



1222 • 2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



ARISTOTLE
UNIVERSITY
OF THESSALONIKI

YourTerm: A Collaborative & FAIR European Project

Giorgio Maria Di Nunzio, Pascale Elbaz, Elpida Loupaki, Rodolfo Maslias, Federica Vezzani

University of Padova

Institute of Intercultural Management and Communication

Aristotle University of Thessaloniki

Terminology Coordination Unit, European Parliament



yourterm.eu

Translating and the Computer - TC44
European Convention Center Luxembourg (ECCL)
November 24th, 2022

YourTerm

A Collaborative & FAIR European Project

Translating and the Computer - TC44

Panel Session

European Convention Center Luxembourg (ECCL)

November 24th, 2022

Giorgio Maria Di Nunzio, Federica Vezzani

giorgiomaria.dinunzio@unipd.it, federica.vezzani@unipd.it

Department of Information Engineering, Università degli Studi di Padova

Department of Linguistics and Literary Studies, Università degli Studi di Padova

Theoretical premise

- “Se demander si la terminologie est au centre de l’activité langagière du traducteur spécialisé, c’est comme se demander si un être vivant respire. La question ne se pose pas; la terminologie fait partie de sa vie professionnelle. La qualité de son travail dépend même en grande partie de sa compétence à faire une bonne recherche terminologique.”

How many data every minute?



How many data every minute?



What to do with these data?

Zuckerberg also announced that Meta was working on a universal speech translator, aiming to provide instant speech-to-speech translation across all languages. The company previously set a goal for its AI system to translate all written languages.

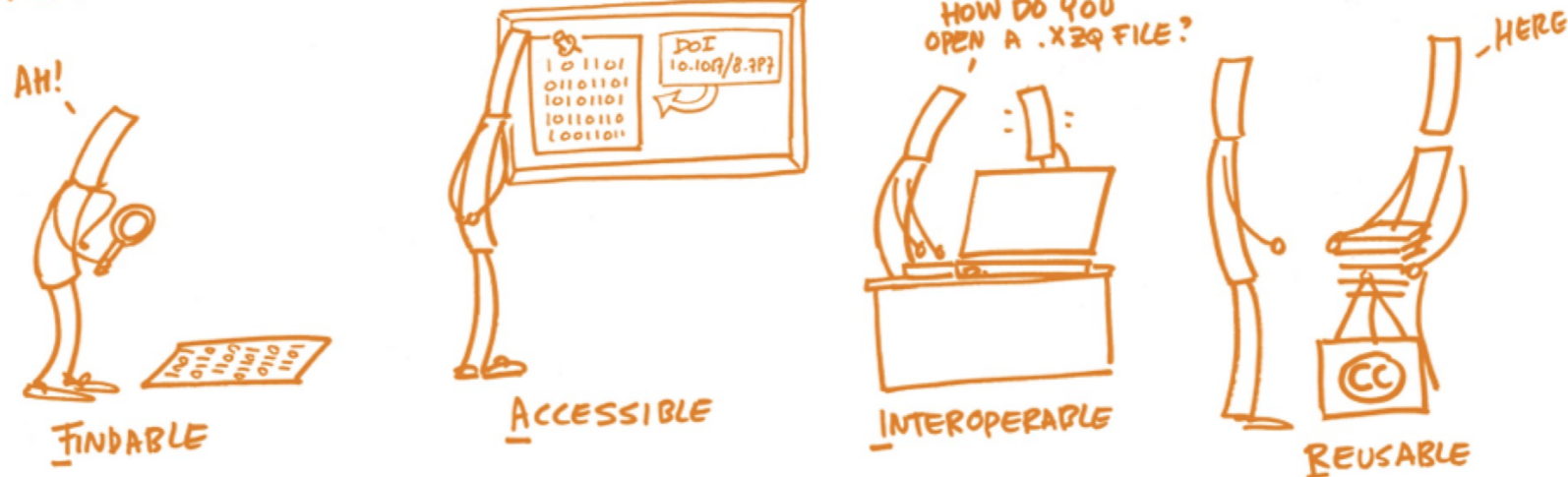


Small toy figures are seen in front of displayed Facebook's new rebrand logo Meta in this illustration taken, October 28, 2021. REUTERS/Dado Ruvic/Illustration/File Photo

**How to organise multilingual
terminology data?**

FAIR Approach

FAIR DATA PRINCIPLES



Findability, Accessibility, Interoperability, Reusability

FAIR - Guiding Principles

- “**Good data management** is not a goal in itself, but rather is the key conduit leading to **knowledge discovery and innovation**, and to subsequent data and knowledge **integration** and **reuse** by the community after the data publication process.”
- “Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments.”

