



Triple

Publishing SSHOC Multilingual Terminologies

Cesare Concordia, ISTI - CNR

Use of vocabularies for metadata curation and quality assessment in
Social Sciences and Humanities | **TRIPLE** | 26/3/23 | Berlin



Objectives of the SSHOC activities on Multilingual Terminologies

- Test state of the art Terminology Extraction and Machine Translation
- **Use state of the art technologies to speed up and optimise the creation and publishing of multilingual vocabularies**
- Produce interoperable multilingual resources
- Make them available via the CLARIN ERIC, CLARIN-IT and SSHOC channels



SSHOC Multilingual Terminologies: case studies

- Case study 1: translating CLARIN Core metadata
- Case study 2: creating and translating a terminology for Data Stewardship



Case study 1- Multilingual Metadata

- Translating CLARIN Core metadata
 - 232 approved metadata concepts from the CLARIN Concept Registry
 - Dutch, French, Greek, Italian
- MT services used
 - Google Translate, DeepL
 - CUNI, Reverso
- Evaluation method:
 - Assessment of translations by experts

The screenshot shows the CLARIN Concept Registry Browser interface. At the top, there is a search bar with the placeholder "Please type one or more space separated search terms". Below the search bar are sections for "Search terms mode" (set to "Or") and "Search terms matching" (set to "Whole word"). On the left, there are "Search field filters" for "Labels", "Definition", and "Default documentation fields". Under "Facet filters", there are options for "Status" (Approved: 236, Candidate: 2847, Expired: 80, Any selected), and "Concept Schemes" (Dialogue Acts: 1, Language Codes: 4). The main results table lists 25 concepts found out of 3163. The columns in the table are "Label", "Definition", and "Status". The first result, "winien", is detailed in the definition column: "Winien [winien] contains the flexemes winien 'should', powinien 'should' and rad 'eager, pleased', inflecting for number and gender, with only analytical past tense and conditional forms. Grammatical categories of winien [winien]: number (<http://www.iscat.org/datcat/DC-2709>), gender (<http://www.iscat.org/datcat/DC-2728>), aspect. (source: Przepiórkowski, A. (2004). The IPI PAN Corpus: Preliminary version. Institute of Computer Science, Polish Academy of Sciences, Warsaw.)". Other results include "withdrawn", "within-subjects design", "without declension", "wizard-of-oz", "vocalic", and "wordgroup".

Case study 1- Validation

	UFAL			DEEP-L (https://www.deepl.com/en/tr)			Google translate (https://translate.google.com)			REVERSO			List better system	
term	definition	sous term translati	te definition translation (UFAL)	def	definition val	ter definition tra	def	term translation	ter definition i de	term translatio	terr definition t definitio	term (U,D,G, R)	definition (U,D,G, R)	
wizard-of-oz	A research experiment (soi magicien de N	Expérience de recherche dans laquel Y	magicien d'oz N	Une expérienc Y	magicien d'oz	N	Une expérie Y	magicien de l'oz N	Une expérier Y	//	D,G,R			
writing systems	The visual representation (soi systèmes d'é Y	La représentation visuelle de la lang Y	systèmes d'écri Y	La représentat Y	systèmes d'écriture Y	Y	La représent Y	systèmes d'écritu Y	La représent Y	D,G,R	D,G,R			
persistent identifi	Specification of a persis (soi identifiant pr M	Spécification d'un identificateur pers M	identifiant pers M	Spécification c Y	identifiant persistant M	M	Spécificatio M	identifiant persis M	Spécification Y	//	D,R			
public	The access to the comm (soi public	Y L'accès à l'événement de communica Y	public	Y L'accès à l'évé Y	publique	Y	L'accès à l'év Y	public	Y L'accès à l'év Y	D,G,R	D,G,R			
address	The address of an organ (soi adresse	Y L'adresse d'une organisation qui a pa Y	adresse	Y L'adresse d'un Y	adresse	Y	L'adresse d'i Y	adresse	Y Adresse d'une Y	D,G,R	D,G,R			
age	The number of years th (soi âge	Y Le nombre d'années de vie de quelqu Y	âge	Y Le nombre d'a Y	âge	Y	Le nombre c Y	âge	Y Le nombre d Y	D,G,R	D,G,R			
elicited	Investigator asks speake (soi élimé	N L'enquêteur demande au ou aux loc Y	a suscité	M L'enquêteur di Y	suscité	M	L'enquêteur Y	obtenue	M L'enquêteur i Y	//	D,G,R			
email	The email address of a p (soi courriel	M L'adresse électronique d'une person Y	email	Y L'adresse élect Y	email	Y	L'adresse e-i Y	courriel	M L'adresse col Y	D,G	D,G,R			
event structure	Indicates the structure (soi structure évé Y	Indique la structure de l'événement (Y	structure des é M	Indique la stru Y	structure de l'événe M	Indique la st Y	structure d'événe M	Indique la sti Y	//	D,G,R				
execution locatio	Identification of the loc (soi lieu d'exécut Y	Identification de l'endroit où l'outil o Y	lieu d'exécutio Y	Identification (Y	lieu d'exécution	Y	Identificatio Y	emplacement d'e M	Identificatio Y	D,G	D,G,R			
experimental sett	A transmission of the cc (soi cadre expéri Y	Une transmission du contenu se dérc Y	milieu expérimental M	Une transmiss Y	cadre expérimental Y	Y	Une transmi Y	cadre expérimen Y	Transmission Y	G,R	D,G,R			
face to face	The transmission of the (soi face à face	Y La transmission du message assure u Y	face à face	Y La transmissio Y	face à face	Y	La transmiss Y	face à face	Y La transmissi Y	D,G,R	D,G,R			
false	Contrary to what is true (soi faux	Y Contrairement à ce qui est vrai, error Y	faux	Y Contraire à ce Y	faux	Y	Contraireme Y	faux	Y Contraireme Y	D,G,R	D,G,R			
family	The access to the comm (soi famille	Y L'accès à l'événement de communica Y	famille	Y L'accès à l'évé Y	famille	Y	L'accès à l'év Y	famille	Y L'accès à l'év Y	D,G,R	D,G,R			
fax number	The Fax number of a pe (soi numéro de t M	Le numéro de télécopieur d'une pers M	numéro de fax Y	Le numéro de Y	numéro de fax	Y	Le numéro c Y	numéro de télécc M	Numéro de t M	D,G	D,G			
monologue	Communication event v (soi monologue	Y Événement de communication avec l Y	monologue	Y Événement de Y	monologue	Y	Événement i Y	monologue	Y Événement c Y	D,G,R	D,G,R			
morphology	The study of the structu (soi morphologie Y	L'étude de la structure et de l'élector N	morphologie	Y L'étude de la s Y	morphologie	Y	L'étude de l; Y	morphologie	Y L'étude de la Y	D,G,R	D,G,R			
planned	The speaker prepares in (soi prévu	N L'intervenant prépare en détail la str Y	prévu	N L'orateur prép Y	prévue	N	L'orateur pri Y	prévue	N L'orateur pré Y	//	D,G,R			
planning type	Indicates in how far the (soi type de plan Y	Indique dans quelle mesure le consu Y	type de planific Y	Indique dans c Y	type de planification Y	Indique dan Y	type de planificat Y	Indique dans Y	D,G,R	D,G,R				

