Addressing multilingualism in the GoTriple discovery platform

Suzanne Dumouchel | **CNRS**| NISO+ Conference, February 15th @ 7:30 PM EST
"Bibliodiversity in transition: how open infrastructures support multilingualism (or not)"

GoTriple platform :
The discovery service of OPERAS RI

OPERAS is a Research Infrastructure, with a legal entity (AISBL) dedicated to Open Scholarly Communication in the field of SSH and providing several services for the SSH communities (researchers, librarians, publishers).

The RI is organised in Special Interest Groups (SIGs) which constitute the Assembly of the Commons at the General Assembly. It promotes largely multilingualism especially through its services and its national nodes (that constitute OPERAS Executive Assembly)

Discover Connect Collaborate

Triple

Triple

Transforming Research Through Innovative Practices for Linked Interdisciplinary Exploration

TRIPLE was launched in October 2019.

21 partners in Europe, coordinated by CNRS (France)

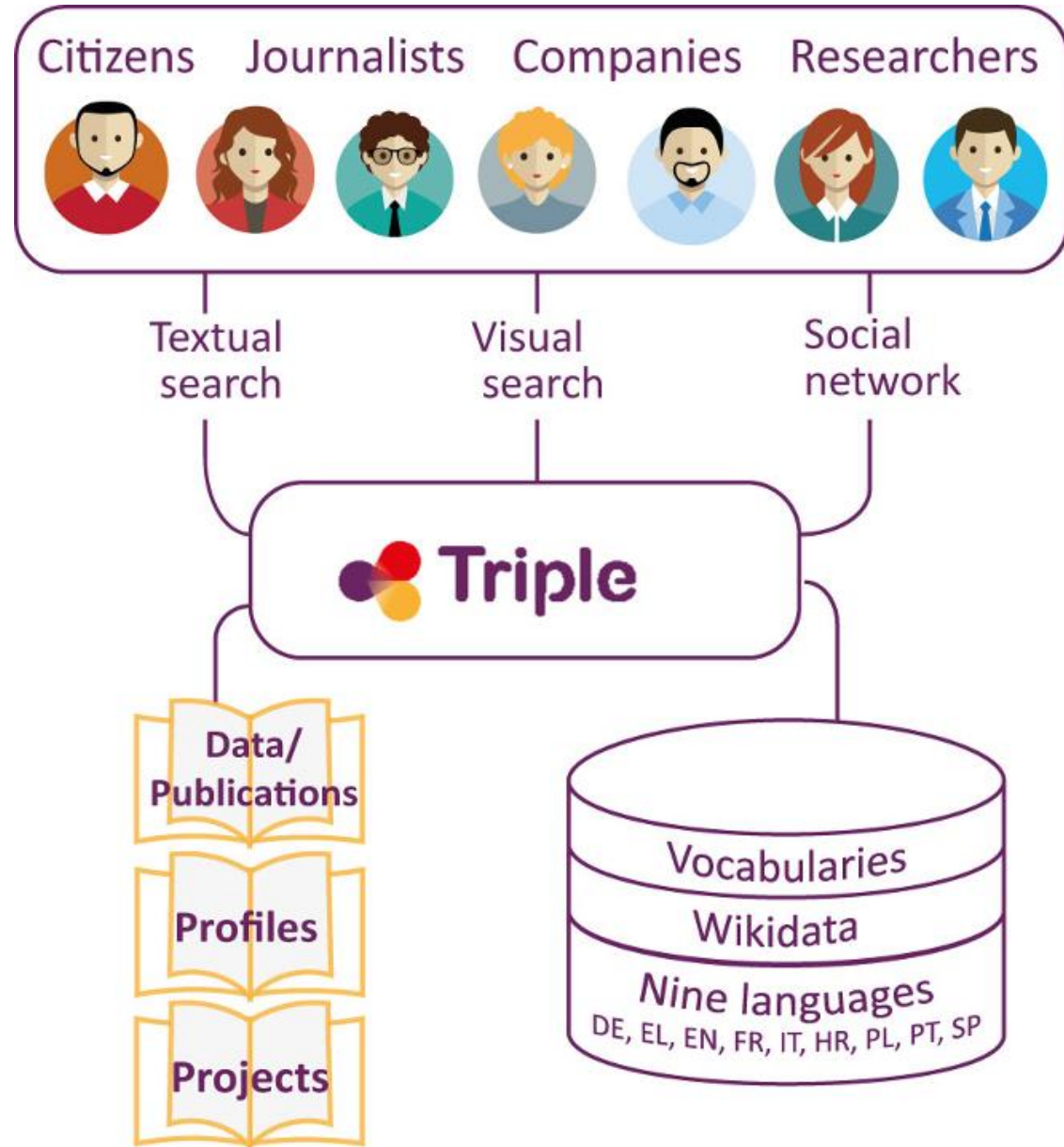**Duration of 42 months and a budget of 5.6M€**