FAIR & EOSC

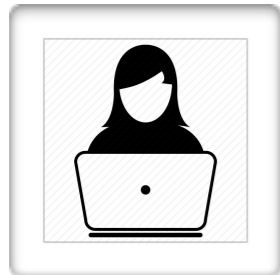
European Open Science Cloud

Data Source

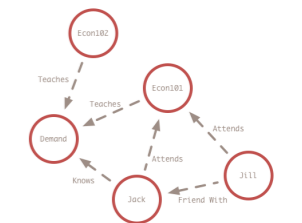
Conversion and Enrichment

Preservation

Re-use



*Systems and Services
Digital Ecosystem*



“FAIR” questions

- Where are the (meta) data published? With what tools and how to search for them?
- In what format are the (meta) data available? Are they findable and how?
- Can the (meta) data be downloaded?
- Can (meta) data be integrated (?) “automatically”?
- Are the (meta) data distributed with any license to use?

FAIR Terminology Paradigm

FAIR Terminology Paradigm




- Objective: to provide guidelines for structuring FAIR terminological data and metadata
- Implemented via *de jure* standards in the context of terminology management (ISO TC 37/SC 3)


FAIR Terminology

- Paradigm to create a terminological resource which:
 - follows an **interoperable** structural model (TMF)
 - allows **access** to terminological data through standard communication protocols
 - provides (meta) data rigorously documented, and therefore **findable**, through a data categories repository
 - ensures data **reuse** through the application of terminology exchange formats (TBX)

ISO standards


International Standard Organization

Standards All about ISO **Taking part** Store   EN 



STANDARDS BY ISO/TC 37

Language and terminology

SUBCOMMITTEE 	SUBCOMMITTEE TITLE	PUBLISHED STANDARDS	STANDARDS UNDER DEVELOPMENT
ISO/TC 37/SC 1	Principles and methods	6	2
ISO/TC 37/SC 2	Terminology workflow and language coding	14	3
ISO/TC 37/SC 3	Management of terminology resources	5	3
ISO/TC 37/SC 4	Language resource management	24	13
ISO/TC 37/SC 5	Translation, interpreting and related technology	13	5

International Standard Organization



International
Organization for
Standardization

- ISO 16642, 2017 - *Computer applications in terminology - Terminological Markup Framework (TMF)*
- ISO 12620, 2019 - *Management of terminology resources - Data category specifications*
- ISO 30042, 2019 - *Management of terminology resources - TermBase eXchange (TBX)*

ISO 30042: 2019 - *TermBase eXchange* (TBX)

