Prêt-à-LLOD

D5.5 Community Building Report

Author(s): Thierry Declerck (DFKI),
John McCrae (NUIG), Jorge Gracia
(UNIZAR), Julia Bosque-Gil
(UNIZAR)
Date: 31.03.2021

H2020-ICT-29b

Grant Agreement No. 825182

Prêt-à-LLOD - Ready-to-use Multilingual
Linked Language Data for Knowledge
Services across Sectors

D5.5 Community Building Report

Deliverable Number: D5.5
Dissemination Level: PU
Delivery Date: 31.03.2021
Version: Final
Author(s): Thierry Declerck (DFKI), John McCrae (NUIG), Jorge Gracia (UNIZAR),
Julia Bosque-Gil (UNIZAR)

Document History

Version Date	Changes	Authors
02.03.2021	First version	Thierry Declerck (DFKI)
15.03.2021	Second version	Thierry Declerck, Julia Bosque Gil (UNIZAR)
23.03.2021	Third version	Thierry Declerck, Jorge Gracia (UNIZAR)
25.03.2021	Pre-Final version	Thierry Declerck
30.03.2021	Internal Review	John McCrae (NUIG)
31.03.2021	Final version	Thierry Declerck

Table of Contents

D5.5 Community Building Report

1 Introduction

This is the second deliverable reporting on activities pursued in the context of Task 5.4. "Community Building", which is a part of WP5 "Language Resource and Service Sustainability". While there might be some overlap with activities reported in the first version of this report (D.5.4), the focus is on activities taking place in the time between January 2020 and March 2021.

Task 5.4 of Prêt-à-LLOD deals with raising awareness on the potentials of Linguistic Linked Open Data (LLOD) driven language resources and services, beyond the consortium and related initiatives. It is complementing the dissemination strategy of Prêt-à-LLOD, which is described in the deliverables reporting on the dissemination activities of the project (see D6.1 and D6.2: "Project Dissemination Report", v1 and v2).

A way to achieve this goal is to contribute to the wider application and the extension of existing standards, or even in the creation of new ones. Partners of Prêt-à-LLOD have intensively explored their relations to the relevant ISO and W3C committees. Further, discussions with the Text Encoding Initiative (TEI) community have been initiated and pursued in this period, also in cooperation with the ELEXIS project.[1] Further, Prêt-à-LLOD is actively contributing to the new OASIS standard "Lexicographic Infrastructure Data Model and API (LEXIDMA)". Attention has been given to the re-use and RDF transformation of ISO standards dealing with linguistic annotations, as this is described in Section 11. Such activities aim to establish a community extension across standardisation bodies. Prêt-à-LLOD standardisation activities were reported for the first year of the project in the deliverable D5.1 "Report on Vocabularies for Interoperable Language Resources and Services''. From year 2 on they are reported in the corresponding versions of the deliverable for task 5.4: D5.5 and D5.6.

A second type of activities consisted in organizing LLOD-specific hackathons or datathons and in contributing to spring and summer schools, complementing thus the T6.1 dissemination activities (see D6.1 and D6.2) with in-depth training activities. While the pandemic has caused the cancellation of various events, alternatives have been developed and the project can be considered as back on track and a relevant number of events are already in place for 2021, as documented in Section 3.

A third type of community building efforts consisted of presenting the methodologies and results of Prêt-à-LLOD in industry events, especially those relevant to the WP4 use cases. An industry event planned as a satellite event to the Semantics 2020 conference had to be cancelled, but industrial partners organized Webinars, which are described in Deliverable D6.2 (Section 2.2).
Finally, Prêt-à-LLOD is engaged in cooperation with the other projects resulting from the call "ICT-29-2018: A multilingual Next Generation Internet".

---

[1] See https://elex.is for more information.

The following sections will describe in detail the activities of the project reflecting the points made above.

2 Relations to the other projects resulting from the call "ICT-29-2018: A multilingual Next Generation Internet"

From the very beginning, Prêt-à-LLOD aimed to establish strong relations to the other projects funded in the context of the "ICT-29-2018: A Multilingual Next Generation Internet" call. Two types of projects were foreseen in this call: 1) an Innovation Action, called "A European Language Grid", corresponding to the call "29a", and 2) a Research and Innovation Action, called "Domain-specific/challenge-oriented Human Language Technology", corresponding to call "29b". We describe below briefly the main objectives of the projects that have been selected for this call.