Discover Connect Collaborate

What is TRIPLE

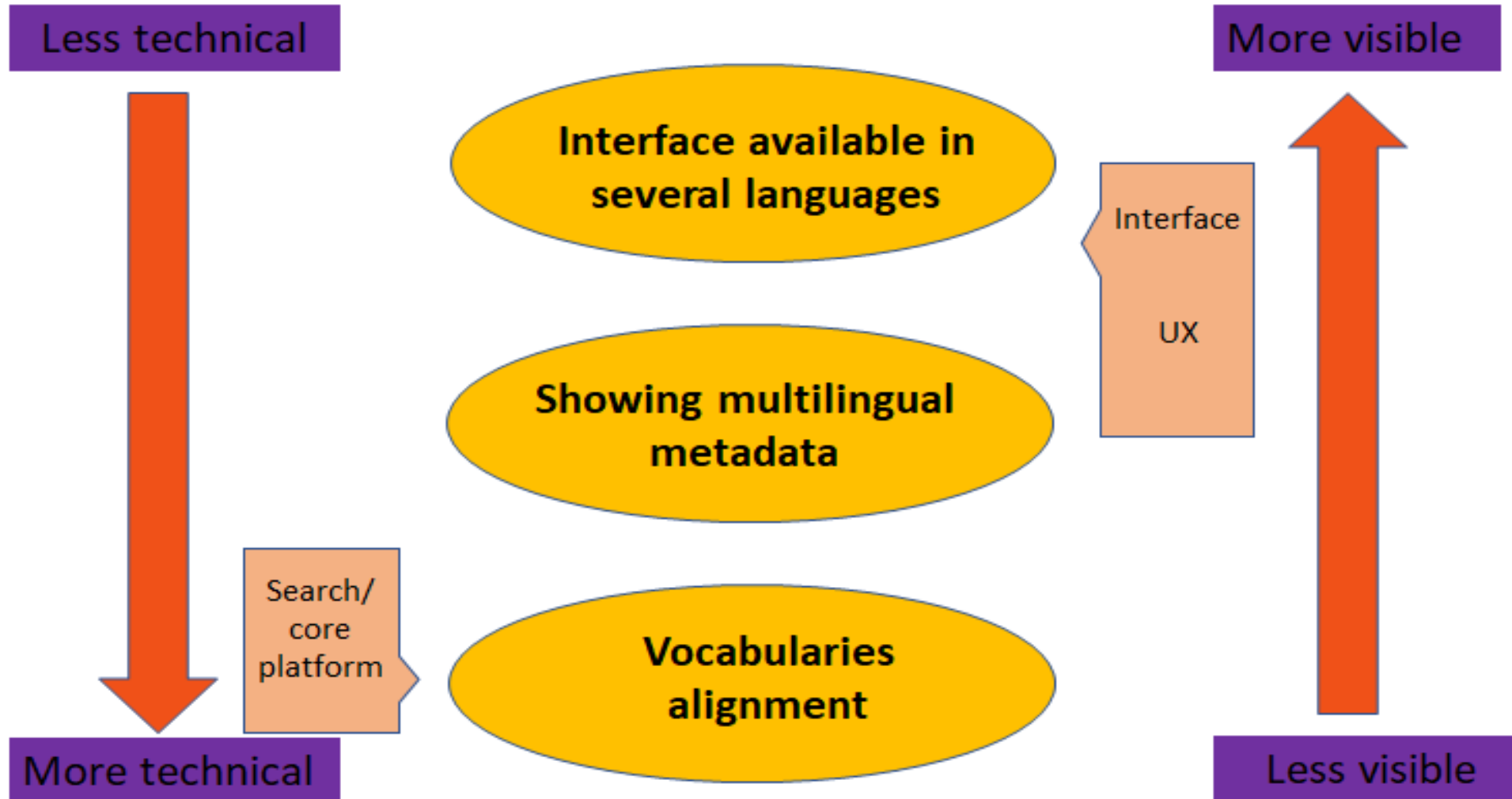TRIPLE platform will promote a **"Community of Practice"**

