Adopting Linguistic Linked Data Principles: Insights on Users' Experience

Verginica Barbu Mititelu
Romanian Academy, Research Institute for Artificial Intelligence
vergi@racai.ro

Maria Pia di Buono
Univ. of Naples "L'Orientale"
Napoli, Italy
mpdibuono@unior.it

Hugo Gonçalo Oliveira
Univ. of Coimbra, Portugal
CISUC, DEI
hroliv@dei.uc.pt

Blerina Spahiu
University of Milan-Bicocca, Italy
blerina.spahiu@unimib.it

Giedrė Valūnaitė Oleškevičienė
Mykolas Romeris university, Lithuania
gvalunaite@mruni.eu

Abstract

Despite the advantages, Linguistic Linked Data (LLD) best practices and principles seem to be far away from being widely adopted. Such a situation can be related to existing challenges in the creation, reusing, and exposing of LLD resources. In this paper, we present the results of a survey which investigated users' perspective and experience in the use and application of LLD principles, to evaluate the impact, prospects, requirements, or challenges encountered in LLD adoption. The survey was organized in several sections to collect information about participants' background, LLD knowledge, use, development, publishing, and metadata use. The results show that some bounds have to be overstepped to ensure the penetration of LLD principles in a wider community and fully exploit their potential.

1 Introduction

Linguistic Linked Data (LLD) best practices and principles aim at describing language resources and conveying useful linguistic information about them, allowing linking among resources, interoperability across datasets and systems, as well as their federation (Chiarcos et al., 2020).

Despite their advantages, including for under-resourced languages (Bosque-Gil et al., 2022), LLD best practices and principles seem to be far from being widely adopted. Such a situation can be related to some challenges in the creation, reusing, and exposing of LLD. In this paper, we present the results of a survey, conducted within the COST Action "CA18209 - European network for Web-centred linguistic data science" (Nexus Linguarum, NL CA), Working Group (WG) 1 - Task 1.2 in collaboration with Tasks 1.4 and 1.5, which investigated the users' perspective and experience in the use and application of LLD principles, in order to evaluate the impact, prospects, requirements, or challenges encountered in LLD adoption.

Such an evaluation complements another survey carried out within NL CA (Khan et al., 2022), as it offers another (i.e., the (potential) user's) perspective on the adoption of LLD and could be of interest not only to other WGs within NL CA, but also to other stakeholders, including people and categories involved in European initiatives and projects, such as the European Language Grid[1] and the Prêt-à-LLOD[2] projects.

The paper is organized as follows: in Section 2 we report on related work; in Section 3 we describe the survey aims and structure, while in Section 4 we present the results. Section 5 is devoted to discussing some of our findings and, finally, in Section 6 we conclude and envisage future work.

2 Related Work

LLD is known to offer numerous advantages and opportunities. Lezcano et al. (2013) observed that the simple syntactic model of RDF, which allows organizing structured data into a set of simple triples, makes linguistic data suitable for carrying out tasks combining data from different sources. Also, as Linked Data (LD) is comparatively straightforward, data discovery and harvesting become an accessible task for performing without full knowledge of the data structure. While discussing their survey, Lezcano et al. (2013) pointed out that RDF requires a standardized representation of the annotation semantics. The authors identified some legal and economic issues concerning

[1] https://live.european-language-grid.eu/

[2] https://pret-a-llod.github.io/

copyrighting and pricing of Language Resources (LR), that act as barriers to LR interoperability and propose that the adoption of LLD approaches to LR exchange may have a positive impact on these matters. They also identified an open issue – the development of mechanisms and knowledge to support the alignment of different features and aspects of LRs which allow for ensuring semantic and conceptual interoperability in the LOD cloud[3]. Some other areas of LLD are to be considered for improvement concerning the languages covered and types of linguistic datasets presented in the LOD cloud.

Geddes (2019) acknowledged that LLD provides the opportunity to use the data freely and connect the data to other existing data; however, the focus on the user, user's needs and capacities is of key importance in the process of sustaining a healthy data ecosystem. As LLD technologies facilitate information integration and interoperability, they require making the entities addressed in an unambiguous way, so that they could be accessed and interpreted. Also, it should be ensured that entities associated on a conceptual level are physically associated with each other as well.