Concept

Language

Term

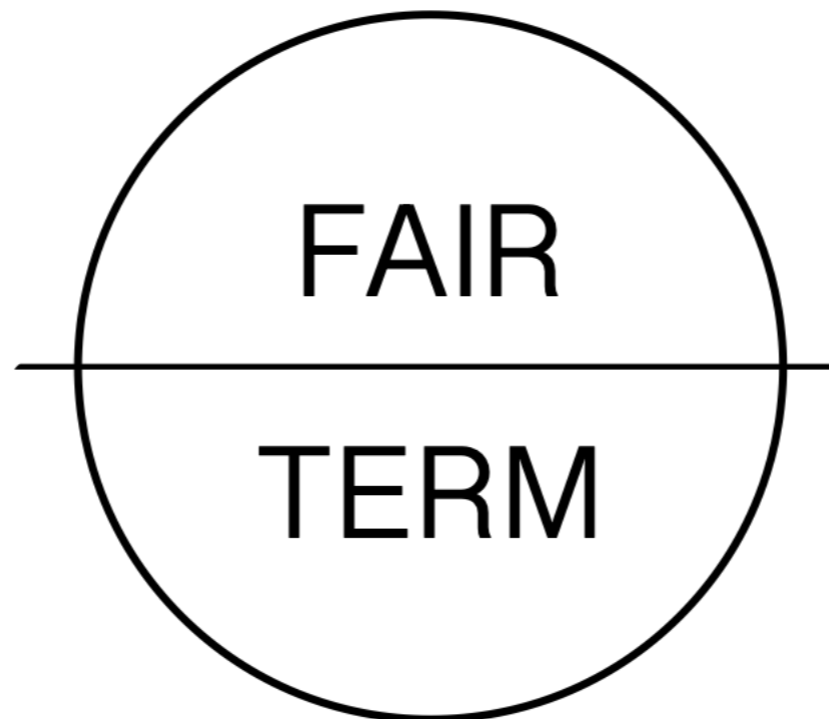
Language

Term

```
<conceptEntry id="c1">
  <subjectField>medicine</subjectField>
  <langSec xml:lang="en">
    <definition>The identification of diseases by the examination of...</definition>
    <source>https://www.thefreedictionary.com/diagnosis</source>
    <termSec>
      <term>diagnosis</term>
      <partOfSpeech>noun</partOfSpeech>
      <grammaticalGender>neuter</grammaticalGender>
    </termSec>
  </langSec>
  <langSec xml:lang="it">
    <definition>Procedura consistente nell'interpretazione di segni e sintomi...</definition>
    <source>http://www.treccani.it/enciclopedia/diagnosi_%28Dizionario-di-Medicina%29/</source>
    <termSec>
      <term>diagnosi</term>
      <partOfSpeech>nome</partOfSpeech>
      <grammaticalGender>femminile</grammaticalGender>
    </termSec>
  </langSec>
</conceptEntry>
```

FAIRterm WebApp

FAIRterm Web Application



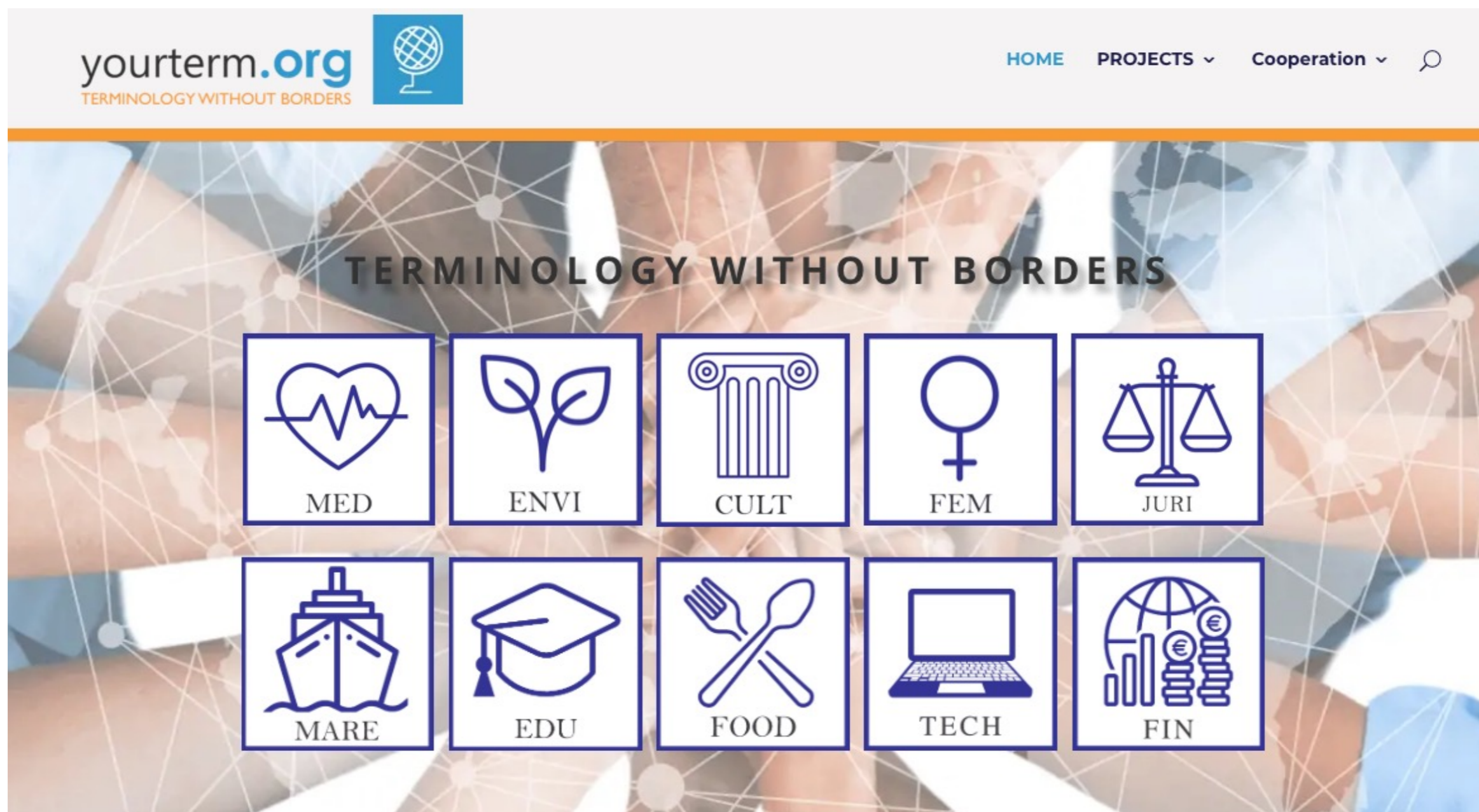
www.purl.org/fairterm

FAIRterm Web Application

- FAIRterm is a free tool for compiling multilingual terminological records
 - Designed to meet FAIR principles
 - Implemented according to ISO TC/37 SC/3 standards
- Working languages: 24 official EU languages + Turkish, Georgian, Russian, Chinese, Japanese and Korean

