

## Abstract

CLARA project aims to empower teachers to facilitate the creation of licensable Educational Resources (ER) based on existing ones. Our approach will suggest a relevant set of ERs that are coherent with a course sketch and have compatible licenses. The main challenges we face are (1) how to enrich a network of ERs using AI algorithms, and (2) how to guarantee a minimal set of license-compatible ERs relevant to a given course goal with query relaxation techniques. We will exploit ERs provided by the French Ministry of Education and the X5-GON\* project.

## Goals

- WP1.** To build a corpus of linked ERs with a rich semantic representation (available metadata, licenses, similarities, dependencies, etc.).
- WP2.** To enrich the semantic annotations with machine learning algorithms (concepts, temporality of concepts, prerequisites, etc.).
- WP3.** To design a query engine that facilitates finding relevant license-compatible educational resources.
- WP4.** To involve teachers to test our solutions all along the project.

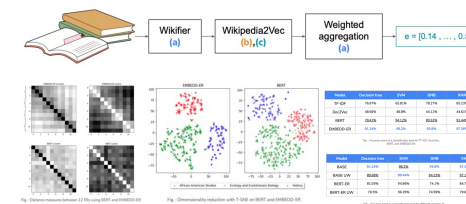
## Ongoing planning

Workpackage	Partners	Year 1				Year 2				Year 3				
		T3	T6	T9	T12	T15	T18	T21	T24	T27	T30	T33	T36	
WP0	Project management													
WP1	Building a collection of OER													
Task 1.1 (OER corpus)														
Task 1.2 (JPIP graphs)														
Task 1.3 (License analysis)														
WP2	Enriching the graph of OER													
Task 2.1 (State-of-the-art analysis)														
Task 2.2 (Representation learning)														
Task 2.3 (Identification of learning paths)														
WP3	OERs integration for a target course													
Task 3.1 (Sketch definition)														
Task 3.2 (Planning results)														
Task 3.3 (Planning goals)														
WP4	User case testing													
Task 4.1 (Repository testing)														
Task 4.2 (Intermediate prototype testing)														
Task 4.3 (Final prototype testing)														

## Main achievements

### EMBEDD-ER: EMBEDDing ERs using LOD [3]

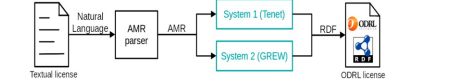
**Goal:** Describe efficiently ERs' content and unseen ERs.  
**Approach:** Based on ER'  
**Target:** Embeddings with entities in Wikipedia.



Several experiments validate our approach.

### Workflow for Extracting ODRL Licenses from Text [4]

**Goal:** Automatically analyse license texts.  
**Approach:** Symbolic and formal methods.  
**Target:** RDF representation (ODRL vocabulary\*).



We developed 2 systems for this problem. They use AMR as an intermediate representation. They successfully process simple sentences.

	Modalities			Actions			Global Ratio
	P	R	F	P	R	F	
Tenet System	0.780	0.640	0.703	0.989	0.810	0.891	0.630
GREW System	0.872	0.820	0.845	0.990	0.910	0.948	0.810

\*https://www.w3.org/TR/odrl-vcocab/

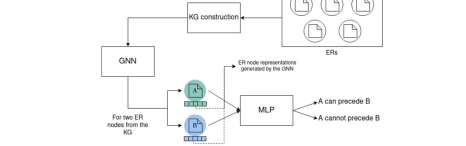
### Benchmark for Evaluating RDF Reification [5]

**Goal:** Evaluation of the impact of RDF reification models and multi-valued properties on several triplestores.  
**Approach:** Experiments with four reifications models, 28 SPARQL queries over four triplestores.  
**Target:** A reification model and triplestore for our KG and a benchmark (datasets and queries).

Triplestores	RDF reification models						
	1	2	3	4	5	6	7
Virtuoso	1	1	1	1	1	2	1
Jena	2	4	4	3	4	3	4
GraphDB	2	2	2	2	2	1	2
Oxigraph	4	3	3	4	3	4	3

### PreSAGE: Precedability Prediction Between ERs [13]

**Goal:** To help educators and learners to navigate through ERs.  
**Approach:** Based on ERs' textual expression, an enriched KG is built. Thanks to this KG and GraphSAGE (an existing GNN model), PreSAGE is able to predict precedability.  
**Target:** Predict precedability between ERs. We define precedability as the possibility of one element to precede another.



