



Mappings in Social Sciences and Humanities: the GoTriple case

"Why Mappings Matter and how to make them FAIR?" Workshop

Luca De Santis - Net7 / TRIPLE project

April 13th 2023

About me *(and why I'm presenting the SSH perspective...)*

R&D director of the Italian SME **Net7 Srl**.

Degree in **Computer Science** at the University of Pisa, Italy

Working experience of ~ 30 years; large experience in research projects

Some of Net7's **SSH-related solutions**

- the Pundit semantic web annotation tool (thepund.it)
- The Muruca platform for Scholarly Digital Editions (muruca.org/en/)

Involved in several **OPERAS** projects (operas-eu.org)

- Technical director of the TRIPLE project => GoTriple.eu
- Chief architect and PM for the implementation of the VERA (vera.operas-eu.org) and Pathfinder platforms (COESO and OPERAS-PLUS projects)



About me *(and why I'm presenting the SSH perspective...)*

R&D director of the Italian SME **Net7 Srl**.

Degree in **Computer Science** at the University of Pisa, Italy

Working experience of ~ 30 years; large experience in research projects

Some of Net7's **SSH-related solutions**

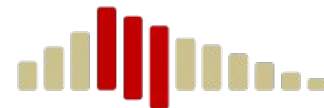
- the Pundit semantic web annotation tool (thepundit.it)
- The Muruca platform for Scholarly Digital Editions (muruca.org/en/)

Involved in several **OPERAS** projects (operas-eu.org)

- Technical director of the TRIPLE project => GoTriple.eu
- Chief architect and PM for the implementation of the VERA (vera.operas-eu.org) and Pathfinder platforms (COESO and OPERAS-PLUS projects)



PUNDIT



MURUCA



About me *(and why I'm presenting the SSH perspective...)*

R&D director of the Italian SME **Net7 Srl**.

Degree in **Computer Science** at the University of Pisa, Italy

Working experience of ~ 30 years; large experience in research projects

Some of Net7's **SSH-related solutions**

- the Pundit semantic web annotation tool (thepundit.it)
- The Muruca platform for Scholarly Digital Editions (muruca.org/en/)

Involved in several **OPERAS** projects (operas-eu.org)

- Technical director of the TRIPLE project => GoTriple.eu
- Chief architect and PM for the implementation of the VERA (vera.operas-eu.org) and Pathfinder platforms (COESO and OPERAS-PLUS projects)



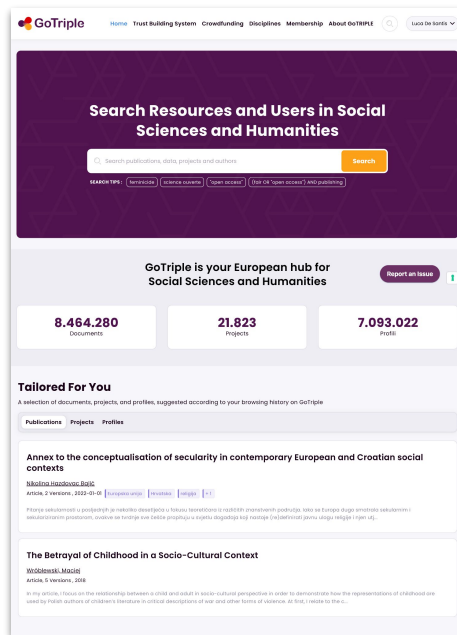
About GoTriple.eu

Multilingual discovery platform for the SSH domain

Search engine to discover Publications, Projects, Authors and Researchers Profiles

A **configurable harvesting and processing pipeline** to import and process publications and projects metadata

- ~8.5 million documents, 21.800 projects, 7 million authors... and counting
- Documents automatically harvested by small repositories and large aggregators alike
- Almost 1.300 OAI-PMH endpoints managed, dump imports from OpenAIRE, Isidore, CORDIS



FAIR in GoTriple

- **Findable**

- Searchable
- Detailed metadata description with the TRIPLE Data Model
- Metadata enrichment

- **Accessible**

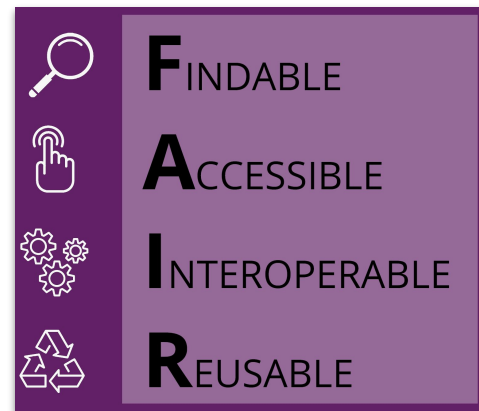
- Standard access protocols

- **Interoperable**

- A classification system for SSH based on 27 disciplines
- The TRIPLE Vocabulary of ~ 3,300 concepts in 12 languages
- The TRIPLE Ontology; linked data

- **Reusable**

- Explanation about data provenance
- Data license clearly expressed
- Free APIs; detailed documentations



FAIR should be considered as a journey...