The LLD applications reveal the potential of the technology in linguistics, but there is still a considerable barrier for linguists who are not advanced users of RDF and related technologies. Since the early days of the Semantic Web, the "cognitive overhead" of learning RDF and related technologies was pointed out as an obstacle to its adoption by a broader community (Marshall and Shipman, 2003). This identifies the necessity of the technology to achieve a certain level of user-friendliness suitable for its non-advanced users (Chiarcos et al., 2020).

An overview of the existing guidelines and best practices in LLD development, interlinking, publication, and validation was given by the data collection carried out as part of the survey on LLD models (Khan et al., 2022) performed as part of Task 1.1 of the NL CA. The process included the compilation of a survey of LLD-relevant projects and other relevant initiatives (i.e. W3C community groups). Khan et al. (2022) identified that the advantages of LLD and the numerous opportunities it offers as a means of publishing linguistic data require a certain level of technical appreciation of the Semantic Web, of RDF and other formalisms as well as a number of other technologies. In order to increase the uptake of LLD amongst non-specialists, it is important to make sure the available materials are made accessible to non-specialists and provide clear instructions and ways of doing common tasks which could be ensured by Guidelines (GLs) and Best Practices (BPs). The authors provided a list of the areas for improvement for LLD GLs/BPs supported by the experience of the authors, consumers, and compilers of the documents:

- access to documents should be provided to speakers of more (ideally any) languages, not only English;
- the documents should be easily findable and freely accessible;
- the documents should be clear and self-contained;
- the documents should be designed for different levels of expertise and for covering at least the types of resources listed in the LLOD cloud and the four tasks (generation, interlinking, publication, and validation);
- the documents should refer to existing tools that can be integrated into the workflow;
- the documents should be regularly updated with the latest technology/models/tools.

The provided list of important areas helps to evaluate the already existing materials and the trends of use which we have found in the survey, as well as to suggest the directions to prioritize in the process of producing new materials.

3 Survey

With the aim of identifying potential obstacles preventing (potential) users from adopting LLD principles, we developed a survey, whose structure is rendered in Figure 1, to collect information about participants' background, LLD knowledge/use, development, publishing, and metadata use.
The insights coming from the survey results are relevant for:

- the penetration of LLD, especially among linguists and language professionals/experts;
- the causes preventing potential contributors/users from applying/(re)using LLD principles/resources;

[3] https://linguistic-lod.org/llod-cloud

- the causes preventing potential developers from creating LLD resources or converting resources to LD format, as well as from publishing them;
- highlighting possible limitations of LLD resources/technologies (including current vocabularies);
- the extension/integration of vocabularies and models suitable to describe different linguistic information and language phenomena;
- the extent to which metadata are used to describe resources, as well as the user's preference with respect to their type.

The survey was open from July 2021 to February 2022, with two main calls for participation, distributed through social media, i.e., Twitter, and mailing lists, e.g., Corpora list, NL CA mailing list, and personal contacts. The total number of responses is 84, received from different participants.

4 Results

We present here the results of the survey with respect to the four major lines of interest (LLD use, development, publishing and metadata), as shown in Figure 1.

The survey reached both witting and unwitting researchers in LLD. From the former group, there were 58 participants (≈69%) to the survey, while from the latter there were 26 participants (≈31%). The results presented below are based on the responses provided by the 58 participants, because, as can be seen in Figure 1, the other 26 did not answer the questions related to LLD experience.

The distribution of the 58 participants according to their declared background is shown in Figure 2, where we notice that this distribution is quite balanced.

4.1 Use

Although aware of LLD, about one third (19) of the 58 participants never used LLD.

When inspecting the reasons provided[4] for not using LLD resources, we see that the main one is that the tools and resources they work with do not support this format (≈50%). Two other reasons are that they did not find a useful resource (≈37%) and they were not familiar with LLD (≈32%). To some extent, these are all related: i.e., for someone not familiar with LLD, even if they do not assume it, it will be harder to find useful resources. This relation may also explain why no participant gave both reasons.

