

## 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 L I D G S R C T 2 I S N N S U	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 (RDF graph)														
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 (Ranking results)														
Task 3.3 (Relaxing goals)														
WP4	Use case testing													
Task 4.1 (Repository testing)														
Task 4.2 (Intermediate prototype testing)														
Task 4.3 (Final prototype testing)														

## Work so far

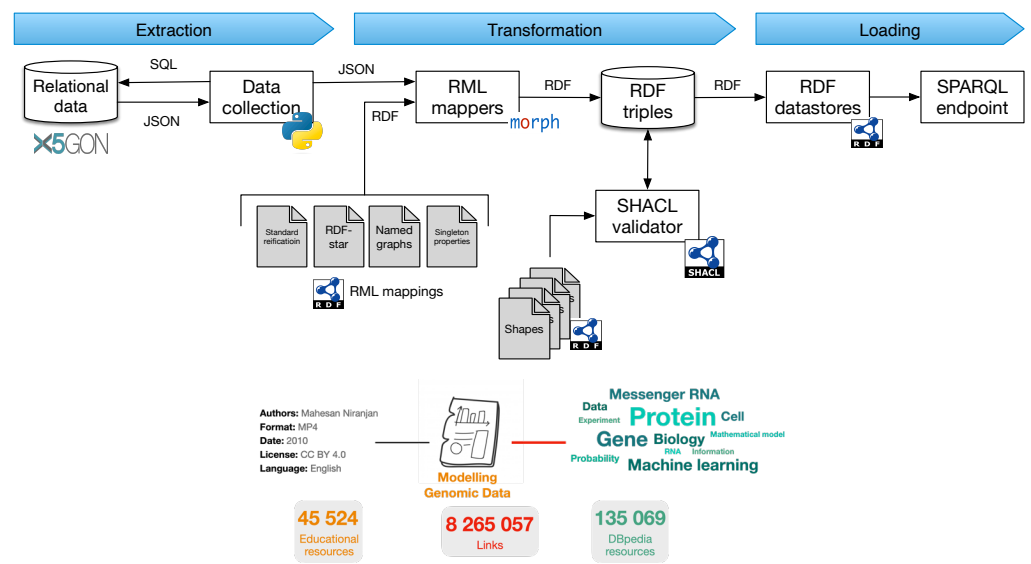
### Research contributions

- ✓ A knowledge graph of educational resources [1] available as a SPARQL endpoint <https://clara.univ-nantes.fr/sparql>.
- ✓ Two systems to automatically produce machine readable licenses from licenses in natural language [2].
- ✓ A method to generate representations for educational resources based on document annotation and Linked Open Data [3].
- ✓ A state of the art of SPARQL query relaxation works [4].

### Demonstrations

- ✓ A search engine for educational resources with licenses verification <https://clara.univ-nantes.fr>.
- ✓ The Florilege platform, a catalogue of French language educational resources with the indispensable metadata and a firefox extension <https://florilege.ls2n.fr>.

## Knowledge graph pipeline (WP1)



## Evaluating RDF reification [1] (WP1)

**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 knowledge graph of educational resources.

	Triplestores							Total
	Q1: Property path	Q2: Join	Q3: Filter	Q4: Join & filter	Q5: Optional	Q6: Union	Q7: Group by	
Virtuoso	1	1	1	1	1	2	1	1 <sup>st</sup>
Jena	2	4	4	3	4	3	4	3 <sup>rd</sup>
GraphDB	2	2	2	2	2	1	2	2 <sup>nd</sup>
Oxigraph	4	3	3	4	3	4	3	4 <sup>th</sup>
Best models	SR NG	SR SR	SR NG	SR NG	SR NG	SR NG	SP NG	NG SR

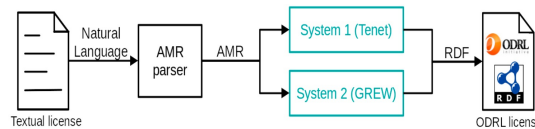
SR: Standard reification  
NG: Named graph  
SP: Singleton properties  
RB: RDF-star

## ODRL license extraction [2] (WP1)

**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

P: Precision, R: Recall, F: F-measure

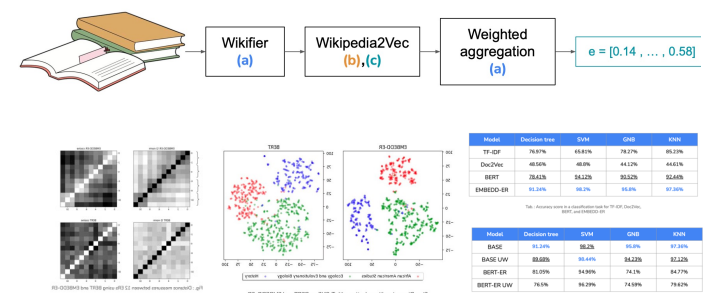
\*<https://www.w3.org/TR/odrl-vocab/>

## Embedding ERs [3] (WP2)

**Goal:** describe efficiently ERs' content (subjects) and unseen ERs.

**Approach:** use Wikipedia2Vec over ERs annotated with Wikipedia resources (wikifier.org).

**Target:** embeddings with entities in Wikipedia.



Several experiments validate our approach.

## Survey on query relaxation [4] (WP3)

Survey of 12 works (2006 to 2021).

General comparison of the analysed works: query shapes, necessary information, terms relaxed, lattice pruning, techniques to avoid redundancy of results, techniques for query ranking, etc.

Our conclusions make evident that query relaxation methods do not behave as expected for queries over reified triples.

## Demonstrations (WP4)

## Publications

- [1] Manoé Kieffer, Ginwa Fakh, Patricia Serrano-Alvarado. Evaluating Reification with Multi-valued Properties in a Knowledge Graph of Licensed Educational Resources. In *International Conference SEMANTICS*, September 2023, Leipzig, Allemagne.
- [2] Malo Revel, Aurélien Lamerclerie, Annie Foret and Zoltan Miklos. Extracting ODRL Digital Right Representations from License Texts using AMR. In *International Workshop ASAIL*, June 2023, Braga, Portugal.
- [3] Aymen Bazouzi, Mikael Foursov, Hoel Le Capitaine, and Zoltan Miklos. EMBEDD-ER: EMBEDDing Educational Resources Using Linked Open Data. In *International Conference CSEDU*, April 2023, Lisbon, Portugal.
- [4] Ginwa Fakh, Patricia Serrano-Alvarado. A Survey on SPARQL Query Relaxation under the Lens of RDF Reification. In *French Conference BDA*, October, 2023, Montpellier, France.