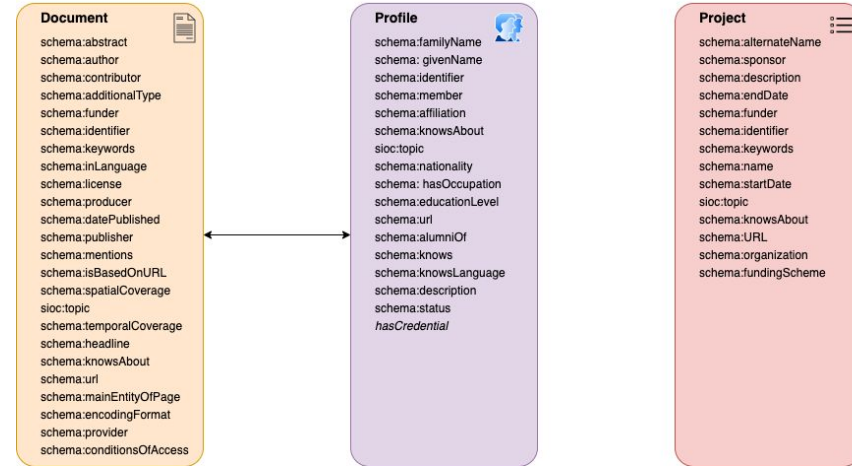


Data Model

Defined through **standard and generic ontologies**: schema.org, [SIOC](http://sioc.org)

Data acquisition:

- **Documents**: Publications harvested from sources
- **Profiles**: *Authors* automatically detected from publications, with a disambiguation algorithm. *Registered users* can “claim” their documents”
- **Projects**: harvested from CORDIS or imported from other OPERAS initiatives (Crowdfunding, VERA)

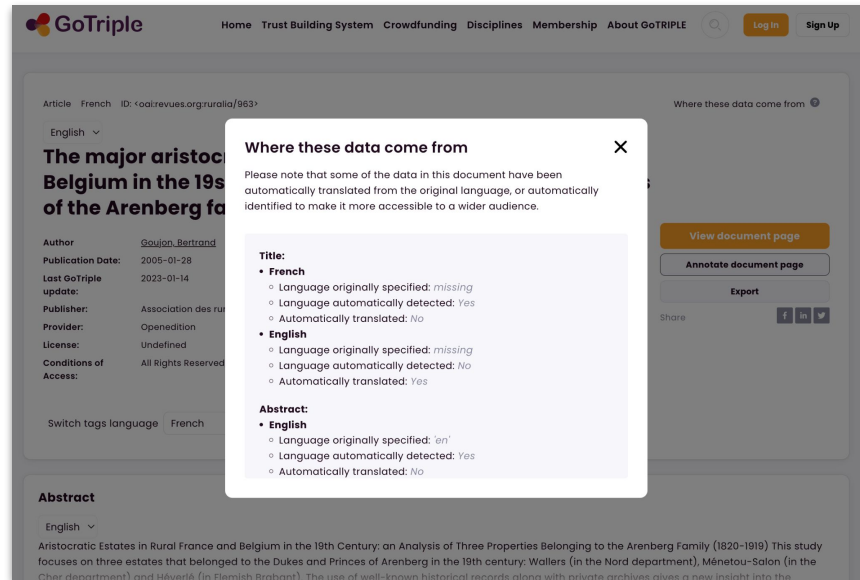


Data Enrichment

Multiple **new metadata** created by:

- **normalisation** rules
 - *controlled vocabularies* for languages, document types, licences, conditions of access
- ad-hoc **algorithms**
 - *clustering* to group duplicate documents
 - authors disambiguation
- **machine learning and NLP** techniques
 - classification and automatic annotation of documents and projects

Detailed presentation of the origin of data



The screenshot shows the GoTriple interface for a document. The document title is "The major aristocratic estates in Belgium in the 19th century: an analysis of three properties belonging to the Arenberg family". The modal window "Where these data come from" provides details on the data's origin:

- Title:**
 - **French**
 - Language originally specified: *missing*
 - Language automatically detected: *Yes*
 - Automatically translated: *No*
 - **English**
 - Language originally specified: *missing*
 - Language automatically detected: *No*
 - Automatically translated: *Yes*
- Abstract:**
 - **English**
 - Language originally specified: *en*
 - Language automatically detected: *Yes*
 - Automatically translated: *No*



How to share and reuse all this



Classification of SSH disciplines

We proposed in TRIPLE a **classification for SSH** based on 27 disciplines.