Reasons like the lack of documentation (2, ≈10%), and, consequently, not knowing how to access this data (1, ≈5%) were also given. Both participants that refer to the lack of documentation also answer that they did not find a useful resource that fits their needs. The lack of documentation seems to be an obstacle to the adoption of LLD resources and technologies. As recently highlighted by Khan et al. (2022), there are not enough materials available fulfilling the role of guidelines and best practices for LLD, and, moreover, a lot of what exists has not been updated for years, thus being unable to reflect the latest developments in the field.

Another relevant reason for not using LLD resources was that the dump or SPARQL endpoint of a resource they were interested in was not working (≈20%). This is not surprising: di Buono et al. (2022) recently noted that in the metadata of the 136 linguistic datasets in the LLOD Cloud, only 41 included a SPARQL endpoint and none included the URL of their dump. This is more related to the maintenance of LLD, which can be quite complex for the creators of this kind of data. The fact that many resources listed in the LLOD Cloud and other hubs are not accessible is definitely not good advertising for LLD, and may push potentially interested users away. Together with the lack of documentation, this contributes to one last reason: not understanding the advantages of LD over other formats (2, ≈10%). Both of these participants also say that they are not familiar with LLD and SPARQL. All the above mentioned reasons for not using LLD resources are presented in Table 1.

The long discussed "cognitive overhead" of learning underlying technologies (Marshall and Shipman, 2003) plays a role here, i.e., it requires time to become familiar with technologies like RDF and SPARQL.

All 58 participants were asked about the conditions (one or more) under which they would use LLD resources more frequently (see Table 2). Among the multiple choices, 30 (≈52%) highlight the need for more documentation to help

[4]The participants could provide more than one reason for not using LLD resources.

