PRODUCTION_R5 Global Data Management ↻

VocBench Data Metadata ▾ SPARQL History Tools ▾

Class Concept Scheme Collection Property Datatype

ABSOLVENTI VYSOKÝCH ŠKOL (cs), ABSOLVENTEN (de), ΠΤΥΧΙΟΥΧΟΙ (el), GRADUATES (en), PERSONAS LICENCIADAS (es), KORKEAKOULUSTA VALMISTUNEET (fi), DIPLÔMÉS (fr), DIPLOMÁS (hu), ÚTSKRIFTARNEMAR UNIVERSITETS OG HØYSKOLEUTDANNET (no), ABSOLVENTI (ro), DIPLOMANT (sl), EXAMINERADE (sv)

ANALÝZA (cs), ANALYSE (de), ΑΝΑΛΥΣΗ (el), ANALYSIS (en), ANÁLISIS (es), ANALYYSI (fi), ANALYSE (fr), ELEMZÉS (hu), GREINING (is), ANALIZÉ (it), ANALYSE (nl), ANALYSE (no), ANALIZE (ro), ANALIZA (sl), ANALYSER (sv)

ANALÝZA DAT (cs), DATENANALYSE (de), ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ (el), DATAANALYSIS (en), ANÁLISIS DE DATOS (es), AINEISTON ANALYYSI (fi), ANALYSE DE DONNÉES (fr), ADATELEMZÉS (hu), GAGNAGREINING (is), DATAANALYSE (no), ANALIZĂ DE DATE (ro), ANALIZA PODATKOV (sl), DATAANALYS (sv)

ANTROPOLOGIE (cs), ANTHROPOLOGIE (de), ΑΝΘΡΩΠΟΛΟΓΙΑ (ΚΟΙΝΩΝΙΚΗ) (el), ANTHROPOLOGY (en), ANTROPOLOGÍA (es), ANTROPOLOGIA (fi), ANTHROPOLOGIE (fr), ANTROPOLÓGIA (hu), MANNFRÆÐI (is), ANTROP (no), ANTROPOLOGIE (ro), ANTROPOLOGIJA (sl), ANTROPOLOGI (sv)

BEHAVIORÁLNÍ VĚDY (cs), VERHALTENSWISSENSCHAFTEN (de), ΕΠΙΣΤΗΜΕΣ ΜΕΛΕΤΗΣ ΤΗΣ ΑΝΘΡΩΠΙΝΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ (el), BEHAVIOURAL SCIENCES (en), CIENCIAS DEL COMPORTAMIENTO (es), KÄYTTÄYTYMIS' VISELKEDÉSTUDOMÁNY (hu), ATFERLISVÍSINDI (is), ELGSEÑOS MOKSLAI (lt), GEDRAGSWETENSCHAPPEN (nl), ATFERDSVITENSKAPENE (no), ȘTIINȚE COMPORTAMENTALE (ro), VEDENJSKA ZNANOST (sl), BETEENDE

BIOGRAFIE (cs), LEBENSGESCHICHTEN (de), ΙΣΤΟΡΙΕΣ ΖΩΗΣ (el), LIFE HISTORIES (en), HISTORIA BIOGRÁFICA (es), ELÄMÄNHISTORIA (fi), HISTOIRES DE VIE (fr), ÉLETTÖRTÉNET (hu), LÍFSSÖGUR (is), GYVENIMO ISTO (no), ISTORII DE VIAȚĂ (ro), ŽIVLJENJSKA ZGODOVINA (sl), LEVNADSHISTORIA (sv)

BIOLOGICKÁ OCHRANA (cs), PERSISTENTE ORGANISCHE SCHADSTOFFE (de), ΕΜΜΟΝΟΙ ΟΡΓΑΝΙΚΟΙ ΡΥΠΟΙ (el), PERSISTENT ORGANIC POLLUTANTS (en), CONTAMINANTES ORGÁNICOS PERSISTENTES (es), PYSYV, ORGANIQUES PERSISTANTS (fr), KÖRNYEZETBEN TARTÓSAN MEGMARADÓ SZERVES SZENNYEZŐ ANYAG (hu), ÞRÁVIRKIR LÍFRÆNIR MENGUNARVALDAR (is), PATVARIEJI ORGANINIAI TERŠALAI (lt), ORGANISCHE VI ORGANISCHE FORURENSNINGER (no), CONTROL BIOLOGIC (ro), OBSTOJNI ORGANSKI ONESNAŽEVALEC (sl), LÄNGLIVADE ORGANISKA FÖRORENINGAR (sv)