One project was selected for the 29a call and the project has the name "European Language Grid" (ELG: see https://www.european-language-grid.eu/, or at the Cordis site: https://cordis.europa.eu/project/rcn/219378/factsheet/en). "The ELG will be a scalable cloud platform, providing, in an easy-to-integrate way, access to hundreds of commercial and non-commercial Language Technologies for all European languages, including running tools and services as well as data sets and resources. It will enable the commercial and non-commercial European LT community to deposit and upload their technologies and data sets into the ELG, to deploy them through the grid, and to connect with other resources." (quoted from https://cordis.europa.eu/project/rcn/219378/factsheet/en).

In the first period of the projects, regular telcos were organized among the 6 projects selected in the call 29b, also in cooperation with ELG. One of those 6 projects, EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media), has been recognized as particularly relevant to Prêt-à-LLOD. A specific cooperation will be established in integrating cross-lingual word embeddings, represented as dense real-valued vectors, with lexical data that are encoded in the OntoLex-Lemon model, based on a symbolic representation of lexical knowledge, but which in the context of the development of the FRaC module for OntoLex-Lemon (see the paragraph on FRaC in Section 12) is opening to also considering the symbolic encoding of real-valued vectors representation of linguistic data, and an in-depth (online) meeting will be organized in June or July 2021.

Behind this technological cooperation, the "ICT call 29" projects have addressed the topic of sustainability, aiming at establishing a joint document collecting all the sustainability strategies to be implemented in the various projects. Preliminary results of this initiative have been presented at the Meta-Forum 2019 event,[2] within a session dedicated to "Sustainability". This plenary session was jointly organized by Prêt-à-LLOD and ELG. Prêt-à-LLOD was additionally present at META-FORUM with a plenary presentation and a booth. A similar joint session on

---

[2] https://www.european-language-grid.eu/meta-forum-2019/

sustainability was organized at the (online) META-FORUM 2020.[3] Prêt-à-LLOD was also presented at this event with a poster.

Another community building cooperation has been established between Prêt-à-LLOD and ELG, with both projects contributing to the further development of the so-called Meta-Share Ontology (ms-ontology) dealing with the refinement and extension of the current metadata infrastructure used in the context of the past Meta-Share project and being further developed for ELG and other projects (see http://www.meta-share.org/p/93/Documentation#model, Gavrilidou et al. (2012)). Prêt-à-LLOD partners UNIZAR, NUIG, UNIBI und UPM are directly involved in this work.

3 Training Activities of Prêt-à-LLOD

3.1 Online Training School on "Introduction to Linked Data for Linguistics"

Under the umbrella of the 15th edition of the Eurolan School (http://eurolan.info.uaic.ro/2021/) and the NexusLinguarum COST Action (https://nexuslinguarum.eu/), Prêt-à-LLOD partners contributed with a series of lectures, hand-on sessions, and supervised exercises to this event, which took place online from February 8th to 12th. Partners involved were University of Zaragoza (UZAR), Goethe University Frankfurt (GU), National University of Ireland Galway (NUIG) and the German Research Centre for Artificial Intelligence (DFKI). With 3 contributors from the University of Rome and one from the Czech Technical University, a total of 12 lecturers/tutors were involved. The local organisation (setting up of the registration system, website, and teleconference systems) was carried out by University "Alexandru Ioan Cuza" Iaşi, Romania, and by the Romanian Academy. The representative of the University of Prague was also instrumental in setting up the Zoom framework so that enough "rooms" were available for group working during the exercise sessions.

The event was organized as follows: in the morning 2 theoretical lectures, followed by a hands-on session. The afternoon was dedicated to exercises, which take place in 10 different "rooms", with tutors "circulating" between the "rooms" and re-acting to calls for help.

The event had 99 registrations, and we counted a regular attendance of 70+ participants at the lectures. 40 to 50 participants joined the exercise sessions.

