

Report on the outcomes of a Short-Term Scientific Mission¹

Action number: CA18209

Grantee name: Anas Fahad Khan

Details of the STSM

Title: Domain Labels in Linked Lexicographic Resources

Start and end date: 20/03/2023 to 30/03/2023

Description of the work carried out during the STSM

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

(max. 500 words)

During the period of the STSM, Dr Khan was able to collaborate with Professor Costa's group at CLUNL on the compilation of a set of draft guidelines for encoding domain labels for linked data lexical resources in RDF -- as was proposed in the original work plan*. This collaborative effort built on previous work carried out both by Prof. Costa's group as well as by Dr Khan. The intention from the start was for these draft guidelines to contribute to Deliverable D1.2 of the COST action. In addition the work also suggested other potential sets of guidelines for encoding retro-digitized dictionaries.

More practically, during his research visit Dr Khan worked on the conversion into RDF of a number of representative entries from *Dicionário da Língua Portuguesa*, the Portuguese language dictionary of the Academia das Ciências, as well the retro-digitised *Diccionario da lingua portuguesa* compiled by de Moraes Silva; this work made use of the XSLT stylesheet developed by John McCrae and Laurent Romary to convert the original TEI-XML source into OntoLex**. These entries will form the basis of future work aiming at making both resources available as linked open data, and ultimately in helping to make more high quality resources available for the Portuguese language on the Linked Open Data cloud. It will also hopefully inform future work on publishing retrodigitized dictionaries in LLOD: especially those which contain extensive domain label information. In addition, Dr Khan gave two presentations of his work during the period of his STSM: first at the University of Coimbra where he was hosted by Hugo Gonçalo Oliveira as well as at the Universidade Nova of Lisbon itself.

¹ This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

In the latter case, Dr Khan was able to meet and interact with Professor Manuel Portela's group in the Department of Languages, Literatures, and Cultures and discuss shared research interests.

*These draft guidelines can be found here:
<https://github.com/anasfkhan81/EncodingDomainLabelsRDF/blob/main/Guidelines.md>

**These examples can be found here:
<https://github.com/anasfkhan81/EncodingDomainLabelsRDF/tree/main/Examples>

Description of the STSM main achievements and planned follow-up activities

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

(max. 500 words)

The STSM achieved its planned goals: the draft guidelines which have been produced during this STSM are, the authors feel, of a sufficient quality to be proposed as one of the series of the guidelines and best practices currently being developed as part of D1.2 of Nexus Linguarum. They will also be published on Zenodo and proposed as a set of ontology design patterns (Dr Khan has already requested an account on the Ontology Design Pattern wiki in order to submit his proposed patterns). As mentioned above, the collaboration between Dr Khan and Professor Costa's group has also suggested other sets of guidelines which could also contribute to D1.2, especially when it comes to retrodigitised legacy dictionaries. In addition, this collaboration has resulted in suggestions towards the improvement of the XSLT guidelines produced by McCrae and Romary for mapping TEI-XML lexicons to OntoLex-Lemon. It has also brought to light several inconsistencies and oddities in the definitions of classes and properties in lexinfo 3.0 which will be raised as issues on the lexinfo github. The work on converting entries into RDF and in particular the conversion of the Morais dictionary is ongoing. In particular, further virtual meetings between Dr Khan and Professor Costa are already planned for the rest of the year.