BIOLOGIE (cs), BIOLOGIE (de), ΒΙΟΛΟΓΙΑ (el), BIOLOGY (en), BIOLOGÍA (es), BIOLOGIA (fi), BIOLOGIE (fr), BIOLÓGIA (hu), LÍFFRÆÐI (is), BIOLOGIJA (lt), BIOLOGIE (nl), BIOLOGI (no), BIOLOGIE (ro), BIOLOGIJA (sl), BIOLOGI

BLAHOBYT (SPOLEČENSKÝ) (cs), SOZIALES WOHLBEFINDEN (de), ΕΥΗΜΕΡΙΑ (ΚΟΙΝΩΝΙΑΣ) (el), WELL-BEING (SOCIETY) (en), BIENESTAR (SOCIEDAD) (es), YHTEISKUNNAN HYVINVOINTI (fi), BIEN-ÊTRE (SOCIÉTÉ) (fr), (is), GEROVÉ (VISUOMENÉ) (lt), WELZIJN (MAATSCHAPPIJ) (nl), VELVÆRE (SOSIALT) (no), BUNASTARE (SOCIETATE) (ro), BLAGINJA (sl), VÄLBEFINNANDE (SOCIALT) (sv)

CESSDA in ELSST, postopek dela

- pregled terminov po različnih jezikih

The screenshot shows a web browser window with the address bar displaying the URL: <https://elsst.cessda.eu/id/5/54561e90-8d00-4d87-9187-adc81f6c9902>. The browser tab is titled "BIOGRAFIE (cs), LEBENSGESCHICH...". The main content area is titled "Lexicalizations:" and displays a list of terms under the label "skos:prefLabel". Each term is preceded by a small flag icon representing its language. The list includes:

- BIOGRAFIE
- LEBENSGESCHICHTEN
- ΙΣΤΟΡΙΕΣ ΖΩΗΣ
- LIFE HISTORIES
- HISTORIA BIOGRÁFICA
- ELÄMÄNHISTORIA
- HISTOIRES DE VIE
- ÉLETTÖRTÉNET
- LÍFSSÖGUR
- GYVENIMO ISTORIJS
- LEVENSGESCHIEDENIS
- LIVSLØP
- ISTORII DE VIAȚĂ
- ŽIVLJENJSKA ZGODOVINA
- LEVNADSHISTORIA

CESSDA in CVS

35 Vocabularies

Results per page **30**

Sort by **CV name (A-Z)**

▼ Agency

Search agencies

gesis

Central concepts of R|EX research (*rexconcepts*)

This CV, developed within the framework of the DP-R|EX joint project, maps the central concepts and theoretical approaches in research on racism and right-wing extremism. The compilation is based on a systematic evaluation of the relevant national and international empirical research literature. The CV equally takes into account the different thematic (racism, right-wing extremism, discrimination) as well as methodological (qualitative research, standardised surveys, data from social media and messaging services) research strands.

Version: 1.6.0 [Download](#)

DE **EN**



Aggregation Method (*AggregationMethod*)

Identifies the type of aggregation used to combine related categories, usually within a common branch of a hierarchy, to provide information at a broader level than the level at which detailed observations are taken. (From: The OECD Glossary of Statistical Terms)

Version: 1.1.2 [Download](#)

EN NO



Analysis Unit (*AnalysisUnit*)

Describes the entity being analyzed in the study or variable. This vocabulary can also be used to describe the unit of observation, which is the unit being observed, or from which data are collected. The unit of observation can be the same as, or different from the unit of analysis.

Version: 2. [Send feedback](#)

VIR

Comparing SSH vocabularies and their applications in different systems

TRIPLE Event on Vocabularies, Berlin 2023-03-28

Nina C. Rastinger, Massimiliano Carloni, Klaus Illmayer, Matej Ďurčo

<https://doi.org/10.5281/zenodo.7849428>

Triple Vocabulary: an SSH multilingual vocabulary based in LCSH

URI: <http://semantics.gr/authorities/vocabularies/SSH-LCSH>

[RDF/XML](#)
[JSON-LD](#)
[N-triples](#)
[CSV \(Triple\)](#)

The Triple Vocabulary is a multilingual and hierarchical set of SSH-related concepts. It is a subset of LCSH (Library of Congress Subject Headings) that cover popular SSH aspects and is enhanced with labels in Greek, French, Polish, German, Italian, Portuguese, Spanish and Croatian. The vocabulary is used for the automatic annotation of the publications hosted in the GoTriple platform (<https://www.gotriple.eu/>).
doi: 10.12681/semantics.gr/SSH-LCSH

Category	Concepts
Semantic class	skos:Concept
Provider	Triple Project Consortium
Creator	Triple Project Consortium
Default language	
License	Attribution (CC BY 4.0)



SSHOC

Social Sciences *plural noun* • the study of society and the way people live; the subjects connected with this, for example history, economics, etc.