3.2 Contributions to Spring and Summer Schools

Thierry Declerck (DFKI) and John McCrae (NUIG) have been invited to give a lecture at the Spring School LiSeH, Graz, Austria, April 2020. The topic of the Spring School was "Linked Data and the Semantic Web for Humanities research". But the event was cancelled. The event will instead take place this year, April 6-9 in a virtual environment.[4]

---

[3] See session 7 at the third day of Meta-Forum: https://www.european-language-grid.eu/meta-forum-2020/meta-forum-2020-programme/

[4] https://informationsmodellierung.uni-graz.at/en/institute/events/spring-school-liseh-2021/

The same situation was encountered for an invitation to teach in July 2020 at the Lisbon Summer School on Linguistics, where John McCrae (NUIG) and Thierry Declerck (DFKI) were invited to teach on "Linked Open Data in Linguistics". The lectures will be given instead this year at the beginning of July, with Julia Bosque Gil (UNIZAR) and Thierry Declerck (DFKI) as lecturers.

A tutorial to be given by Gilles Sérasset (University of Grenoble, not a Prêt-à-LLOD partner) and Thierry Declerck (DFKI) has been accepted to the French national conference on NLP (to be held in June/July 2021 in Lille, https://talnrecital2021.inria.fr/en/)

4 Academic Conferences and Workshops organized or supported by Prêt-à-LLOD

4.1 The Language Data and Knowledge (LDK) Conference supported by Prêt-à-LLOD and co-located Events

Language, Data and Knowledge (LDK) is a bi-annual conference series on matters of human language technology, data science, and knowledge representation, initiated in 2017 by a consortium of researchers from two Prêt-à-LLOD partners (NUIG, GUF) and the University of Leipzig. It has been supported by an international Scientific Committee of leading researchers in Natural Language Processing, Linked Data and Semantic Web, Language Resources and Digital Humanities. The second edition of the LDK conference was hosted 2019 by the Institut für Angewandte Informatik (InfAI) in Leipzig (May 20-23), and co-organized by the two Prêt-à-LLOD partners NUIG and GUF.

LDK conferences aim at bringing together researchers from across disciplines concerned with the acquisition, curation and use of language data in the context of data science and knowledge-based applications. The efficient and meaningful exploitation of this data in scientific and commercial innovation is at the core of data science research, employing natural language processing and machine learning methods as well as semantic technologies and knowledge graphs.

The third edition of LDK is organized by UNIZAR, under the umbrella of the NexusLinguarum COST Action and with active support by Prêt-à-LLOD. Thierry Declerck (DFKI) and John McCrae (NUIG) are the co-chair of LDK this year. Webpage, call-for-papers, etc. have been finalized, and the conference will have 4 satellite workshops and two tutorials. Reviewing process is starting on the 1st of April (71 submissions are to be reviewed). The event will take place in Zaragoza, September 1st to September 3rd (2021.ldk-conf.org/). In addition, there will be a post conference day ("W3C day") on 4th September devoted to host two half-day meetings of the W3C Ontolex and W3C LD4LT community groups, respectively, where active participation by Prêt-à-LLOD members is expected.

4.2 The Translation Inference Across Dictionaries (TIAD 2020/2021) shared Task

The Prêt-à-LLOD partner UNIZAR (Jorge Gracia and team) organises the shared task for Translation Inference Across Dictionaries (TIAD) with the support of Prêt-à-LLOD. This initiative is aiming at exploring methods and techniques for automatically generating new bilingual (and multilingual) dictionaries from existing ones, by supporting a coherent experiment framework that enables reliable validation of results and solid comparison of the processes used.[5] The participant teams submit a system description paper including: a description of their system, the way data was processed, the applied algorithms, the obtained results, as well as the conclusions and ideas for future improvements. Other works from non-participant teams can be submitted as regular papers describing any technique or approach that is relevant for the topic of the workshop.