- foster new collaborations and networking
- encourage interdisciplinarity
- take into account multilingual and multicultural diversity
- facilitate crowdfunding

Multilingualism in GoTRIPLE

Less technical

More visible

Interface available in several languages

Interface

UX

Showing multilingual metadata

Search/ core platform

Vocabularies alignment

More technical

Less visible

Discover Connect Collaborate

Triple

1. The vocabularies in SSH and their alignment

Discover
Connect
Collaborate

Triple

Process for resource's determination:

1. Classification by discipline (from the MORESS categories) and by language

1. Annotation of the resource (tagging item or metadata) in one of the 9 languages => more searchable for audiences that speak any of the other languages

1. Enrichment with links that help in the disambiguation of the term and increases its semantic interoperability.

Discover
Connect
Collaborate

Triple

What is a Controlled Vocabulary?

- A controlled vocabulary is an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. It typically includes preferred and variant terms and has a defined scope or describes a specific domain.

- The purpose of controlled vocabularies is to organize information and to provide terminology to catalog and retrieve information.
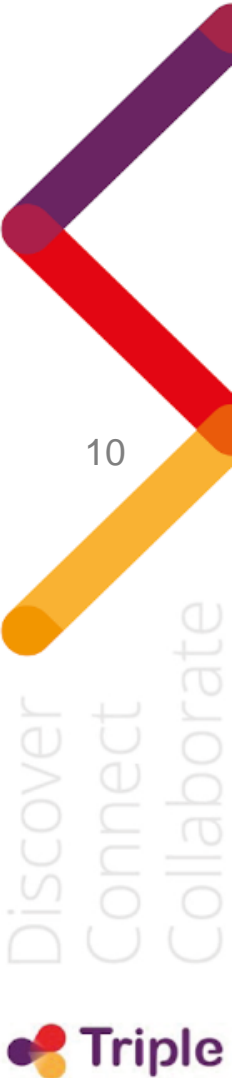
Discover
Connect
Collaborate

Triple

The need for a reliable SSH Vocabulary

- SSH Researchers struggle for better access to resources relevant to their research, publications and data.

- In this effort reliable SSH vocabularies are essential: necessary at the indexing phase because without them catalogers will not consistently use the same term to refer to the same concept, person, place, or thing.

- In the retrieval process, various end users may use different synonyms or more generic terms to refer to a given concept. End users are often not specialists and thus need to be guided because they may not know the correct term.

Discover
Connect
Collaborate

Triple

The GoTriple Vocabulary

To ensure an effective annotation mechanism for publications of all the 9 languages, GoTriple vocabulary must contain a sufficient number of concepts and, at the same time, these concepts must have labels in as many of these languages as possible.

Discover
Connect
Collaborate

Triple

How to build the GoTriple vocabulary

1) Existing classifications and controlled vocabularies have been gathered and compared.
2) From this first analysis, creation of new vocabularies required for the description in new languages.
3) Selection of an existing vocabulary (the **Library of Congress Subject Headings** System) used as a basic resource in English for the construction of the TRIPLE Vocabulary.
4) Automatic translation for the SSH concepts in each of the 9 languages (after testing several automatic translation tools by human validation and checking)
5) Human validation of the translation by experts of each language to confirm the translations or to suggest better ones if needed.