The starting point was the taxonomy proposed in the **MORESS research project**

- *Mapping of research in European Social Sciences and Humanities, 2003-2005*

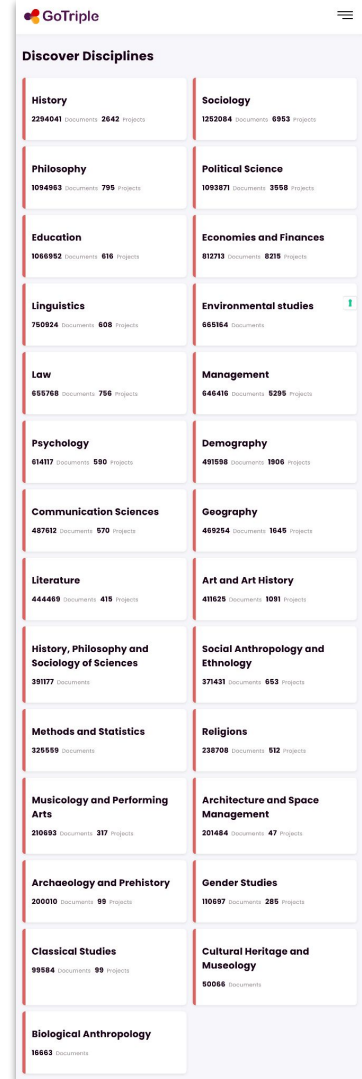
Plus

- Reusable! It is used for classification in other OPERAS services

Minus

- Not linked... until now... (stay with me!)

<https://gotriple.eu/disciplines>



GoTriple

Discover Disciplines

History 2294041 Documents 2642 Projects	Sociology 1252084 Documents 6953 Projects
Philosophy 1054983 Documents 795 Projects	Political Science 1093871 Documents 3558 Projects
Education 1066992 Documents 616 Projects	Economics and Finances 81273 Documents 8219 Projects
Linguistics 750924 Documents 608 Projects	Environmental studies 685164 Documents
Law 650768 Documents 756 Projects	Management 646416 Documents 5295 Projects
Psychology 61417 Documents 590 Projects	Demography 49159 Documents 1806 Projects
Communication Sciences 487812 Documents 570 Projects	Geography 489254 Documents 1645 Projects
Literature 444469 Documents 415 Projects	Art and Art History 41625 Documents 1091 Projects
History, Philosophy and Sociology of Sciences 38177 Documents	Social Anthropology and Ethnology 37431 Documents 653 Projects
Methods and Statistics 325599 Documents	Religions 28708 Documents 912 Projects
Musicology and Performing Arts 210893 Documents 317 Projects	Architecture and Space Management 20184 Documents 47 Projects
Archaeology and Prehistory 20010 Documents 89 Projects	Gender Studies 110697 Documents 285 Projects
Classical Studies 98584 Documents 99 Projects	Cultural Heritage and Museology 90036 Documents
Biological Anthropology 16663 Documents	

Classification of SSH disciplines

We proposed in TRIPLE a **classification for SSH** based on 27 disciplines.

The starting point was the taxonomy proposed in the **MORESS research project**

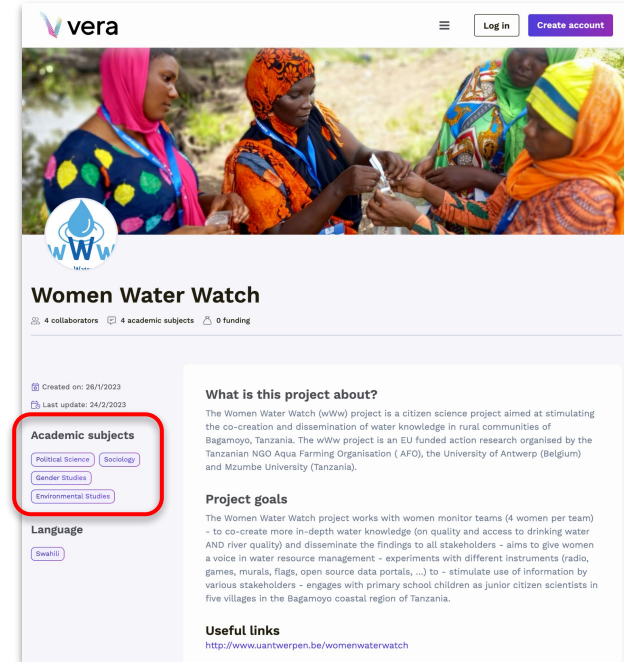
- *Mapping of research in European Social Sciences and Humanities, 2003-2005*

Plus

- Reusable! It is used for classification in other OPERAS services

Minus

- Not linked... until now... (stay with me!)