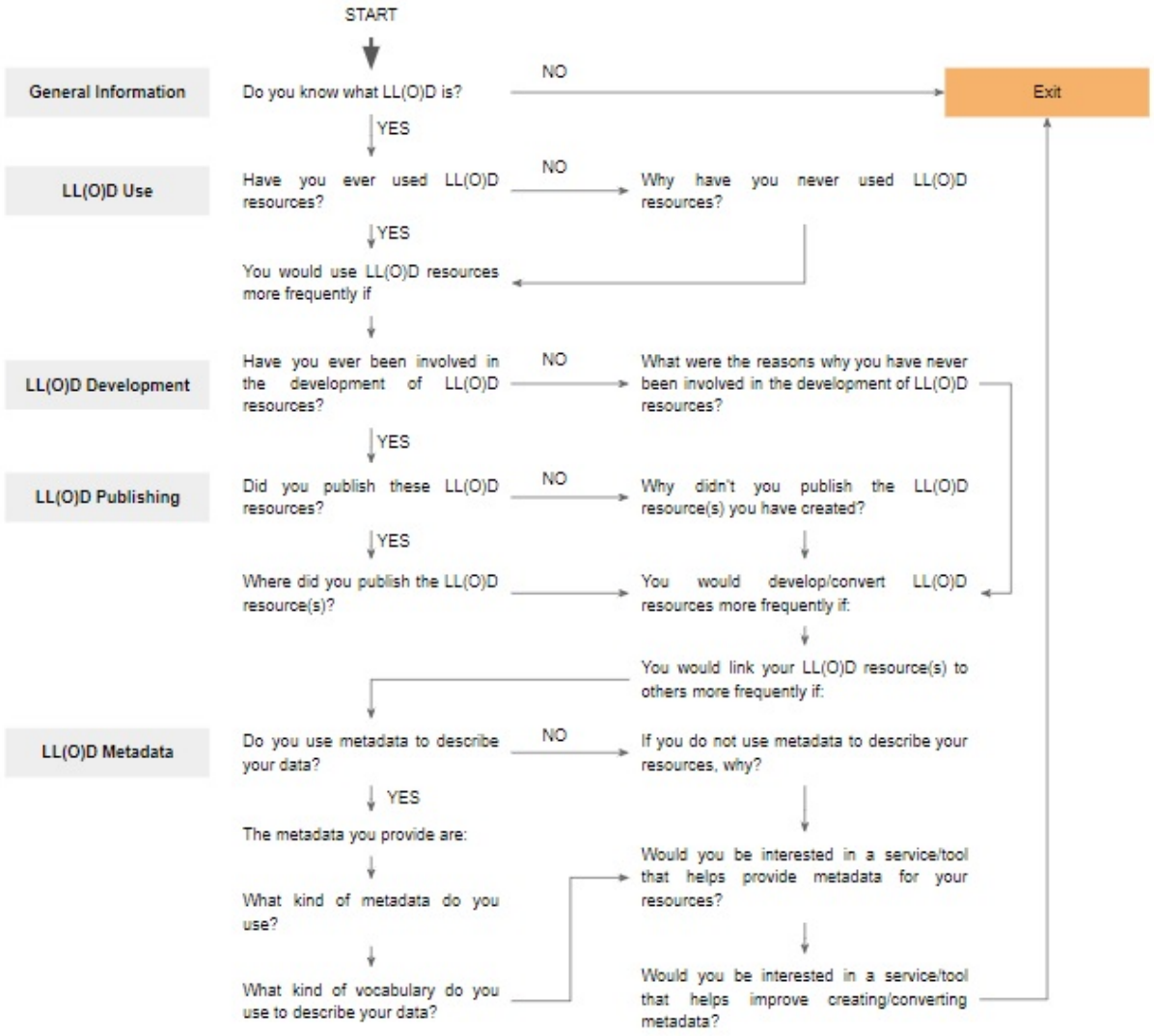

Figure 1: Diagram of survey flow. Some questions in the General Information section have been omitted given space constraints.

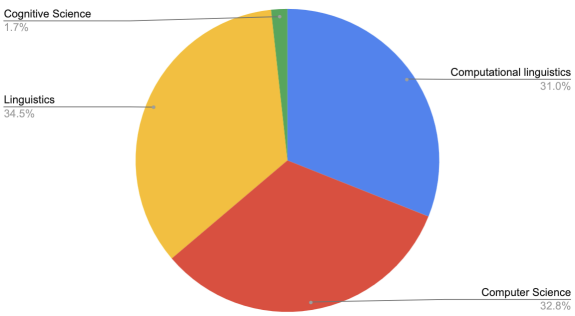

Figure 2: The distribution of the 58 LLD-aware participants to the survey according to their declared background.

Table 1: Reasons for never using LL(O)D resources

Reason	#
The tools/resources I work with do not support this format	9
I didn't find a useful resource that fits my needs	7
I am not familiar with the LLD models/SPARQL	6
The dump/SPARQL endpoint of the resource I was interested in was unavailable	4
I don't understand the advantages of Linguistic Linked (Open) Data resources over other formats (e.g. CoNLL-U)	2
I don't know much on how to access them	1
The documentation for the resource I was interested in was missing	2

them using LD resources, and 23 (≈40%) say that they would need more documentation about the resources they would potentially use. Moreover, 38 (≈66%) and 29 (50%) participants, respectively, selected the availability of tools/services suitable to use and discover LLD resources.

Table 2: Conditions under which paricipants to the survey would use LL(O)D resources more frequently

Condition	#
You were aware of a user-friendly service/tool to help you use LD resources	38
You were aware of more informative documentation to help you use LD resources	30
You were aware of a user-friendly service/tool to help you discover LD resources	30
The resources I would potentially use had (better) documentation	23
Other	5

4.2 Development

The shares of participants that develop resources in LLD format and of those who do not are almost equal, with a slight dominance of the former: ≈51% of the participants are also developers of LLD resources, while ≈48% do not develop them.

More than one reason could be provided for not developing LD resources and the answers given are summarized in Table 3: incompatibility with other tools or resources is the reason invoked by half of the respondents, 21% of all participants mentioned the lack of knowledge about the appropriate model or vocabulary for the resource under focus, while 17% of them complain about the inability of models to model data thoroughly.