www.purl.org/fairterm

Terminology Without Borders



<https://yourterm.org>

FAIRterm: data category organisation

FAIR Term

- Create new term
- Compile record
- Download TBX
- Download TSV

Save

Source language

en

Target language

it

Search a term

wine

Term (source)

wine

Term (target)

-(add here)-

Categories

Formal features Semantics Variation Usage

www.purl.org/fairterm

FAIRterm: data display

FAIR Term

- Create new term
- Compile record
- Download TBX
- Download TSV
- Save

Source language	Target language
French	English
Search a term	
bibliothèque numérique	
Source Term	Target Term
bibliothèque numérique	online library
Categories	
Formal features Semantics Variation Usage	
Definition	Definition
ensemble de ressources documentaires mis en ligne et accessible à un public précis	a university's online system which provides resources and databases to support distance learners to conduct research or to consult information
External cross-ref (definition)	External cross-ref (definition)
https://www.enssib.fr/bibliotheque-numerique/	https://www.igi-global.com/dictionary/online-library/39120

www.purl.org/fairterm

FAIRterm: data reuse

```
tbx
1 ?xml version="1.0" encoding="UTF-8"?>
2 <tbx xmlns:tbx3="urn:iso:std:iso:30042:ed-2" xmlns:min="http://www.tbxinfo.net/ns/min" xmlns:basic="http://www.tbxinfo.net/ns/basic">
3   <tbxHeader>
4     <fileDesc>
5       <sourceDesc>
6         <p>An instance of the FAIRterm termbase consisting of one concept entry.</p>
7       </sourceDesc>
8     </fileDesc>
9   </tbxHeader>
10  <text>
11    <body>
12      <conceptEntry id="yourterm_79515">
13        <trimed:conceptIdentifier>yourterm_79515</trimed:conceptIdentifier>
14        <transacGrp>
15          <basic:transactionType>origination</basic:transactionType>
16          <date>2021-05-23</date>
17          <basic:responsibility>sun.celine@isitparis.eu</basic:responsibility>
18        </transacGrp>
19        <basic:note>NA</basic:note>
20        <min:subjectField>Education et communication</min:subjectField>
21        <trimed:subDomain>Service d'information</trimed:subDomain>
22        <langSec xml:lang="bg">
23          <basic:definition>NA</basic:definition>
24          <basic:externalCrossReference>NA</basic:externalCrossReference>
25          <basic:source>NA</basic:source>
26          <note>NA</note>
27        <termSec>
28          <term>- (add here) -</term>
29          <trimed:identifier>yourterm_79515_bg</trimed:identifier>
30          <trimed:termReference>NA</trimed:termReference>
31        <transacGrp>
32          <basic:transactionType>NA</basic:transactionType>
33          <date>NA</date>
34          <basic:responsibility>NA</basic:responsibility>
35        </transacGrp>
36        <note>NA</note>

```

Text Grid Author

YourTerm

A Collaborative & FAIR European Project

Translating and the Computer - TC44

Panel Session

European Convention Center Luxembourg (ECCL)

November 24th, 2022

Giorgio Maria Di Nunzio, Federica Vezzani

giorgiomaria.dinunzio@unipd.it, federica.vezzani@unipd.it

Department of Information Engineering, Università degli Studi di Padova

Department of Linguistics and Literary Studies, Università degli Studi di Padova



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



ARISTOTLE
UNIVERSITY
OF THESSALONIKI

A Collaborative & FAIR European Project: a case study



Pascale ELBAZ & Elpida LOUPAKI



Outline

1. Project's ID
2. Background of the research
3. Terminology Tools
4. Workflow & Methodology
5. Results & Evaluation



Project's ID



Project name	Terminology Without Borders	3
Partners:	TermCoord Aristotle University of Thessaloniki ISIT Paris University of Padova	
Working languages:	ENG, FR, EL	
Domain:	Natural Language Processing (NLP)	
Terminology team:	13 postgraduate students [8 AUTH & 5 ISIT]	
Project coordinators:	Pascale ELBAZ, Professor Elpida LOUPAKI, Assistant Professor	
Institutional supervisor:	Rodolfo Maslias, Nadezhda Krasteva, Angelica Marino & Sofia Vigo, TermCoord	
External experts:	<i>Pascal Tartarin</i> <i>Jean-Baptiste Tanguy</i> <i>Dr Stella Markantonatou</i>	