Humanities *plural noun* • literature, language, history, philosophy, and other subjects that are not a science, or the study of these subjects

Open Science *noun* • the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional.

Cloud (Computing) *noun* • the cloud [S] a computer network where files and programs can be stored, especially the internet.

Data *noun* • information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.

Research *noun* • a detailed study of a subject, especially in order to discover (new) information or reach a (new) understanding.

SSH Open Marketplace *noun* • the only fully-integrated discovery portal which pools and harmonises the tools and services useful for the SSH research communities, offering a high quality and contextualised answer at every step of the research data

Library *noun* • a building, room, or organization that has a collection, especially of books, for people to read or borrow.

EOSC *noun* • an initiative aiming to build the European Open Science Cloud, a trusted and open distributed system for the scientific community, providing seamless access to data and interoperable services that address the whole research data cycle, from discovery and mining to storage, management, analysis and reuse across borders and scientific disciplines.

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

Heritage *noun* • features belonging to the culture of a particular society, such as traditions, languages, or buildings, that were created in the past and still have historical importance

Cloud (Computing) *noun* • the cloud [S] a computer network where files and programs can be stored, especially the internet.

Data *noun* • information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.

ESFRI *noun* • the European Strategy Forum on Research Infrastructures, is a strategic initiative to develop the scientific infrastructure and to strengthen international outreach

FAIR Data Principles • a set of guiding principles of data findable, accessible, interoperable and reusable.

Metadata *noun* • information that is given to describe or identify other information

Human-centered artificial intelligence • the study of how to design systems, etc. that are designed to work in ways that people can understand and learn.

Open Science *noun* • the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional.

Research *noun* • a detailed study of a subject, especially in order to discover (new) information or reach a (new) understanding.

Humanities *noun* •

SSHOC

Workshop Notes

CLARIN



SSHOC

SSHOC Considerations for the Vocabulary Platforms

SSHOC Workshop

06 November 2020

09:00 to 13:15

Online



Availability

What is surveycodings.org?

Surveycodings.org is a free service for survey projects, data archives and researchers measuring and coding socio-economic background variables and more general social science classifications. It provides questionnaires, data collection tools, coding frames, and classifications based on standard statistical classifications for a large number of countries and languages.

SurveyCodings covers the following individual and socio-economic variables: industry, occupation, educational attainment, field of education, religious denominations, social networks, cost of living, and region. Provided materials can be used to collect harmonised and cross-nationally comparable data as well as for post-hoc harmonisation of already collected data (e.g. coding of open-ended questions).



Search tree

- ▶ Kmetijstvo, narava, živali, okolje
- ▶ Skrbstvo, otroci, blaginja, socialno delo
- ▶ Avtomobili, mehaniki, tehniki, inženirji
- ▶ Čiščenje, gospodinjstvo, smeti, odpadki
- ▶ Uradniki, tajniki, pošta, telefon
- ▼ **Komerciala, trgovina, nakup in prodaja**
 - ▶ Pomožne storitve (notranje)
 - ▶ Nadzornik v trgovinah
 - ▶ Pomočnik trgovinske prodaje
 - ▶ Lastnik prodajalne
 - ▶ Najemanje, nastanitev
 - ▶ Ulični prodajalec
 - ▶ Supermarket, veleblagovnica
 - ▶ Popravila
 - ▶ Prodaja
 - ▶ Nakup
 - ▶ Lasje, lepota, kozmetika
 - ▶ Razstava
 - ▶ Klicni center, informacije
- ▶ Gradbeništvo, montaža, nastanitev
- ▶ Izobraževanje, znanstveno raziskovanje, usposabljanje
- ▼ **Finance, bančništvo, zavarovanje**
- ▶ Prehrabena proizvodnja
- ▶ Straža, vojska, policija
- ▶ Zdravstvena oskrba, reševalci, laboratorij
- ▶ Gostoljubnost, turizem, prosti čas, šport
- ▶ Kadrovsko upravljanje, posredovanje dela, organizacija
- ▶ Industrijska pridelava, proizvodnja, kovina



D3.9 Report on Ontology and Vocabulary Collection and Publication

Francesca Frontini¹ ; Federica Gamba¹; Monica Monachini¹ ; Daan Broeder² ;
Kea Tijdens³ ; Irena Vipavc Brvar⁴ 