Two editions of TIAD have been organised so far within the duration of Prêt-à-LLOD: on May 20, 2019, collocated with LDK'19 in Leipzig, Germany (https://tiad2019.unizar.es/), and on Tuesday, May 12 2020 collocated with LREC 2020 in Marseille, France (https://tiad2020.unizar.es/). Unfortunately, LREC 2020 was cancelled owing to the COVID-19 crisis and did not take place in Marseille as expected, nor was it replaced by an online event. However, the TIAD 2020 campaign continued as planned, the evaluation results were communicated, and the system papers were published jointly with the Globalex workshop proceedings. As for the physical (or virtual) authors' presentation of such papers to the community, the authors will have the opportunity of presenting them in the next TIAD edition. In fact, a new edition of TIAD is planned for LDK 2021 in Zaragoza on 1st September 2021 (http://tiad2021.unizar.es/).

A description of the participant systems and the overall results can be found in Gracia et al (2019) and Kernerman et al. (2020).

4.3 7th Workshop on Linked Data in Linguistics: Building tools and infrastructures

Prêt-à-LLOD co-organized the 7th Workshop on Linked Data in Linguistics: Building tools and infrastructures (http://ldl2020.linguistic-lod.org/), as a satellite event to LREC 2020. Since the conference was cancelled, the workshop took place later, in June 2020, as an online event, with 90 registrations. Proceedings are published and available in the ACL Anthology.

4.4 1st international Workshop on Multimodal Wordnets

Thierry Declerck (DFKI) co-organized this event (http://www.hitz.eus/multimodalwordnets2020/), which was co-located with LREC 2020, but as LREC was cancelled, the workshop generated only proceedings. The workshop was organized as part of the community building efforts, targeting the WordNet community at large. The interest of Prêt-à-LLOD for the topic of multimodality is also present in the discussion

---

[5] See http://tiad2021.unizar.es for more details.

within the FRaC extension module proposed to OntoLex-Lemon (see the paragraph on FRaC in Section 12).

4.5 6th Semantic Deep Learning (SemDeep-6) Workshop and the Word In Context (WiC) Challenge

Thierry Declerck (DFKI) was co-organizing the 6th Semantic Deep Learning (SemDeep-6), co-located with IJCAI 2020. The workshop was postponed to January 2021 and took place as a virtual event workshop. http://www.dfki.de/semdeep-6/index.html
In the same context, Artem Revenko (SWC) co-organized the Word In a Context challenge (https://www.dfki.de/semdeep-6/challenge.html), which attracted good submissions.

4.6 W3C OntoLex Community Group Meetings

Partners of the Prêt-à-LLOD consortium are organizing regular tele-conferences, in which two extensions to the OntoLex model are discussed: the "Morphology"[6] and the "Frequency, Attestation and Corpus Information"[7] modules. In addition, a series of modelling issues encountered during the conversion of terminological resources and reported by Prêt-à-LLOD partners in the W3C group will serve as basis for the discussion and consideration of a potential new module on Terminology. Also in relation to representation needs, several Prêt-à-LLOD partners, jointly with other members of the NexusLinguarum COST Action and the OntoLex community group in general, are involved in a new series of teleconferences focused on the representation of multimodal information in linguistic resources.
Prêt-à-LLOD partners are organizing a new OntoLex workshop at LDK 2021. The tentative agenda for this half-day workshop includes a discussion on the current modules' under-development in OntoLex and potential new modules, as well as an exchange on use cases and limitations encountered in the use of OntoLex, some of which might require a revision of the specification guidelines. The relationship and scope boundaries of the community group in relation to other initiatives led by Prêt-à-LLOD partners (e.g., LD4LT Linguistic Annotations) is also a potential aspect to address during the meeting.

4.7 TEI-OntoLex Workshop

In collaboration with the ELEXIS H2020 infrastructure project, a two-day workshop connecting the main developers of the OntoLex-Lemon model and the TEI Lex-0 model was held on the 4th and 5th of October in Sintra, Portugal after the eLex 2019 conference. As a result, an authoritative tool for converting from OntoLex to TEI-Lex0 was released on GitHub at https://github.com/elexis-eu/tei2ontolex. Continued discussion between the TEI-Lex0 community and the OntoLex community is ongoing.

---

[6] See https://www.w3.org/community/ontolex/wiki/Morphology for more details.
[7] See https://acoli-repo.github.io/ontolex-frac/ for more details.