Background

EMT COMPETENCE FRAMEWORK - 2022

TRANSLATION competences

- ✓ research = evaluate the relevance and reliability of information
- ✓ thematic and domain-specific knowledge
- ✓ mastering systems of concepts, terminology and phraseology, specialised sources

PERSONAL & INTERPERSONAL competences

- ✓ plan and manage time
- ✓ work both autonomously and in virtual, multicultural and multilingual teams
- ✓ Use appropriate communication technologies

Background Previous cooperation between AUTH & ISIT

yourterm.eu Home PROJECTS Cooperation

ISIT

YourTerm EDU Project on: Intercultural Education
Project Language(s): EN – FR – ES – IT
 Click [here](#) to consult the project terms.

YourTerm EDU Project on: New Technologies on Higher Education
Project Language(s): EN – FR
 Click [here](#) to consult the project terms.

The Project | The Project Coordinators | Uni Course | Experts

YourTerm EDU Project on: New Technologies in Higher Education
Project Language(s): FR – EL
 Click [here](#) to consult the project terms.

ARISTOTLE UNIVERSITY OF THESSALONIKI

New Technologies in Higher Education

Term	Notes [Term]	Language ID	Definition	external cross reference [definition]	source [definition]
archive ouverte		FR	base de données documentaire issue de la recherche scientifique et de l'enseignement, groupe de travail qui se tient en ligne via un outil numérique permettant de mettre en	https://coop.isit.chad.fr/publier-et-difuser/deposer-dans-une-archivage-ouvert	D'après Debols, M.C., Déposer ses publications dans une archive ouverte, en 8 points. Montpellier (FRA) : CRAAD, 10
atelier en ligne interactif		FR	ensemble de ressources documentaires mis en ligne et accessible à un public précis	https://www.anapge.com/fr/workshop-ouvert-ouvert/	D'après Anapge : qu'est-ce qu'un workshop 7, 22.07.2020. Consulté le 17.05.2021
bibliothèque numérique		FR	lieu physique permettant aux usagers universitaires d'accéder via Internet à des dispositifs d'évaluation et de certification permettant d'attester un profil de	https://www.enssib.fr/bibliotheque-numerique/	D'après ENSSIB, La bibliothèque numérique. Consulté le 14.05.2021
campus numérique		FR	environnement d'enseignement et d'apprentissage en ligne dans lequel sont	https://denou.lauf.org/campus/	D'après le site initiative pour le Développement du Numéri dans l'Espace Universitaire Francophone (DENUEF), Réseau
certification des compétences numériques		FR	situation dans laquelle deux individus ou plus acquièrent des connaissances et des	https://disciplines.ac.toulouse.fr/faq/la-certification-des-competences-numeriques/	D'après l'Académie de Toulouse, la certification des compétences numériques.
classe virtuelle		FR	conférence utilisant Internet ou des réseaux informatiques privés, pouvant combiner le	https://www.digiforma.com/definition-classe-virtuelle/	D'après Digiforma, Classe virtuelle. Consulté le 12.05.2021
Computer Supported Collaborative Learning		FR	enseignement universitaire dispensé physiquement par un professeur devant un	http://masterprotoc.free.fr/IMG/pdf/csl.pdf	D'après Stéphane Fournand, Le Computer-Supported Collaborative Learning dans la formation
conférence en ligne		FR	type de communication visuelle qui permet de présenter des informations, des produits ou	https://site.europe.eu/search/standards/result/16107508242471	D'après IATE. Consulté le 12.05.2021
cours magistral en présentiel		FR	ensemble de ressources, outils et pratiques utilisant un cadre de partage ouvert dans le	https://www.cert.fr/definition/magistral/	Consulté le 11.05.2021
design numérique		FR	jeune ayant grandi avec Internet et les médias numériques	https://79designs.fr/blog/design-web-digital/design-numerique/	D'après Jamah Johnson, Qu'est-ce que le design numérique Une étude en profondeur, 2020.
éducation ouverte	Le concept d'éducation ouverte est sujet à de nombreux débats car il cherche à	FR	enseignement combinant un apprentissage à distance et apprentissage en présentiel	https://www.openmedproject.eu/fr/etcon-1-1-quel-ce-que-l-education-o/	D'après Cours OpenMed, Qu'est-ce que l'éducation ouverte Consulté le 12.05.2021
enfant du numérique		FR	ensemble intégré de services numériques mis	https://habilmedias.ca/blog/experts/ta-amateurs-louer-comp3C23A/	D'après Matthew Johnson, Experts ou Amateurs? Jouer le compétences en littératie numérique des jeunes canadiens
enseignement hybride		FR		https://webologie.ac.cennes.fr/leip/hb7article4201	D'après l'Académie de Rennes, Enseignement hybride : Comment articuler présentiel et distanciel 7, 09.05.2021 et
espace numérique de	Le Learning Management System (LMS) est	FR		https://eduscol.education.fr/1540/ou	D'après Eduscol, Qu'est-ce qu'un ENT 7, mai 2021.