5

Discover
Connect
Collaborate



Case study 1- the SSHOC Multilingual Metadata

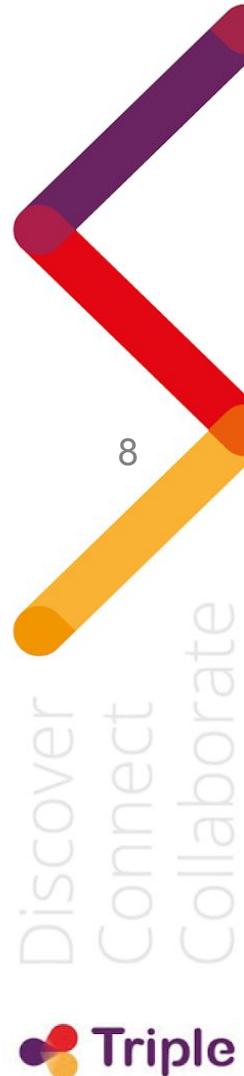
English				Dutch		French		Greek		Italian	
term	definition	source	URI	term	definition	term	definition	term	definition	term	definition
wizard-of-oz	A research experiment (source: CLARIN)		http://hdl.handle.net/11459/0	tovenaar van Oz	Een onderzoeks	expérience du N	Une expérience	Τεχνική wizard	'Eva ερευνητι	esperimento de	Un esperimento
writing systems	The visual representations (source: CLARIN)		http://hdl.handle.net/11459/0	schrijfsystemen	De visuele weer	systèmes d'écrit	La représentatio	συστήματα γρ	Η οπτική αναγ	sistemi di scritt	La rappresentazi
persistent identi	Specification of a persi (source: CLARIN)		http://hdl.handle.net/11459/0	permanente ide	Specificatie van	identifiant péren	Spécification d'u	μόνιμο αναγν	Προσδιορισμό	identificatore pe	Specificazione di
public	The access to the com (source: CLARIN)		http://hdl.handle.net/11459/0	openbaar	De toegang tot h	public	L'accès à l'événe	δημόσια	Η πρόσβαση στο	il pubblico	L'accesso all'eve
address	The address of an orga (source: PWittenburg)		http://hdl.handle.net/11459/0	adres	Het adres van ee	adresse	L'adresse d'une	διεύθυνση	Η διεύθυνση ε	indirizzo	L'indirizzo di un'
age	The number of years t (source: www.macmi		http://hdl.handle.net/11459/0	leeftijd	Het aantal jaren	âge	Le nombre d'ani	ηλικία	Ο αριθμός των età		Il numero di ann
elicited	Investigator asks speak (source: CLARIN)		http://hdl.handle.net/11459/0	uitgelokt	Onderzoeker vra	élite	L'enquêteur der	προκλητός	Ο ερευνητής elicitato		Lo sperimentato
email	The email address of a (source: CLARIN)		http://hdl.handle.net/11459/0	e-mail	Het e-mailadres	email	L'adresse électri	email	Η διεύθυνση τ email		L'indirizzo e-mail
event structure	Indicates the structure (source: CLARIN)		http://hdl.handle.net/11459/0	evenement stru	Aanduiding van	structure événe	Indique la struct	δομή εκδήλω	Υποδεικνύει τ	struttura dell'ev	Indica la struttur
execution locati	Identification of the loc (source: CLARIN)		http://hdl.handle.net/11459/0	uitvoeringslocat	Identificatie van	lieu d'exécution	Identification de	τοποθεσία εκ	Προσδιορισμό	luogo di esecuzi	Identificazione d
experimental se	A transmission of the c (source: CLARIN)		http://hdl.handle.net/11459/0	experimentele s	Een transmissie	cadre expérime	Une transmissio	Πειραματικό	Μετάδοση το	impostazione sp	Una trasmissione
face to face	The transmission of the (source: CLARIN)		http://hdl.handle.net/11459/0	face-to-face	De transmissie	v face à face	La transmission	πρόσωπο με π	Η μετάδοση τι	faccia a faccia	La trasmissione