The screenshot shows a project page for 'Women Water Watch' on the TRIPLE platform. The page features a header with the 'vera' logo, a navigation menu, and 'Log in' and 'Create account' buttons. Below the header is a large image of women in colorful headscarves. The project title 'Women Water Watch' is prominently displayed, followed by statistics: '4 collaborators', '4 academic subjects', and '0 funding'. A red box highlights the 'Academic subjects' section, which lists 'Political Science', 'Sociology', 'Gender Studies', and 'Environmental Studies'. The 'Language' section shows 'Swahili'. The main content area includes a 'What is this project about?' section with a detailed description of the project's goals and a 'Project goals' section. At the bottom, there is a 'Useful links' section with the URL 'http://www.uantwerpen.be/womenwaterwatch'.

<https://gotriple.eu/disciplines>



The TRIPLE Vocabulary

A multilingual **SKOS-based taxonomy** of concepts related to **SSH**

3.375 concepts, manually curated

Labels localised in **11+ languages**

Linked to the Library of Congress Subjects Headings (**LCSH**) vocabulary (*exactMatch*) and others, including Wikidata, CESSDA ELSST (*closeMatch*)

<https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/vocabulary-entries>

The screenshot displays the Semantics.gr platform interface. At the top, it identifies the platform as 'Platform for publishing semantic resources as Linked Open Data' and lists 'Vocabularies - Thesauri - Authority files'. The main content area is titled 'Triple Vocabulary: an SSH multilingual vocabulary based in LCSH'. It provides the LURI: <http://semantics.gr/authorities/vocabularies/SSH-LCSH> and lists metadata such as 'Category: classConcept', 'Provider: Triple Project Consortium', and 'License: Attribution (CC BY 4.0)'. A descriptive paragraph states: '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.go-triple.eu/>); doi: 10.12081/semantics.gr/SSH-LCSH'. Below this is a search bar with a 'Search' button and a 'Clear' button. The results section shows '3375 semantic resources' and lists three entries with their respective labels in multiple languages and associated URIs.



The TRIPLE Ontology

Started as an experiment! A Master thesis of Alessandro Bertozzi, one of my colleagues at Net7.

Main goal: to document and link the controlled vocabularies used during enrichments

Created through an **iterative approach**.

- Currently in its 2nd iteration

<https://gotriple.eu/ontology/triple/index-en.html>

language [en](#)

Triple ontology

This version:
<https://gotriple.eu/ontology/triple/2.0.0>

Latest version:
<https://gotriple.eu/ontology/triple>

Authors:
Alessandro Bertozzi

Download serialization:
[Format: JSON LD](#) [Format: RDF/XML](#) [Format: N-Triples](#) [Format: TTL](#)

License:
[License: license name goes here](#)

Cite as:
Alessandro Bertozzi. Triple ontology. Retrieved from: <https://gotriple.eu/ontology/triple/2.0.0>

Abstract

The ontology for the GoTriple infrastructure

Table of contents

- 1. [Introduction](#)
 - 1.1. [Namespace declarations](#)
- 2. [Triple ontology: Overview](#)
- 3. [Triple ontology: Description](#)
- 4. [Cross-reference for Triple ontology classes, object properties and data properties](#)
 - 4.1. [Classes](#)
 - 4.2. [Object Properties](#)
 - 4.3. [Data Properties](#)
 - 4.4. [Annotation Properties](#)
 - 4.5. [Named Individuals](#)
- 5. [References](#)
- 6. [Acknowledgments](#)



The TRIPLE Ontology

What's there:

- formalisation of Document entities
- schema.org mapping
- definition of the controlled vocabularies for document types, languages, disciplines
- document types linked to COAR entities
- disciplines linked to corresponding TRIPLE Vocabulary entities
- TRIPLE Vocabulary keywords properly linked

Coming soon(ish...):

- formalisation of Profile and linking with Document entities
- definition of the concept of "cluster document"
- formalisation of Project



Reusing data in GoTriple

Data export in Document landing pages

- Multiple formats including JSON and JSON-LD (compliant to the TRIPLE Ontology)

REST APIs

- Again output in JSON or JSON-LD

Documentation available

- See GoTriple.eu footer

The GoTriple platform

Your single access point to discover and reuse open scholarly SSH resources and find peers.
The project is funded by the European Commission, under Grant Agreement No. 863420.

General Information

- [About TRIPLE](#)
- [TRIPLE Data Enrichment](#)
- [TRIPLE Ontology](#)
- [For Content Providers: Your data in GoTriple](#)
- [GoTriple Content Providers Handbook](#)
- [TRIPLE APIs documentation](#)
- [Contact](#)
- [Terms and conditions](#)
- [Privacy policy](#)

Links

- [OPERAS website](#)
- [OPERAS blog](#)
- [European Open Science Cloud - Portal Catalogue and Marketplace](#)
- [Horizon 2020](#)

English ▾



Thanks for your attention!

Luca De Santis

desantis@netseven.it

<https://gotriple.eu/profile/ldesantis>