Linked Pasts 12 Hybrid and Satellite Events

23.November.2026, 14:00-16:30, FLUP

GOBLIN Use Case 4.1.4 Integrating heterogeneous lexicographic datasets in Wikidata Lexemes

David Lindemann (EHU University of the Basque Country), Ana Salgado (CLUP – Centre of Linguistics of the University of Porto)

This GOBLIN COST Action use case builds on work undertaken in summer 2026 by the Institute of Lexicology and Lexicography of the Portuguese Language (ILLLP), Lisbon Academy of Sciences, with the University of the Basque Country (EHU). Data from selected lexical entries, from the Dicionário da Língua Portuguesa (DLP), including sense descriptions, were transformed from TEI Lex-0 into OntoLex-Lemon and stored in ILLLP Wikibase. From there, approximately 70,000 entries were integrated into Wikidata lexemes, a multilingual collection whose data model incorporates core OntoLex-Lemon elements, with persistent identifiers linking lexemes and senses to DLP records. Manual validation included homograph disambiguation and peer review of sense alignments (see project report).

Further enrichment of Portuguese lexemes in Wikidata will combine content from lexical resources, computational methods, and community contributions. An already ongoing initiative aims to add inflected forms from a Portuguese morphological resource used in NLP. Other priorities include etymologies, corpus-based usage examples, pronunciation, demonyms and validated links between senses and Wikidata items. Enrichment should preserve source attribution at entry, sense and form levels, including regional and register information.

The use case will document and evaluate reusable workflows for data selection, modelling, alignment, validation and updating. Lexicographers, linguists, language-technology researchers and Wikidata contributors will enrich the Portuguese lexemes collection on Wikidata, and assess an adaptation of such workflows to other languages and resources. Expected outcomes include documented procedures, validation guidelines, reusable queries and scripts, and prioritised enrichment tasks.

The preliminary timetable for this Use Case includes the following:

- Months 1–3: document the existing workflow, agree on priorities and establish validation criteria.
- Months 4–6: carry out a Portuguese enrichment pilot and assess the results.
- Months 7–9: identify interested partners and a suitable resource for testing the workflow in another language.
- Months 10–12: undertake that adaptation, subject to partner participation and data availability, and consolidate the procedures, queries and scripts.

25.November.2026, 10:00-12:00, UPDigital Building

Discovering Cultural Heritage Knowledge Graphs

Uldis Bojārs (National Library of Latvia)

Museums, archives, libraries and research projects have published a large number of cultural heritage knowledge graphs (KG) but many of them can be hard to find. Information about them, including their description, data model and access methods, is scattered across the web and is not documented consistently. Generic registries such as the LOD Cloud (https://lod-cloud.net/) do not describe these KGs in sufficient detail and their information is often outdated. To address these shortcomings, the GOBLIN COST Action (CA23147 - Global Network on Large-Scale, Cross-domain and Multilingual Open Knowledge Graphs; https://www.cost.eu/actions/CA23147/) is building a catalogue of cultural heritage knowledge graphs, with detailed KG descriptions and metadata.

A workshop on Cultural Heritage Knowledge Graphs, organized by GOBLIN in Coimbra in January 2026, let its participants collaborate in person in order to improve the state of information about cultural heritage KGs and collect detailed dataset descriptions. This session brings that work to the Linked Pasts community.

The main goals of this session are:

- to bring together people interested in improving the state of information about cultural heritage KGs and in making them easier to discover;

- to extend the catalogue with information about knowledge graphs that have been created by the Linked Pasts community and are important to it;

- to give newcomers a guided, practical introduction to the cultural heritage linked open data.

The session will be divided into three parts and delivered in a hybrid format, with both onsite and online participation:

- Part 1 (15 min): Introduction to the task, the metadata model and what has already been done in the GOBLIN Cultural Heritage Knowledge Graphs activity.

- Part 2 (60 min): Cataloguing. Participants will work in teams, select KGs from their field and add the information about these KGs to the catalogue.

- Part 3 (15 min): Reporting back and further exploration. Each team will give a brief report on what it catalogued in Part 3. After the reports, participants will discuss how the catalogued KGs and the metadata catalogue can be used, and will explore the KGs that they are interested in. At the end of the session we will collect participant feedback in order to continue this work after the symposium.

Main audience: This session is for anyone who works with cultural heritage knowledge graphs or is interested in this topic. Newcomers are welcome, no previous semantic technology experience is necessary (we will form mixed teams where newcomers can learn from the more experienced participants).

26.November.2026, 10:00-17:00, FLUP

Wikidata and Wikibase for the Digital Humanities

Thursday, 26th November 2026 | 10:00-17:00 | University of Porto

Camillo Carlo Pellizzari di San Girolamo (Scuola Normale Superiore, Pisa)

This workshop aims to provide medium to advanced training regarding the use of Wikidata, and other Wikibase instances, for Digital Humanities projects. The main themes covered will be gadgets and tools which can make editing, data modelling and authority control more efficient. The workshop can be adapted according to the requests of the participants, adding more introductory information if needed. In the afternoon there will be time for questions from the public regarding specific issues in Wikidata and/or other Wikibase instances. Both the morning and the afternoon will alternate theoretical explanations and Wikidata editing.