New Technologies in Higher Education

FR	EN	EL
archive ouverte	open archive	
atelier en ligne interactif	interactive online workshop	
bibliothèque numérique	online library	ηλεκτρονική βιβλιοθήκη
campus numérique	virtual campus	
certification des compétences numériques		
classe virtuelle	virtual classroom	εικονική τάξη
Computer Supported Collaborative Learning	Computer Supported Collaborative Learning	
conférence en ligne	online conference	
cours magistral en présentiel	in-person mass lecture	
design numérique	digital design	
éducation ouverte	open education	

Background

Importance of NLP in everyday communication

Examples: speech recognition, automatic translation, terminology extraction, word segmentation

Lack of terminological resources in this domain

Terminology Tools for term extraction, corpus management & text analysis



TermoStat Web 3.0

FAIR Term

Project
Nothing selected

Subproject
Nothing selected

+ Create new record

🔍 Compile record

📄 Download TSV


📄 Download TBX

Save changes

Source language of your term
Nothing selected

Term
(write term here)

Add new term



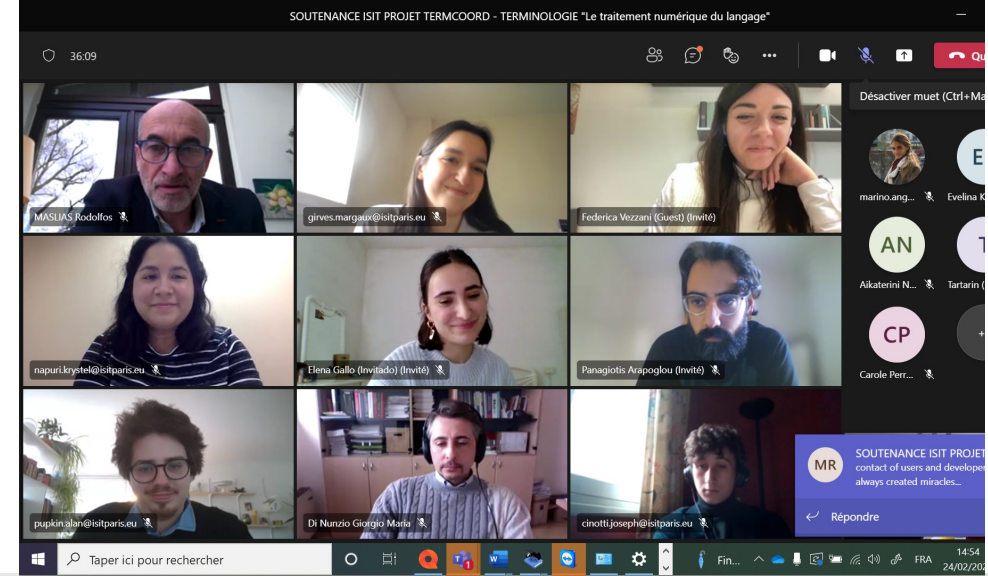
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