Case study 2- Corpus

- Creating and translating a terminology for Data Stewardship
 - 211 domain-specific concepts were automatically extracted from a corpus of Data Curation and Stewardship (English)
 - Translated into Dutch, French, German, Greek, Italian, Slovenian
- Terminology extraction methods used:
 - TermoStat (Drouin, 2003) statistical and linguistic
 - TBXTools (Oliver and Vázquez, 2015)
- MT:
 - DeepL

Case study 2- Corpus

- 70 open access documents
 - standards and recommendations for Data Stewardship and Curation
 - deliverables
 - technical documents
- Source: RDA and OpenAIRE
- All documents included are in English, and they amount to a total of 746,084 tokens

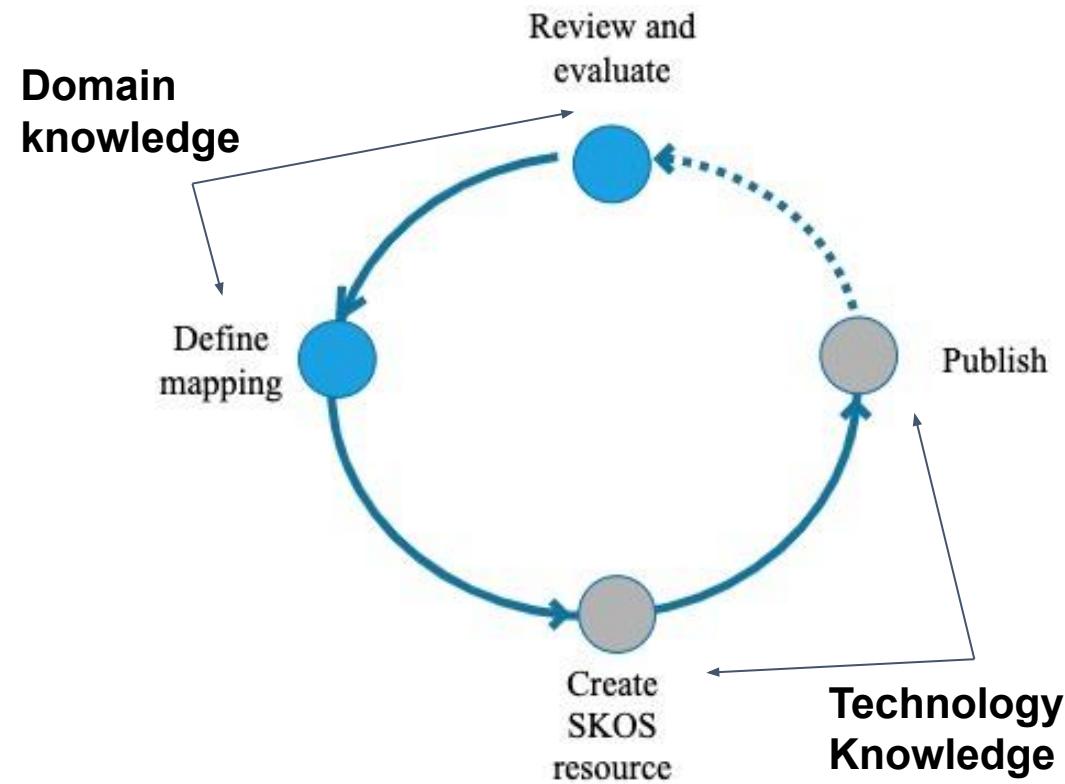


Case study 2- SSHOC Multilingual Data Stewardship Terminology

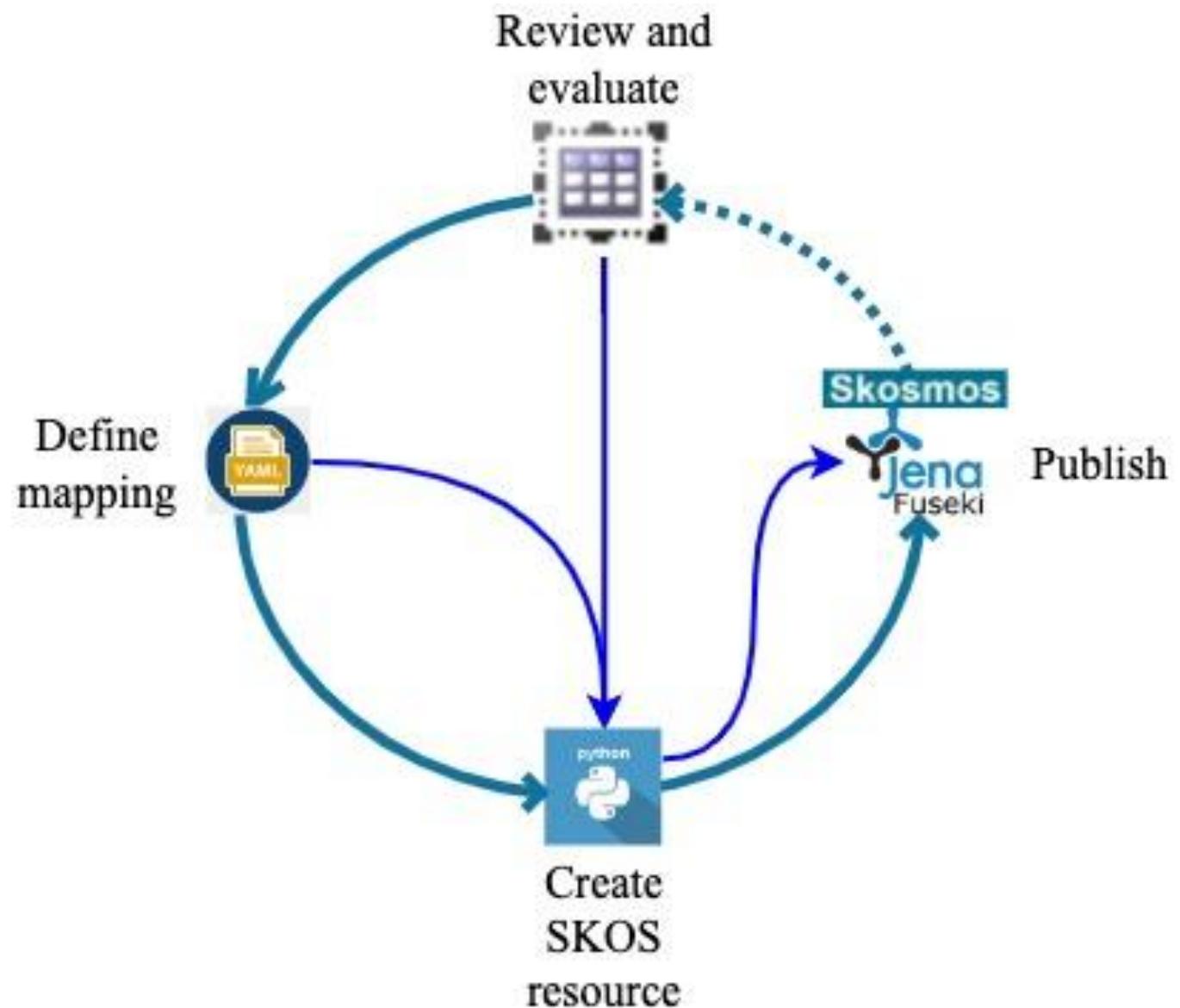
Concept ID		English	Source of definition	Dutch	French	German	Greek	Italian	Slovenian	Loterre Open Science Thesaurus	Terms4FAIRs
accessible_data_1	PrefLabel	accessible data	European Com	toegankelijke gegevens	données accessibles	zugängliche Daten	προσβάσιμα δεδομένα	dati accessibili	dostopni podatki	http://data.loterre.fr/ark:/13030/20190101-0001	9
	Definition	Data that conform to the		Gegevens die de standaard volgen	Données conformes à la norme	Daten, die den Standard einhalten	Δεδομένα που αποτελούνται από δεδομένα που είναι σύμφωνα με την πρότυπη	Dati conformi al standard	Podatki, ki ustrezajo standardu		
administrative_data_2	PrefLabel	administrative data	European Com	administratieve gegevens	données administratives	Verwaltungsdaten	διοικητικά δεδομένα	dati amministrativi	uradni podatki	http://data.loterre.fr/ark:/13030/20190101-0002	9
	Definition	Administrative data refer to data that are created, maintained and used by governments and other public authorities.		Administratieve gegevens verwijzen naar gegevens die worden gecreëerd, gehandhaafd en gebruikt door overheden en andere publieke instanties.	Les données administratives sont celles qui sont créées, maintenues et utilisées par les gouvernements et autres autorités publiques.	Verwaltungsdaten sind jene, die von Regierungen und anderen öffentlichen Ämtern erstellt, gewartet und verwendet werden.	Τα διοικητικά δεδομένα είναι αυτά που δημιουργούνται, διατηρούνται και χρησιμοποιούνται από κυβερνήσεις και άλλες δημόσιες αρχές.	I dati amministrativi sono quelli che vengono creati, mantenuti e utilizzati da governi e altre autorità pubbliche.	Uradni podatki se referirajo na podatke, ki jih ustvarjajo, vzdržujejo in uporabljajo vlade in druge javne uradne institucije.		