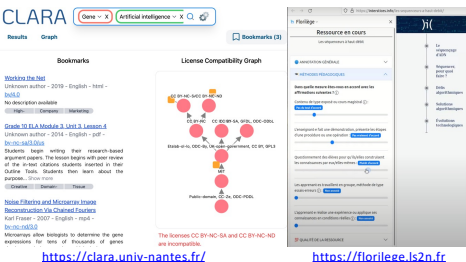
### A Survey on SPARQL Query Relaxation [12]

**Goal:** Overview of SPARQL query relaxation literature with a focus on challenges posed by RDF annotation models: standard reification, named graphs, n-ary relations, singleton properties, and RDF-Star.  
**Approach:** Comparative analysis of 12 works (published between 2006 and 2021) organized in two parts: logical relaxation (ontology-based) and similarity of instances.  
**Open issues:** (a) Querying both data and metadata generates larger queries, increasing the possibilities for relaxation. Therefore, ontology-based relaxation solutions should optimize their pruning methods. (b) Different RDF reification models result in varied forms of RDF graphs, which means that similarity-based approaches that follow graph structures need to be adapted accordingly. (c) Overall, this survey reveals that no existing query relaxation solution effectively leverages both RDF triples and their annotations.

### A Hybrid Approach for SPARQL Query Relaxation

**Goal:** An approach for query relaxation that efficiently leverages RDF data and their annotations.  
**Approach:** Hybrid query relaxation model that integrates both entity-based and ontology-based relaxation techniques.  
**Target:** (1) Develop a new lattice pruning technique that uses a similarity function, optimally integrating class and property similarity with instance and literal similarity, relative to the original query. (2) Create a new benchmark for evaluating relaxation methods for queries that target both data and metadata. This benchmark will support the experimental evaluation, enabling comparison of our model with state-of-the-art approaches.

## Demonstrations



## Main publications

We have 13 publications: <https://project.inria.fr/clara/publications-2/>

[13] A. Bazouzi, H. Le Capitaine, Z. Miklos, M. Foursov. *Precedability Prediction Between Open Educational Resources*. In GoodIT, Sept. 2024. hal-04654407.

[12] G. Fakh, P. Serrano-Alvarado. *A Survey on SPARQL Query Relaxation under the Lens of RDF Reification*. In Semantic Web Journal, IF 3, CiteScore 2024 8.3, SJR 2023 0.86, Q2 Sept. 2024. hal-04677368.

[11] M. Kieffer, H. Chabane, M. Lécivain, P. Serrano-Alvarado. *CLARA search engine: Linking Licensed Educational Resources*. Demo paper. In ESWC, May 2024. hal-04586246.

[8] M. Kieffer, G. Fakh, P. Serrano-Alvarado in collaboration with Corentin Follenfant and Colin de la Higuera. *CLARA Knowledge Graph of licensed educational resources*. March 2023, 10.5281/zenodo.8403141. hal-04754656v1.

[5] M. Kieffer, G. Fakh, P. Serrano-Alvarado. *Evaluating Reification with Multi-valued Properties in a Knowledge Graph of Licensed Educational Resources*. In SEMANTICS, Sept. 2023. hal-04160947v3

[4] M. Revel, A. Lamercurie, A. Foret and Z. Miklos. *Extracting ODRL Digital Right Representations from License Texts using AMR*. In ASAIL, Jun. 2023. hal-04449668

[3] A. Bazouzi, M. Foursov, H. Le Capitaine, Z. Miklos. *EMBEDD-ER: EMBEDDing Educational Resources Using Linked Open Data*. In CSEDU, Apr. 2023, ISBN 978-989-758-641-5. hal-04037990

## Follow-up PARTAGE

**PARTAGE** is a CominLabs innovation action that began in September 2024, aimed at providing teachers with integrated tools to find, access, and remix Open Educational Resources. Building on the achievements of the CLARA and Florilège projects, PARTAGE will leverage Semantic Web technologies and AI-enhanced metadata to implement **FAIR** principles (Findable, Accessible, Interoperable, Reusable) for educational resources.