Multilingual Vocabularies and Ontologies

- Multilingual occupation ontology

 - Description of the ontology

 - Use case: Slovenia

 - Gendered occupational titles

- Multilingual terminology: case study on Data Stewardship

 - Corpus creation

 - Automatic term extraction

 - External validation

 - Definitions, linking and translation

SKOSifying Results

- Implementing the SKOS mapping

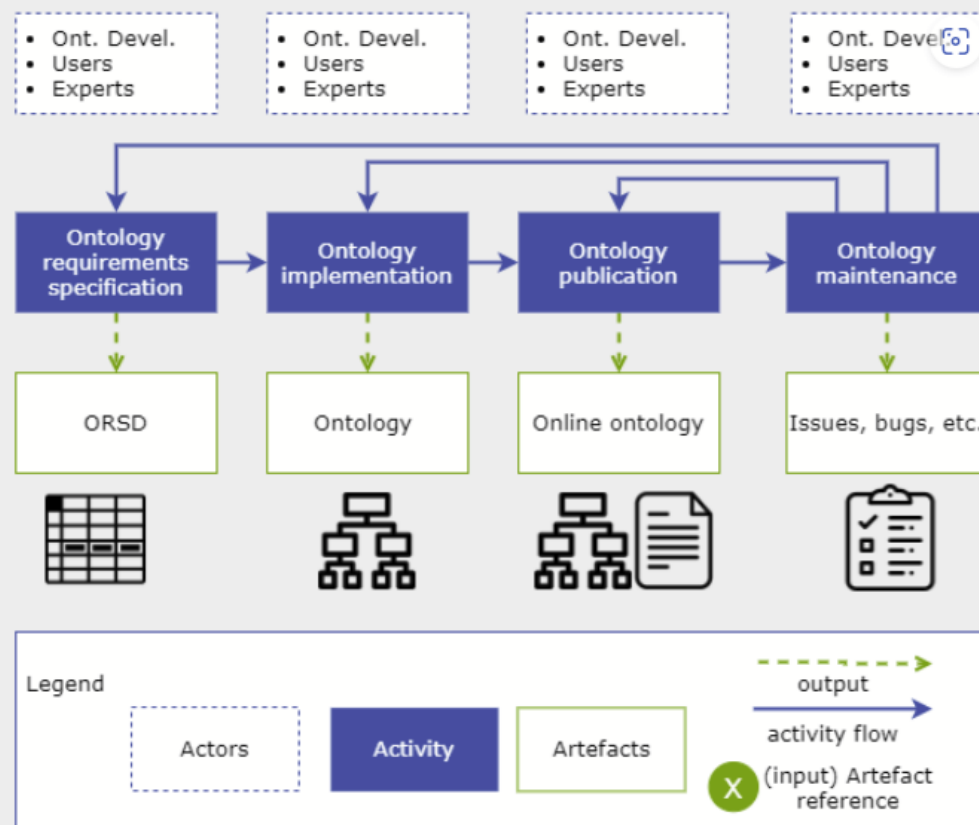
- SKOS Description of the resources

- Language and local specificities of the SKOS-ifying process

Linked Open Terms

Industrial version

LOT (Linked Open Terms) Methodology is an industrial method for developing ontologies and vocabularies. This website will provide support for following the LOT methodology in your developments.



SSHOC Multilingual Data Stewardship Terminology

Content language

English ▼

✕

Search

Alphabetical

Hierarchy

A B C D E F H I L M N O

P R S T U V

access data → data access
accessibility → data accessibility
 accessible data
 administrative data
aggregate data → data aggregation
 anonymisation
 anonymised data
anonymization → anonymisation
anonymized data → anonymised data
 auxiliary data
availability of (research) data → data
 availability
 available data

Vocabulary information

TITLE

SSHOC Multilingual Data Stewardship Terminology

DESCRIPTION

The SSHOC Multilingual Data Stewardship Terminology is a multilingual terminology that collects terms specific to the domain of Data Stewardship, as well as their definitions. A list of domain-specific terms was automatically extracted from a corpus pertaining to the domain of Data Stewardship and Curation, validated by domain experts, assigned a definition, and linked to other existing terminologies (Loterre Open Science Thesaurus, terms4FAIRskills, Linked Open Vocabularies, ISO terms and definitions). Each term-definition pair was then automatically translated into multiple languages (Dutch, French, German, Greek, Italian, Slovenian) by employing Deep-L. The Multilingual Data Stewardship Terminology thus consists of 211 concepts available in Dutch, French, German, Greek, Italian, Slovenian. This resource was created within the frame of the SSHOC (Social

[VIR](#)

RAZPRAVA

Zahvala kolegoma Mihi Pečetu in Mihi Seručniku z ZRC
SAZU za pomoč pri pripravi prosojnic, in vroči debati



Univerza v Ljubljani