anonymisation_3	PrefLabel	anonymisation	European Com	anonymisering	anonymisation	Verwaltungsdaten	Τα διοικητικά δεδομένα	I dati amministrativi	Uradni podatki se referirajo na podatke, ki jih ustvarjajo, vzdržujejo in uporabljajo vlade in druge javne uradne institucije.	http://data.loterre.fr/ark:/13030/20190101-0003	9
	AltLabel	anonymization									
anonymised_data_4	Definition	Anonymisation is the process of removing personally identifiable information from datasets.	Anonymisering is de proces van het verwijderen van persoonlijk identificerend informatie uit datasets.	L'anonymisering	L'anonymisation	Unter Anonymisierung	Η ανωνυμοποίηση	L'anonymizzazione	Anonimizacija je postupak uključivanja	http://data.loterre.fr/ark:/13030/20190101-0004	9
	PrefLabel	anonymised data		geanonimiseerde gegevens	anonimiseerde données	anonimisierte Daten	ανωνυμοποιημένα δεδομένα	anonimizata	anonimizirani podatki		
auxiliary_data_5	AltLabel	anonymized data	ISO 5127:2017								
	Definition	Personal data modified or removed according to ISO 5127:2017.		Persoonsgegevens die zijn bewerkt of verwijderd volgens ISO 5127:2017.	Données personnelles modifiées ou supprimées conformément à l'ISO 5127:2017.	Personenbezoeg gegevens die zijn bewerkt of verwijderd volgens ISO 5127:2017.	Δεδομένα προσώπου που έχουν επεξεργαστεί ή έχουν απομονωθεί σύμφωνα με την ISO 5127:2017.	Dati personali modificati o rimossi secondo l'ISO 5127:2017.	Osebni podatki, spremenjeni ali odstranjeni po navodilih ISO 5127:2017.	http://data.loterre.fr/ark:/13030/20190101-0005	9
	PrefLabel	auxiliary data		hulpgegevens	données auxiliaires	Hilfsdaten	βοηθητικά δεδομένα	ausiliarie dati	spremljajoči podatki		
	Definition	The term auxiliary data refers to Gotz, A., Perrin De term hulpgegevens. Le terme "données auxiliaires" est le terme "dati ausiliari". Der Begriff Hilfsdaten. O όρος βοηθητικά δεδομένα. Il termine dati ausiliari. Izraz spremnjajoči podatki.									