5 Relations of Prêt-à-LLOD to the Lexicography Community

The relation to the European Lexicographic Infrastructure (ELEXIS) project (https://elex.is) is not only relevant due to the Lexicography use case of Prêt-à-LLOD (led by the partner Oxford University Press (OUP)), but also to the fact that Linked Data are playing an increasing role in eLexicography, and as such the OntoLex-Lemon model, which is at the core of Prêt-à-LLOD, is getting more and more used in this context (McCrae et al., 2019; Bosque-Gil et al., 2019; Declerck et al., 2017 or Stolk, 2019). While OUP has strong relations with some partners in ELEXIS, NUIG is also a partner of this project and Thierry Declerck (DFKI) is involved as an external expert in ELEXIS. This connection to ELEXIS is important for Prêt-à-LLOD, as an increasing community of lexicographers are making use of OntoLex-Lemon and other LLOD technologies, thus ensuring sustainability of methods developed by the main technological partners of Prêt-LLOD, but also giving feedback on certain issues.

Cooperation with ELEXIS is also concerned with the discussion on standards, for example establishing bridges between OntoLex-Lemon,[8] as the result of a W3C Community Group, and the TEI Lex-0 encoding guidelines in development within the TEI community.[9]

Prêt-à-LLOD partners submitted successfully to the Euralex 2020 conference (to be held in September 2021, due to the pandemic). And submissions by Prêt-à-LLOD partners have been accepted for eLex 2021.

Additionally, the Prêt-à-LLOD partner NUIG is a member of the OASIS standardisation community and is contributing to the new standard being developed by this working group. This group is developing open standards based framework for internationally interoperable lexicographic work. They are developing a simple, modular, and easy to adopt data model that will be attractive for all lexicographic industry actors across companies and academia as well as geographic locations. Adoption of that model will facilitate exchange of lexicographic and linguistic corpus data globally and enable effective exchange with adjacent industries such as language services, terminology management, or technical writing.[10]

6 Relations of Prêt-à-LLOD to the WordNet Community

The WordNet community is an important group of users and producers of language data, delivering so-called synsets that are a representation of lexical semantics properties associated with words. Prêt-à-LLOD is aiming at increasing the representation of WordNets data in the formats at the core of the Linguistic Linked Open Data cloud. John P. McCrae (NUIG) is a board member of the Global WordNet Association (GWN)[11] and Thierry Declerck

---

[8] See https://www.w3.org/2016/05/ontolex/ for the full specification of the model.

[9] TEI stands for Text Encoding Initiative (https://tei-c.org/). For TEI Lex-0, which is used more in the DARIAH community, see https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html.

[10] For more details see https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=lexidma

[11] See http://globalwordnet.org.

(DFKI) is chairing the GWN working group on Multimodal Wordnets.[12] Both John McCrae and Thierry Declerck (DFKI) attended the (online) Global WordNet Conference in January 2021 presenting different works on porting WordNets in the OntoLex-Lemon model (John McCrae et al. 2021); Thierry Declerck, Lenka Bajčetić, 2021). NUIG is actively involved in the work of the WordNet community, firstly through the development of the Collaborative Interlingual Index[13] and secondly through the development of the English WordNet.[14]

As such, Prêt-à-LLOD project aims have been disseminated through this community and the members of the community are active stakeholders in developing linguistic linked open data. Within the WP3 "Transforming, Linking and Workflows for Language Resources" of Prêt-à-LLOD active work is dedicated to the transformation of data included in the Open Multilingual WordNet initiative,[15] as this can be read in Deliverable 3.1 "Research Challenges Report v1" (section 2.2 Case study: Open Multilingual Wordnet in TSV).

7 Relations of Prêt-à-LLOD to the Terminology Community

Multilingual terminologies are an important component of the use cases defined for Prêt-à-LLOD. To include such terminologies in the Linked Data based workflows of the use cases, the project is also making use of the OntoLex-Lemon model. This model supports the use of the SKOS[16] vocabulary for linking terms listed in terminological bases to a full lexical description to be encoded in the corresponding elements of OntoLex-Lemon. The cooperation with the terminology community is therefore an important aspect of the community building activities of the project.