Discover
Connect
Collaborate

Triple

Focus on the selection of concepts:

1) 14 basic concepts were identified   from the Frascati taxonomy under Social Sciences and Humanities.

1) Based on these 37 broad concepts from LCSH were identified.

1) For each of these we used the Linked Data API of the Library of Congress and we retrieved their SKOS representation.

1) For each of these we followed the skos:narrower property and we extracted their children in SKOS.

1) We ended up with **2513** concepts.

Discover
Connect
Collaborate

Triple

Example of spreadsheets

Hierarchy	en	translation_1	validated_1	translation_2	validated_2	Final (read-only)
▶ Anthropology ▶ Ethnology ▶ Folklore	Communication in folklore	Communication dans le folklore	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Ethnic folklore	Folklore ethnique	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Intercultural communication in folklore	Communication interculturelle dans le folklore	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Mass media and folklore	Médias de masse et folklore	☑	Les médias et le folklore	☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Music and folklore	Musique et folklore	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	National socialism and folklore	National-socialisme et folklore	☑	Le national-socialisme et le folklore	☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Public folklore	Folklore public	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Radio in folklore	Radio dans le folklore	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Folklore	Symbolism in folklore	Symbolisme dans le folklore	☑		☐	☑
▶ Anthropology ▶ Ethnology ▶ Forest people	Rain forest people	Les gens de la forêt tropicale	☑	Peuples de la forêt tropicale	☑	☑
▶ Civilization	Aeronautics and civilization	Aéronautique et civilisation	☑		☐	☑
▶ Civilization	Climate and civilization	Climat et civilisation	☑	Le climat et la civilisation	☐	☑
▶ Civilization	Comparative civilization	Civilisation comparée	☑	civilisation comparative	☐	☑
▶ Civilization	Fishes and civilization	Poissons et civilisation	☑	Les poissons et la civilisation	☐	☑
▶ Civilization	Nature and civilization	Nature et civilisation	☑		☐	☑
▶ Civilization	Ocean and civilization	Océan et civilisation	☑		☐	☑
▶ Civilization	Soil and civilization	Sol et civilisation	☑	Le sol et la civilisation	☐	☑
▶ Civilization ▶ Education	After-school programs	Programmes parascolaires	☑	Parascolaire	☐	☑
▶ Civilization ▶ Education	Architecture in education	L'architecture dans l'éducation	☐	Architecture dans l'éducation	☑	☑
▶ Civilization ▶ Education	Archives and education	Archives et éducation	☑	Archives et de l'éducation	☐	☑

Validated so far:

Discover Connect Collaborate

Triple

Current coverage per language

		October 2020		June 2021	
		# of Concepts with labels in language (out of 2565)	language coverage percentage	# of Concepts with labels in language (out of 2565)	language coverage percentage
Greek	el	248	9.67%	2276	88.73%
French	fr	1624	63.31%	2219	86.51%
Polish	pl	336	13.10%	2561	99.84%
German	de	1028	40.08%	2561	99.84%
Italian	it	686	26.74%	2560	99.81%
Portuguese	pt	347	13.53%	2560	99.81%
Spanish	es	406	15.83%	2560	99.81%
Croatian	hr	153	5.96%	2228	86.86%

2. Highlight translated keywords and titles in the metadata document

Discover
Connect
Collaborate

Triple

· Article · English · DOI: oai:doaj.org/article:6d926108c0264303ac8208ac10657f23

Feminicide in Latin America: legal vacuum or deficit in the rule of law?

Authors Cecilia Brunetti

Publication date 2017-12-01

Publication Barcelona Centre for International Affairs (CIDOB)

Licence Licence CC BLA BLA ⓘ

Switch metadata language

- English (title, keywords, abstract)
- French (title, keywords)
- Spanish (keywords)
- Portugese (keywords)
- Italian (title)
- Polish (title, keywords, abstract)
- German (abstract)