Publishing the vocabularies: workflow

- Review the source vocabulary, individuate: entities, classes, properties, individuals...
- Define mapping rules:
Source format → SKOS
- Create the SKOS resource
- Publish the SKOS resource



Publishing the vocabularies: technologies



11

Discover
Connect
Collaborate



Case study 1 - evaluate and individuate entities

English				Dutch		French		Greek		Italian	
term	definition	source	URI	term	definition	term	definition	term	definition	term	definition
wizard-of-oz	A research experiment (source: CLARIN)		http://hdl.handle.net/11459/0	tovenaar van Oz	Fen onderzoeks	expérience du N	Une expérience	Τεχνική wizard	'Ένα εσευντι	esperimento de Un esperimento	
writing systems	The visual representation (source: CLARIN)		http://hdl.handle.net/11459/0	schrijfsystemen	De visuele weer	systèmes d'écrit	La représentatio	συστήματα γρ	Η οπτική αναγν	sistemi di scritt	
persistent ident	Specification of a persi (source: CLARIN)		http://hdl.handle.net/11459/0	permanente ide	Specificatie van	identifiant pére	Spécification d'u	μόνιμο αναγν	Προσδιορισμό	identificatore p	
public	The access to the comr (source: CLARIN)		http://hdl.handle.net/11459/0	openbaar	De toegang tot	l public	L'accès à l'événe	δημόσια	Η πρόσβαση σ	pubblico	
address	The address of an orga (source: PWittenburg		http://hdl.handle.net/11459/0	adres	Het adres van ee	adresse	L'adresse d'une	διεύθυνση	Η διεύθυνση έ	indirizzo	
age	The number of years th (source: www.macm		http://hdl.handle.net/11459/0	leeftijd	Het aantal jaren	âge	Le nombre d'an	ηλικία	Ο αριθμός τω	Il numero di ann	
elicited	Investigator asks speak (source: CLARIN)		http://hdl.handle.net/11459/0	uitgelokt	Onderzoeker vr	élicite	L'enquêteur der	προκλητός	Ο ερευνητής	elicitato	
email	The email address of a (source: CLARIN)		http://hdl.handle.net/11459/0	e-mail	Het e-mailadres	email	L'adresse électr	email	Η διεύθυνση τ	L'indirizzo e-mail	
event structure	Indicates the structure (source: CLARIN)		http://hdl.handle.net/11459/0	evenement stru	Aanduiding van	structure événe	Indique la struct	δομή εκδήλω	Υποδεικνύει τ	struttura dell'ev	
execution locati	Identification of the loc (source: CLARIN)		http://hdl.handle.net/11459/0	uitvoeringslocat	Identificatie van	lieu d'exécution	Identification de	τοποθεσία εκ	Προσδιορισμό	luogo di esecuzi	
experimental se	A transmission of the c (source: CLARIN)		http://hdl.handle.net/11459/0	experimentele s	Een transmissie	cadre expérime	Une transmissio	Πειραματικό	Μετάδοση το	Impostazione sp	
face to face	The transmission of th (source: CLARIN)		http://hdl.handle.net/11459/0	face-to-face	De transmissie v	face à face	La transmission	πρόσωπο με π	Η μετάδοση τ	faccia a faccia	

rdf:type skos:Concept

skos:prefLabel

language tag

skos:exactMatch

...

