

## SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

**Action number: 18209**

**STSM title: Rights Management of Linguistic Linked Data**

**STSM start and end date: 2021-07-12 to 2021-07-31**

**Grantee name: Víctor Rodríguez Doncel**

### PURPOSE OF THE STSM:

#### **Objective**

To design representation methods of legal rights in the context of Linguistic Linked Data, with a focus on licenses for Language Resources and Technologies and licensing compatibility issues.

#### **Specific activities**

- 1) Identification and review of the policies of interest for ILST.
- 2) Update of the existing mappings to RDF (which is incomplete, based on the old ODRL 2.1 and contain errors)
- 3) Definition of a some examples / references to serve as testcases.
- 4) Implement license-related algorithms: (i) defeasibility (weak/strong restrictions in contradiction); (ii) better modelling of cascade restrictions in workflows and software-related restrictions and (iii) combination of licenses.
- 5) Set up of a demo/live HTTP REST API, which is already in place but must be much beautified.

#### **Expected outcome**

An academic research paper to be submitted to either LREC or JURIX conferences (Language Resources and Evaluation and Legal Knowledge and Information Systems respectively).

### DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

The STSM took place in the expected time frame and no COVID-19 related issues affected its successful development. Researchers regularly met at the ILSP premises, counting with the participation of: Penny Labropoulou (ILSP), Katerina Gkirtzou (ILSP), Dimitris Galanis (ILSP). Besides the visitor (Víctor Rodríguez-Doncel, UPM), other researchers in his group eventually joined meetings (Rana Saniei, Beatriz Esteves). The following activities were carried out:

- Joint discussions on the requirements
- Joint modelling of the domain information.
- Joint writing of technical specifications
- Development of source code and its cross-validation

Discussions have been based on real use cases of licences and terms of use for services distributed via large infrastructures in the LRT domain, mainly the European Language Grid and CLARIN.

### DESCRIPTION OF THE MAIN RESULTS OBTAINED:

The mission has yielded technological and non-technological results.

The main technological results are a data model, a dataset using the data model, and services consuming the data. The non-technological results include the consolidation of the cooperation lines between the two institutes, a presentation given by the mission visitor to the Athena RC institute (given virtually due to covid 19 restrictions), discussions with other ILSP group members and plans for future cooperation lines. The technological results are:

- **DATA MODEL. A draft ODRL profile for language resources and technologies.** ODRL is a language to represent generic policies and licenses, which lacked elements specific to this domain. As a result of the research visit, a new profile has been created and published at <http://rdflicense.linkeddata.es/profile.html> to be in a future as <https://w3c.github.io/odrl/lr/> The model takes into account features that can be subsequently used for assessing compatibility
- **DATA. Relevant to the language resources and technologies domain**, a dataset of licences with different structured interpretations was collected and published (<https://rdflicense.linkeddata.es/> )
  - Collection of XML rights expressions used in the Greek CLARIN and the European Language Grid.
  - A version of the former in RDF, conformant to the Meta-Share ontology.
  - A version of the former, transformed into valid ODRL policies, and conformant with the profile described above.
  - Mappings to other external efforts to represent licences in a structured form (the results of the DALICC research project, SPDX list of licenses)
- **SERVICES.** Two HTTP REST API services to transform Meta-Share XML rights expressions into (a) instances of the Meta-Share ontology. (b) valid ODRL policies using the data model described above.

#### **FUTURE COLLABORATIONS**

- Continue collaboration between the two institutes
- Continue work for the data model in the framework of the ODRL and LD4LT W3C Community Groups
- Started work on a publication to be submitted in the next LREC conference and possibly a relevant journal.