Disciplines

Law   Geography

Classical studies

Keywords

Number concept 🅰🇹   Enactments, Legislative 🅰🇹

Representative government and representation 🅰🇹

Latin America 🅰🇹   feminicide 🅰🇹   impunity 🅰🇹

rule of law 🅰🇹   +3

View on OpenAire ↗

Annotate publication ↗

Other actions ⌄

Share this paper   f in t

17

The escalating number of violent female homicides committed by men over the last two decades has compelled many Latin American countries to classify (typify) the crime of gender-based homicide as "feminicide" (or "femicide"). Their objective was to raise awareness and thereby decrease the number of

Discover Connect Collaborate

◆ Triple

Free full text avaiable   Article · English · DOI: oai:doaj.org/article:6d926108c0264303ac8208ac10657f23

Kobietobójstwo w Ameryce Łacińskiej: luka prawna czy deficyt praworządności?

Authors   Cecilia Brunetti

Publication date   2017-12-01

Publication   Barcelona Centre for International Affairs (CIDOB)

Licence   Licence CC BLA BLA ⓘ

Switch metadata language

Polish (title, keywords, abstract) ▼

Disciplines

Law   Geography   Classical studies

Keywords

Pojęcie liczby A🌐   Przepisy ustawowe i wykonawcze A🌐   Rząd reprezentacyjny i reprezentacja A🌐   Ameryka Łacińska A🌐   feminicyd A🌐   bezkarność A🌐   praworządność A🌐   +3

View on OpenAire ↗

Annotate publication ↗

Other actions ⌄

Share this paper   f in 🐦

Abstract

Victims of Femicide in Latin America: Legal and Criminal Justice Responses

Rosnąca liczba brutalnych zabójstw kobiet popełnianych przez mężczyzn w ciągu ostatnich dwóch dekad zmusiła wiele krajów Ameryki Łacińskiej do sklasyfikowania (typizacji) przestępstwa zabójstwa na tle płci jako "kobietobójstwa" (lub "femobójstwa"). Ich celem było podniesienie świadomości, a tym samym zmniejszenie liczby kobietobójstw popełnianych rocznie. Jednakże, mimo że w wielu krajach wskaźnik kobietobójstw spadł, w kolejnych latach ponownie wzrósł. W niniejszym opracowaniu zbudowano zbiór danych na temat wskaźnika kobietobójstw w latach 2000–2014 oraz dokonano analizy czynników wpływających na jego wysokość. Stwierdzono, że kryminalizacja (typizacja) kobietobójstwa nie jest istotna dla przewidywania wskaźnika kobietobójstwa. Zamiast tego, niski poziom rządów prawa i brak reprezentacji kobiet w organach decyzyjnych, takich jak parlamenty krajowe, okazują się być najbardziej znaczącymi czynnikami w wyjaśnianiu zmienności tendencji w zakresie kobietobójstw.

Rosnąca liczba brutalnych zabójstw kobiet popełnianych przez mężczyzn w ciągu ostatnich dwóch dekad zmusiła wiele krajów Ameryki Łacińskiej do sklasyfikowania (typizacji) przestępstwa zabójstwa na

Read more

18

3. the whole gotriple.eu interface

19

Discover Connect Collaborate

Triple

A co-design activity

Not planned at the beginning of the project

Translation of the interface in several languages will be made by the participants of a future hackathon and with the support of OPERAS national nodes.

Discover Connect Collaborate

Triple

TRIPLE project: Reconciling standardisation and diversity

- Nurturing wikidata with SSH concepts in different languages (=> promotes cultural and linguistic specificities)

- Highlighting regional diversity of SSH resources, keywords, concerns, topics

- Methods and processes will be used for adding new languages and can be reused for other initiatives in order to promote multilingualism
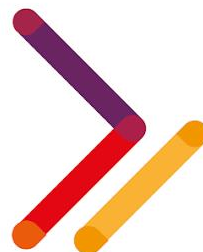
Discover Connect Collaborate

Triple

Join us on www.gotriple.eu !!

project. gotriple.eu

Follow us on:

@TripleEU

TRIPLE Project

Triple

Suzanne Dumouchel

Suzanne.dumouchel@huma-num.fr

The **GoTriple** platform will be the Discovery Service of the OPERAS Research Infrastructure.

OPERAS
open access in the european research
area through scholarly communication