12

Discover
Connect
Collaborate



Case study 2 - evaluate and individuate entities

Concept ID		English	Source of definition	Dutch	French	German	Greek	Italian	Slovenian	Loterre Open Science Thesaurus	Terms4FAIRs
accessible_data_1	PrefLabel	accessible data		toegankelijke gegevens	acces	zugängliche D	προσβάσιμα δ	dati accessibili	dostopni podatki	http://data.loterre.fr/ark:/13013/1	
	Definition	Data that conform to the		Gegevens die	Données confor	Daten, die den	Δεδομένα που	Dati conformi	Podatki, ki ustrejaj		
administrative_data_2	PrefLabel	administrative data		administratieve	données admis	Verwaltungsda	διοικητικά δ	dati amminist	uradni podatki		
	Definition	Administrative data refer European Corr	Administrativi	Les données a	Verwaltungsda	Τα διοικητικά δ	I dati amminis	Uradni podatki se r			
anonymisation_3	PrefLabel	anonymisation		anonimisering	anonymisation	Anonymisierung	Ανωνυμοποίης	Anonymizzazi	anonymizacija		
	AltLabel	anonymization									
	Definition	Anonymisation is the pro Eurito.	(2019).	Anonimisering	L'anonymisatic	Unter Anonym	Η ανωνυμοποί	L'anonymizza	Anonymizacija je pc	http://data.loterre.fr/ark:/13013/1	
anonymised_data_4	PrefLabel	anonymised data		geanonimiseer	données anon	anonymisierte	ανωνυμοποιη	dati anonimiz	anonymizirani poda		
	AltLabel	anonymized data									
	Definition	Personal data modified ir ISO 5127:2017	Persoonsgege	Données pers	Personenbezo	Δεδομένα προ	Dati personali	Osebni podatki, sp			
auxiliary_data_5	PrefLabel	auxiliary data		hulpgegevens	données auxili	Hilfsdaten	βοηθητικά δεδ	dati ausiliari	spremljajoči podat		
	Definition	The term auxiliary data re Gotz, A., Perri	De term hulpge	Le terme "donr	Der Begriff Hilf	O όρος βοηθή	Il termine dati	Izraz spremljajoči p			

rdf:type skos:Concept

skos:prefLabel

language tag

skos:altLabel

skos:exactMatch

...

Define mapping rules

- RDF Mapping Language (RML)
 - based on the W3C standard R2RML
 - may process flat files (json, csv,...)
- YARRRML, text-based representation for declarative RML rules
 - uses the YAML syntax
 - JS parser, Web UI

```
prefixes:  
ex: "http://example.com/"  
schema: "https://schema.org/"  
  
sources:  
books:  
  table: books_table  
  queryFormulation: mysql  
  
mappings:  
BookMapping:  
  sources: books  
  s: ex:$ (book_id)  
  po:  
    - [a, schema:Book]  
    - [schema:name, $(name)]  
    - [schema:isbn, $(isbn)]  
    - [schema:author, $(author)]
```

prefixes:

```
dc: http://purl.org/dc/elements/1.1/
skos: http://www.w3.org/2004/02/skos/core#
sshoccmd: https://vocab.sshopencloud.eu/vocabularies/sshocterm/
xsd: http://www.w3.org/2001/XMLSchema#
```

mappings:**conceptscheme:**

```
sources:
  - [resourcemd.json~jsonpath', '$']
```

s: skos:

```
po:
  - [a, skos:conceptScheme]
  - [dc:title, $(title), en~lang]
  - [dc:description, $(description), en~lang]
  - [dc:version, $(version), xsd:integer]
```

concept:

sources:

- ['metadata.csv~csv']

s: sshoccmd:\$ (ConceptId)

po:

- [a, skos:Concept]
- [skos:exactMatch, \$(URI)]
- [dct:source, \$(source)]
- [skos:preflabel, \$(Englishterm), en~lang]
- [skos:definition, \$(Englishdefinition), en~lang]
- [skos:preflabel, \$(Dutchterm), nl~lang]
- [skos:definition, \$(Dutchdefinition), nl~lang]

...