FAIRTerm, a new collaborative tool

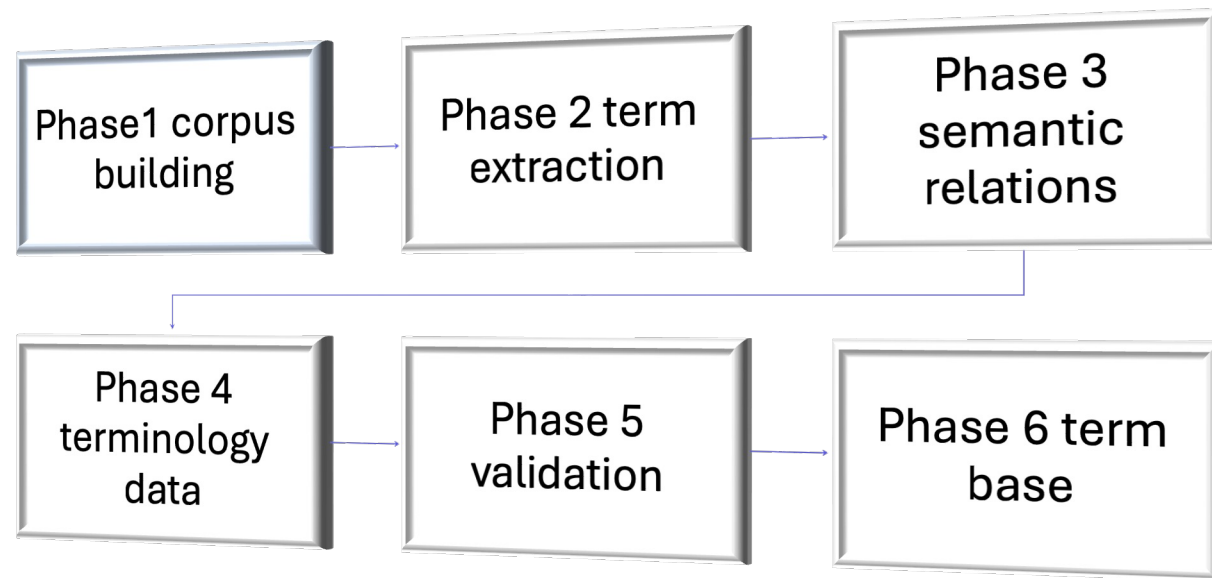
- Collaborative term base
- Developed by G. Di Nunzio & F. Vezzani, University of Padova
- Designed according to the ISO standards
- Designed to meet the FAIR principles
- Free access
- 24 official EU languages + Turkish & Russian

Collaborative methods

- FAIR Term was used.
- All major phases were followed by a videoconference, between the French & Greek groups, using Microsoft Teams, where different issues were discussed.
- Two meetings with TermCoord were organized, using Webex.
- One tutorial & one problem-solving session were organized with the University of Padova via ZOOM.
- One meeting with the experts was organized in each University.
- One teaching mobility was realized.



Workflow



ARISTOTLE
UNIVERSITY
OF THESSALONIKI



PARIS PANTHÉON-ASSAS UNIVERSITÉ



**TERMINOLOGY
COORDINATION**
European Parliament
DG TRAD

computational model	A computational model is a computer implementation of the solution to a (scientific) problem for which a mathematical representation has been formulated. Developing a computational model consists of applying Computer Science concepts, principles and methods.
υπολογιστικό μοντέλο	Μια υπολογιστική προσομοίωση ή υπολογιστικό μοντέλο είναι ένα πρόγραμμα υπολογιστή που προσπαθεί να μιμηθεί ένα αφηρημένο μοντέλο ενός συγκεκριμένου συστήματος. [...]
modèle computationnel	Un modèle computationnel est avant tout l'implantation mathématique d'une conception théorique du système cognitif:
decision tree δέντρο απόφασης	Decision trees are a well-established classification/prediction methodology in machine learning, in which the model is a tree where each node is a decision, and each leaf represents an output class.
	Το δέντρο απόφασης δημιουργεί μοντέλα ταξινόμησης ή παλινδρόμησης με τη μορφή δομής δέντρου. Διασπά ένα σύνολο δεδομένων σε μικρότερα και μικρότερα υποσύνολα ενώ παράλληλα αναπτύσσει σταδιακά ένα σχετικό δέντρο αποφάσεων. Το τελικό αποτέλεσμα είναι ένα δέντρο με κόμβους απόφασης και κόμβους φύλλων
arbre de décision	Représentation hiérarchique de la structure des données sous forme des séquences de décisions (tests) utilisée dans l'exploration de données et en informatique décisionnelle, en vue de la prédiction d'un résultat ou d'une classe.