One aspect concerns the presentation of Prêt-à-LLOD activities in conferences having a focus on terminology. This was the case for example again the 2020 edition of the TOTh conference where DFKI presented work on how to interlink terminology and lexicography data, but also on how to express lexical restriction for terms.[17] Papers dealing with our work on terminology were also successfully submitted at LREC and other conferences. And the successful integration in the ELG Platform of the Term-a-LLOD transformer by the University of Bielefeld was also well-received and documented in (di Buono et al, 2020)

In the context of the development of extensions of current models for phenomena not sufficiently covered, UPM is working together with DFKI to develop an extension to Ontolex-Lemon to represent terminologies generated from heterogeneous data sources. We have already published a draft in the wiki of the W3C community group: https://www.w3.org/community/ontolex/wiki/Terminology. However, discussions are still open, and the extension needs to be presented and agreed by the Ontolex community.

---

[12] See http://globalwordnet.org/resources/working-groups/.

[13] https://github.com/globalwordnet/ili

[14] https://github.com/globalwordnet/english-wordnet

[15] See http://compling.hss.ntu.edu.sg/omw/ for more details.

[16] SKOS stands for "Simple Knowledge Organization System" and is being used by many terminologies, taxonomies and thesauri. See https://www.w3.org/2004/02/skos/ for more details.

[17] http://toth.condillac.org/toth-2020

8 Relations of Prêt-à-LLOD to European Initiatives

John McCrae (NUIG) was an invited speaker and presented Prêt-à-LLOD at ENDORSE 2021 -- the European Data Conference on Reference Data and Semantics --, organized by the Publications Office of the EU, which took place March 16-19.

9 Industry Events

As stated in the introductory section, Prêt-à-LLOD aims to present its methodologies and results at industry events. For this, the project is also committed to organize an "Innovation event" per year. The industrial event planned in 2020, as a satellite event to the Semantics 2020 conference had to be cancelled. But industrial partners of Prêt-à-LLOD organized various Webinars, which are documented in D6.2.

10 The second Sustainability Event of Prêt-à-LLOD

Prêt-à-LLOD has defined a four levels strategy for ensuring sustainability: (1) through the form and representation of data, (2) through a fine-grained provenance, legal information, and smart policies for language data transactions, (3) through accessibility and repository management, and (4) at a community level.

The project started to talk with other projects to see which sustainability strategies are shared or can be considered as complementary, leading to an enlargement of the overall strategy for ensuring sustainability for language technologies. This first round of discussion, in the first year of the project, was leading to the joint organisation by Prêt-à-LLOD and ELG of the session "Sustainability of the ICT-29 Projects" at META-FORUM 2019 dedicated to the sustainability approaches followed by the projects resulting from the call "ICT-29-2018: A multilingual Next Generation Internet". There, John McCrae (NUIG) gave a talk jointly prepared by NUIG and DFKI on "Overview of the Sustainability Plans of the ICT-29b", while Georg Rehm (coordinator of ELG) presented the sustainability plans of ELG leading to the creation of a legal entity. The session, which was attended by many META-FORUM participants, has been very well received.

It was foreseen to have a sustainability meeting at LREC 2020, with representatives of related projects, also in cooperation with the LREC Worksop "LT4Gov", in May 2020, but the events were canceled. Instead, we went again for a joint session with the ELG project at MetaForum 2020,[18] which was held on-line December 1st to December 3rd. There Thierry Declerck (DFKI) gave an overview on the ways Prêt-à-LLOD and related ICT projects are making their data available to the ELG infrastructure, encapsulating those in containers. Prêt-à-LLOD also presented its view on sustainability in the context of iLOD: InterPlanetary File System based Linked Open Data Cloud (Nasir and McCrae, 2020)

---

[18] See https://www.european-language-grid.eu/meta-forum-2020/meta-forum-2020-programme/ (Day 3)

11 The Linked Data for Language Technologies (LD4LT) Community Group Report

The Linked Data for Language Technologies (LD4LT) is a community group that was founded in the previous FP7 project "LIDER" and that is used for the broader discussion of issues related to linked data and its applications in NLP. The group has fallen quite silent in the period after the end of the LIDER project and as a result we asked the former chairs of LD4LT if they would agree to have Prêt-à-LLOD members to be the new chairs of this group and direct it from a Prêt-à-LLOD project view. As such, we are now using this group to discuss some public activities, such as the META-SHARE OWL ontology[19]. In this reporting period, a focus has been in establishing contacts with relevant past and present initiatives in the ISO standardization communities, and work has started, for example, in porting the ISO SynAF standard onto OntoLex-Lemon, using also OLiA and LexInfo vocabularies, as well as insights from the W3C Web Annotation data model (https://www.w3.org/TR/annotation-model/).