Create SKOS resource and publish it

- Transformation code implemented in Python, as a Jupyter Notebook
- The SKOS resource is stored in *turtle* format and uploaded on Fuseki
- A local SKOSMOS instance is used by domain experts to evaluate results

Mapping **SSHOC Multilingual Metadata** to **SKOS** resources

This Notebook implements a simple parser used to transform the SSHOC Multilingual Metadata, created in the Task 3.1 of the SSHOC project and published as spreadsheet, into a SKOS resource. The parser reads the spreadsheet and transforms the content following a set of mapping rules defined using **YARRML**, the result is stored in Turtle files, and downloaded in a Fuseki server.

```
In [1]: import pandas as pd
import rdflib
import itertools
import yaml
import datetime
import json
from jsonpath_ng import jsonpath, parse
from rdflib.namespace import DC, DCAT, DCTERMS, OWL, \
    RDF, RDFS, SKOS, \
    XMLNS, XSD, XMLNS
from rdflib import Namespace
from rdflib import URIRef, BNode, Literal
```

The file `config.yaml` contains the external information used in the parsing, including the path of the spreadsheets. Set the correct values before running the Notebook. The file `mappings.yaml` contains the YARRML mapping rules

```
In [2]: try:
    with open("config.yaml", 'r') as stream:
        try:
            conf=yaml.safe_load(stream)
        except yaml.YAMLError as exc:
            print(exc)
    with open("rules.yaml", 'r') as stream:
        try:
            rules=yaml.safe_load(stream)
        except yaml.YAMLError as exc:
            print(exc)
except FileNotFoundError:
    print('Warning config.yaml file not present! Please store it in the same directory as the notebook')
#print (conf)
```

The following functions implement a basic parser that processes the YARRML rules and creates a RDF graph

```
In [3]: def jsonmapper (json_source, rules):
    ccr_y = rdflib.Graph()
    prefixes=rules['prefixes']
    nss={}
    for key in prefixes:
        myns=Namespace(URIRef(prefixes[key]))
        nss[key]=myns
```

Case study 1- SKOS Resource

 **Vocabs** Vocabularies About Editor SPARQL API Help

SSHOC Multilingual Metadata Content language English ▾

Alphabetical	Hierarchy
A B C D E F G H I L M N O P Q R	S T U V W

PREFERRED TERM availability 

DEFINITION A description of the terms of availability of the resource in simple words.

SOURCE CLARIN

IN OTHER LANGUAGES

beschikbaarheid	Dutch
disponibilité	French
διαθέσιμότητα	Greek
disponibilità	Italian

URI https://vocabsshopencloud.eu/vocabularies/sshocmmsshoc_CCR_C-2453_1f0c3ea5-7966-ae11-d3c6-448424d4e6e8 

DOWNLOAD THIS CONCEPT: RDF/XML TURTLE JSON-LD Last modified 12/17/21

EXACTLY MATCHING CONCEPTS availability availability Metadata

```
1 @prefix dc: <http://purl.org/dc/elements/1.1/> .
2 @prefix dct: <http://purl.org/dc/terms> .
3 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
4 @prefix sshoccmd: <http://sshoc.eu/XXX/> .
5 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
6
7 sshoccmd:CCR_C-2452_604f0c0f-a038-746c-d585-268642eaaba7 a skos:Concept ;
8   dct:source "CLARIN" ;
9   skos:definition "Free"@en,
10    "libre"@fr,
11    "libero "@it,
12    "Gratis"@nl ;
13   skos:exactMatch "http://hdl.handle.net/11459/CCR_C-2452_604f0c0f-a038-746c-d585-268642eaaba7" ;
14   skos:preflabel "free"@en,
15    "libre"@fr,
16    "δωρέα"@gr,
17    "libero "@it,
18    "gratis"@nl .
19
20 sshoccmd:CCR_C-2453_1f0c3ea5-7966-ae11-d3c6-448424d4e6e8 a skos:Concept ;
21   dct:source "CLARIN" ;
22   skos:definition "A description of the terms of availability of the resource in simple words."@en,
23    "Une description des conditions de disponibilité de la ressource en mots simples."@fr,
24    "Una descrizione dei termini di disponibilità della risorsa in parole semplici."@it,
25    "Een beschrijving van de voorwaarden voor beschikbaarheid van de bron in eenvoudige woorden."@nl ;
26   skos:exactMatch "http://hdl.handle.net/11459/CCR_C-2453_1f0c3ea5-7966-ae11-d3c6-448424d4e6e8" ;
27   skos:preflabel "availability"@en,
28    "disponibilité"@fr,
29    "διαθέσιμότητα"@gr,
30    "disponibilità"@it,
31    "beschikbaarheid"@nl .
32
33 sshoccmd:CCR_C-2454_e0e88868-2fc6-a87e-cd0d-d8f69c85b41d a skos:Concept ;
34   dct:source "CLARIN" ;
35   skos:definition "The name of the person that can be contacted to get access to the resource or to the tool/service."
```