Project Results

Three corpora were constructed:

148 000 words = FR corpus

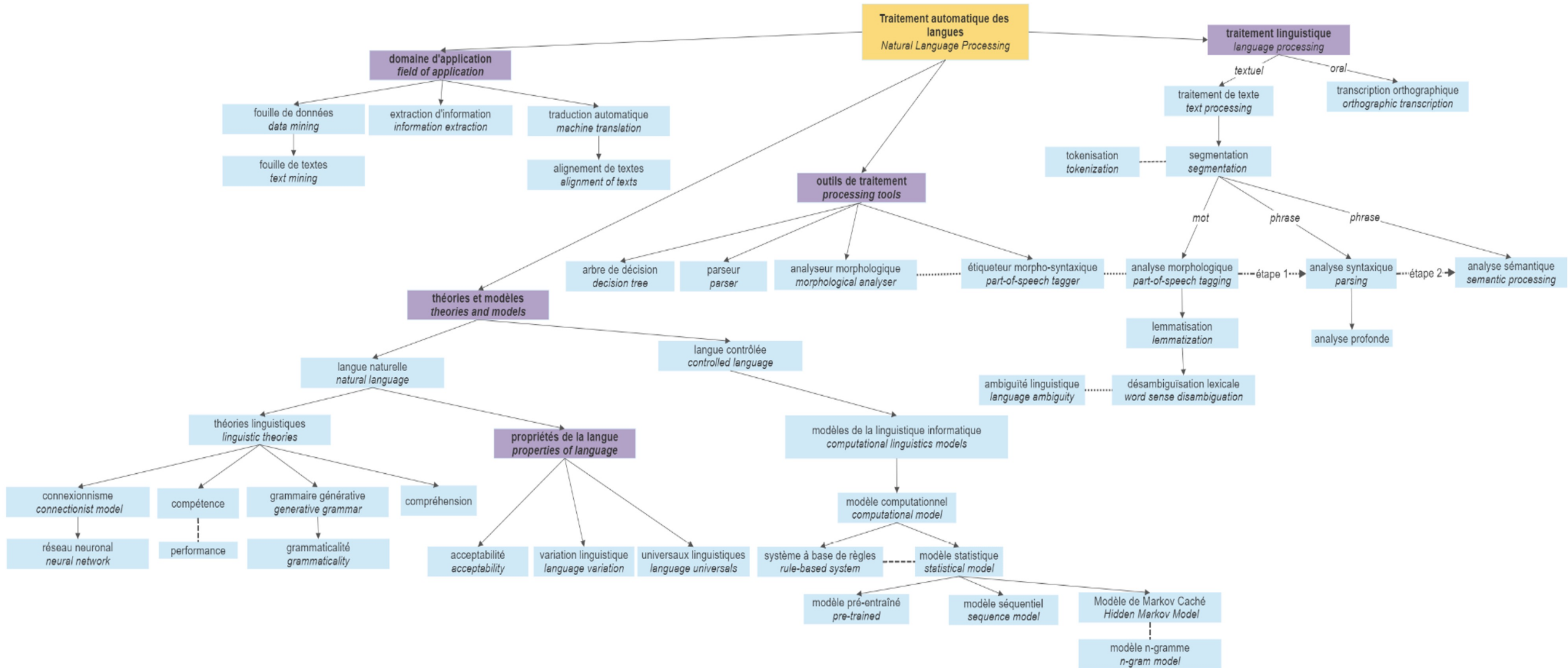
784 346 words = ENG corpus

173 062 words = EL corpus

30 concepts were developed

90 terms (ENG/ FR/ EL) were fully documented

Arborescence



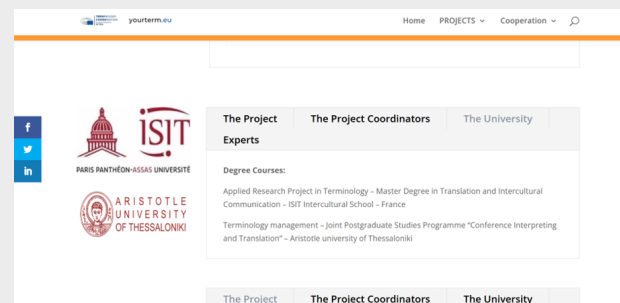
Project evaluation

Development of new competences for students: terminology tools, knowledge of the special field, analytical and synthetical competences, etc.

Development of soft skills for students: working with a European institution, collaborate within an international team, present and support arguments, etc.

Project's dissemination

- <https://termcoord.eu/2022/08/yourterm-aristotle-university-of-thessaloniki-isit-paris-are-going-collaboratively/>
- <https://yourterm.eu/tech-university-projects/>



A photograph showing several hands holding interlocking gears of various sizes. The scene is backlit by a bright, warm light, likely a sunset or sunrise, creating a silhouette effect on the hands and gears. The background is a soft gradient of orange, yellow, and light blue. The text "Thank you!" is overlaid in white on the left side of the image.

Thank you!

For further information & collaborations

Pascale ELBAZ, ISIT, ELBAZ.Pascale@isitparis.eu

Elpida LOUPAKI, Aristotle University of
Thessaloniki, eloupaki@frl.auth.gr