12 Prêt-à-LLOD Standardisation Activities

The use of and the contribution to standards is a way to ensure sustainability of results (algorithms, data, formats, guidelines, etc.) delivered by a project. Prêt-à-LLOD is particularly involved in such standardisation activities, having as a focus the further development of modules for the core model OntoLex-Lemon, developed in the context of a W3C Community Group (https://www.w3.org/2016/05/ontolex/). The former version of this deliverable was just referring to the deliverable D5.1 "Report on Vocabularies for Interoperable Language Resources and Services", in which the relevant existing standards are described in detail, but also the activities that had been already initiated by partners of Prêt-à-LLOD. As there is no update of D5.1 to be delivered, further standardisation activities by Prêt-à-LLOD are described in this section of this deliverable. We list here the main activities:

- Ontolex Module for Morphology: Many academic partners of Prêt-à-LLOD are contributing to the Ontolex module for Morphology, which is now in an advanced stage so that we can expect it to be published in the lifetime of the project. The current state of the module development (and the minutes of telcos) is available at https://www.w3.org/community/ontolex/wiki/Morphology.
- OntoLex Module for Frequency, Attestations and Corpus Information (OntoLex-Frac): GUF is leading this standardisation activities, which aims to (1) extend OntoLex with corpus in-formation to address challenges in lexicography, (2) model lexical and distributional-semantic re-sources (dictionaries, embeddings) as RDF graphs, (3) provide an abstract model of relevant concepts in distributional semantics that facilitates applications that integrate both lexical and distributional information. The current state of development of the FRaC module can be seen at: https://github.com/ontolex/frequency-attestation-corpus-information/blob/master/index.md.
- Candidate for a OntoLex terminology module: UPM is working together with DFKI to develop an extension to Ontolex-Lemon to represent terminologies generated from

---

[19] https://github.com/ld4lt/metashare/tree/dev

heterogeneous data sources. We have already published a draft in the wiki of the W3C community group: https://www.w3.org/community/ontolex/wiki/Terminology.

- LexInfo 3.0: A new version of the LexInfo ontology has been published, under the direction of NUIG. This version is available at https://github.com/ontolex/lexinfo/blob/master/ontology/3.0/lexinfo.owl.
- META-SHARE metadata ontology: As already mentioned, Prêt-à-LLOD partners are contributing to the finalisation of the META-SHARE metadata ontology. The current state of development is documented at: http://www.meta-share.org/p/93/Documentation#model.
- Prêt-à-LLOD was extensively presented at a DFKI internal W3C Workshop (10.03), as DFKI is leading the W3C offices for Austria and Germany.
- UPM continues its contributions to the Open Digital Rights Language (ODRL) W3C Community Group and implements aspects of Task 5.2 "Policy-driven language resource discovery and access" in the new version (2.2) of ODRL.

13 Contributions to the new COST Action CA18209 - European network for Web-centred linguistic data science ("NexusLinguarum")

We quote first from the Action's web page (see https://www.cost.eu/actions/CA18209) "The main aim of this Action is to promote synergies across Europe between linguists, computer scientists, terminologists, and other stakeholders in industry and society, in order to investigate and extend the area of **linguistic data science**. [...] In order to support the study of linguistic data science in the most efficient and productive way, the construction of a mature holistic **ecosystem of multilingual and semantically interoperable linguistic data** is required at **Web** scale. Such an ecosystem, unavailable today, is needed to foster the systematic cross-lingual discovery, exploration, exploitation, extension, curation and quality control of linguistic data. We argue that linked data (LD) technologies, in combination with natural language processing (NLP) techniques and multilingual language resources (LRs) (bilingual dictionaries, multilingual corpora, terminologies, etc.), have the potential to enable such an ecosystem that will allow for transparent information flow across linguistic data sources in multiple languages, by addressing the semantic interoperability problem."