Discover Connect Collaborate



Case study 2- SKOS Resource

 **Vocabs** [Vocabularies](#) [About](#) [Editor](#) [SPARQL](#) [API](#)

Help | Interface language: English ▾

SSHO Multilingual Data Stewardship Terminology

Alphabetical	Hierarchy												
A B C D E F H I L M N O P R S T U V	PREFERRED TERM												
	anonymisation 												
DEFINITION	Anonymisation is the process of removing any information that could lead to the identification of the data subject.												
ENTRY TERMS	<i>anonymization</i>												
NOTE	ISO/TS 17975:2015(en), 3.1												
SOURCE	Eurito. (2019). D3.1 Design of Data Collection Phase. Zenodo. https://doi.org/10.5281/zenodo.3243500 .												
IN OTHER LANGUAGES	<table> <tbody> <tr> <td>anomimisering</td> <td>Dutch</td> </tr> <tr> <td>anonymisation</td> <td>French</td> </tr> <tr> <td>Anonymisierung</td> <td>German</td> </tr> <tr> <td>ανωνυμοποίηση</td> <td>Greek</td> </tr> <tr> <td>anonimizzazione</td> <td>Italian</td> </tr> <tr> <td>anonymizacija</td> <td>Slovenian</td> </tr> </tbody> </table>	anomimisering	Dutch	anonymisation	French	Anonymisierung	German	ανωνυμοποίηση	Greek	anonimizzazione	Italian	anonymizacija	Slovenian
anomimisering	Dutch												
anonymisation	French												
Anonymisierung	German												
ανωνυμοποίηση	Greek												
anonimizzazione	Italian												
anonymizacija	Slovenian												
URI	https://vocabsshopencloud.eu/vocabularies/sshoterm/anonymisation_3 												
DOWNLOAD THIS CONCEPT:	RDF/XML TURTLE JSON-LD												
EXACTLY MATCHING CONCEPTS	Anonymisation Activity http://data.loftere.fr/ark:/67375/TSO-VNNFL8WB-H												

```

42
43      "dati amministrativi"@it,
44      "administratieve gegevens"@nl,
45      "uradni podatki"@sl ;
46      skos:topConceptOf sshoterm: .
47
48 sshoterm:anonymisation_3 a skos:Concept ;
49   dct:source "Eurito. (2019). D3.1 Design of Data Collection Phase. Zenodo.
https://doi.org/10.5281/zenodo.3243500"^^xsd:string ;
50   skos:altLabel "anonymization"@en ;
51   skos:definition "Unter Anonymisierung ist das Entfernen aller Informationen zu verstehen, die zur Identifizierung des
Datensubjekts führen können."@de,
52   "Η ανωνυμοποίηση είναι η διαδικασία αφαίρεσης κάθε πληροφορίας που θα μπορούσε να σημάνει στην ταυτοποίηση του
υποκειμένου των δεδομένων."@el,
53   "Anonymisation is the process of removing any information that could lead to the identification of the data
subject."@en,
54   "L'anonymisation est le processus qui consiste à supprimer toute information pouvant conduire à l'identification
du sujet concerné (par ex. une personne, une institution, etc.)."@fr,
55   "L'anonymizzazione è il processo di rimozione di qualsiasi informazione che potrebbe portare all'identificazione
dell'interessato."@it,
56   "Anonimisering is het proces waarbij alle informatie die tot identificatie van de betrokkenne kan leiden, wordt
verwijderd."@nl,
57   "Anonimizacija je postopek odstranjevanja vseh informacij, na podlagi katerih bi bilo mogoče identificirati
posameznika, na katerega se nanašajo osebni podatki."@sl ;
58   skos:exactMatch <http://data.loftere.fr/ark:/67375/TSO-VNNFL8WB-H>,
59   <https://w3id.org/GDProv#AnonymisationActivity> ;
60   skos:inScheme sshoterm: ;
61   skos:note "ISO/TS 17975:2015(en), 3.1" ;
62   skos:prefLabel "Anonymisierung"@de,
63   "ανωνυμοποίηση"@el,
64   "anonymisation"@en,
65   "anonymisation"@fr,
66   "anonymizzazione"@it,
67   "anonimisering"@nl,
68   "anonymizacija"@sl ;
69   skos:topConceptOf sshoterm: .
70
71 sshoterm:anonymised_data_4 a skos:Concept ;
72   dct:source "ISO 5127:2017(en), 3.1.10.15"^^xsd:string ;

```

Discover Connect Collaborate



Lessons learned

- For small/medium size terminologies the tabular format is a good choice in terms of evaluation/curation tasks and as input format for transformation process
- The mapping process has a crucial role, YARRRML is an interesting candidate but requires some technical knowledge
 - Create templates
- The use of Notebooks for implementing the transformation process helps to document computational steps and improves reproducibility

Next steps

- Build a complete YARRRML parser in Python
- Use the framework to create the SKOS resource for Pan-Latin Textile Fibres Vocabulary (Lessico panlatino delle fibre tessili), developed within the [Realiter network](#)
- Ongoing: creation of a CLARIN terminology (English and Italian) to improve indexing and search of CLARIN content, to be published on the vocabulary platform.



Thank you!

- Frontini, Francesca; Gamba, Federica; Monachini, Monica and Broeder, Daan, 2021, SSHOC Multilingual Data Stewardship Terminology, ILC-CNR for CLARIN-IT, <http://hdl.handle.net/20.500.11752/ILC-567>
- Frontini, Francesca; Gamba, Federica; Monachini, Monica and Broeder, Daan, 2021, SSHOC Multilingual Metadata, ILC-CNR for CLARIN-IT, <http://hdl.handle.net/20.500.11752/ILC-568>
- Trupiano, Luca and Concordia, Cesare, 2021, SSHOC Data Stewardship terminology and Metadata SKOSifying mapping, ILC-CNR for CLARIN-IT, <http://hdl.handle.net/20.500.11752/ILC-566>
- Code: <https://gitea-s2i2s.isti.cnr.it/concordia/sshoc-skosmapping>