Therefore, the role of LD to support NLP is central in this Action, as it is in Prêt-à-LLOD. In fact, there are two working groups in the network that explicitly studies this: WG 1 - Linked data-based language resources and WG 2 - Linked data-aware NLP services. NexusLinguarum (as any other COST Action) will not receive funds for research but for networking activities. Through such networking activities (research visits, conferences/workshops organisation, scientific meetings, training schools, etc.), the Prêt-à-LLOD participants enter in connection with a high number of other researchers and practitioners all across Europe. This constitutes an invaluable opportunity to further exchange ideas, discover new datasets of interest to the project (particularly for minority languages), as well as to discover application scenarios to the outcomes of Prêt-à-LLOD. The joint organisation of several events between NexusLinguarum and Prêt-à-LLOD is also taking place, as for example the Eurolan21 training school or the 3rd LDK conference, already

mentioned in the sections on Training Activities (Section 3) and on the Organisation of Conferences and Workshops (Section 4).

14 Conclusion

In this deliverable we reported on the community building strategy of Prêt-à-LLOD and listed the various actions taken by the project to extend the community of potential users and contributors to the Linguistic Linked Open Data infrastructure. We described the various communities with which we established or enhanced cooperations.

We also reported on the first Innovation and Sustainability events organized by Prêt-à-LLOD.

Together with the report D5.1 and the report on Dissemination activities (D6.1.1), we see the project in a very good way for reaching the goals specified in Work Packages 5 and 6.

15 References

Julia Bosque-Gil, Dorielle Lonke, Jorge Gracia, Ilan Kernermann. Validating the OntoLex-lemon Lexicography Module with K Dictionaries' Multilingual Data. Proceedings of Sixth Biennial Conference on Electronic Lexicography, eLex 2019, (2019).

Maria Pia di Buono, Philipp Cimiano, Mohammad Fazleh Elahi, Frank Grimm. Terme-à-LLOD: Simplifying the Conversion and Hosting of Terminological Resources as Linked Data. In: Proceedings of the 7th Workshop on Linked Data in Linguistics. 2020.

Thierry Declerck, Carole Tiberius, Eveline Wandl-Vogt Encoding lexicographic Data in lemon: Lessons learned. In: Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets, Galway, Ireland, CEURS, 8/2017

Thierry Declerck, Lenka Bajčetić. Towards the Addition of Pronunciation Information to Lexical Semantic Resources. Proceedings of the 11th GlobalWordnet Conference (2021)

Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papageorgiou, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz, Valerie Mapelli. (2012) The META-SHARE Metadata Schema for the Description of Language Resources. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012).

J. Gracia, B. Kabashi, and I. Kernerman (Eds.): *Proceedings of "TIAD-2019 Shared Task – Translation Inference Across Dictionaries*" co-located with the 2nd Language, Data and Knowledge Conference (LDK 2019). Leipzig, Germany, May 20, 2019. See http://ceur-ws.org/Vol-2493/.
I. Kernerman, S. Krek, J. P. McCrae, J. Gracia, S. Ahmadi, and B. Kabashi (Eds.), Proceedings of Globalex 2020 Workshop on Linked Lexicography. ELRA, 2020. See https://www.aclweb.org/anthology/volumes/2020.globalex-1/

John P. McCrae, Alexandre Rademaker, Francis Bond, Ewa Rudnicka and Christiane Fellbaum English WordNet 2019 — An Open-Source WordNet for English. Proceedings of the 10th Global WordNet Conference, (2019).

John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka and Luis Morgado Da Costa. The GlobalWordNet Formats: Updates for 2020. Proceedings of the 10th Global WordNet Conference, (2021).

Jamal A. Nasir, John McCrae. iLOD: InterPlanetary File System based Linked Open Data Cloud. In: Proceedings of MEPDaW'20 - Managing the Evolution and Preservation of the Data Web. 2020

Sander Stolk. A Thesaurus of Old English as Linguistic Linked Data: Using OntoLex, SKOS and lemon tree to Bring Topical Thesauri to the Semantic Web. Proceedings of Sixth Biennial Conference on Electronic Lexicography, eLex 2019, (2019).