Table 3: Reasons for not developing LD resources. % is calculated from the total number of reasons provided.

Reason	#	%
incompatibility with other tools/resources used	14	50
lacking knowledge about adequate model/vocabulary	6	21
models not totally appropriate for representing data	5	18
unclear example or guidelines	8	29
unclear advantages that LD has over other formats	3	11
Other	4	14

4.3 Publishing

Developing LLD resources does not necessarily imply their publishing. According to the results of this survey, only 57% of these resources get published. Figure 3 shows this publication tendency per different types of resources (as classified in the LLOD Cloud): we notice that the few typological databases developed have also been published, two-thirds of the terminologies, thesauri, and databases have been so, only a little more than half of the other types of resources have been published, and less than half of the linguistic resource metadata have been published.

Participants who responded positively to the development of resources (30 respondents, i.e. ≈52%) were then asked to answer about publishing/exposing such resources and only 23 (≈77%) of them published the resources, mainly in local repositories (15 people, i.e., ≈48%) and in the LLOD cloud (8 people, ≈26%). Considering other infrastructures/repositories for linguistic resources/language technologies, ≈17% of the respondents (4) published their resource in CLARIN and only ≈9% (2) in ELG. We note that none of the respondents used META-SHARE to publish their resources – see Figure 4.

With reference to the reasons preventing publishing resources, copyright policies were the main one, as invoked by ≈57% of the respondents not publishing the developed LL(O)D resources. The lack of knowledge about how/where to publish these resources, the cost/effort needed to pub-

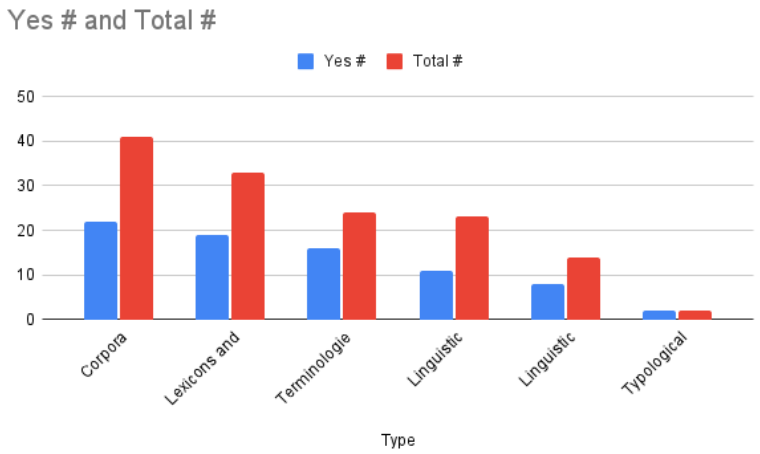

Figure 3: The proportion of users who are only developers of LD resources (in red) and those who are both developers and publishers of LD resources (in blue), for each type of LD resource.

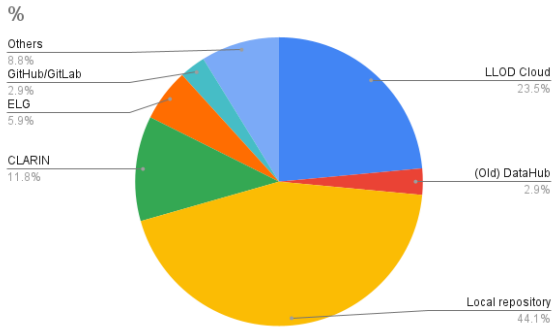

Figure 4: Repositories for publishing LD resources.

lish/maintain the resources, and the lack of motivation have been equally given reasons (≈14% each).

4.4 Metadata

Metadata allows people to organize data in such a way that is meaningful to other people while making their findability easier (Zuiderwijk et al., 2012; Schmachtenberg et al., 2014). It is also a way of keeping the data consistent and enabling decisions in data handling (Spahiu et al., 2019). There are thus many advantages to producing and maintaining metadata.

In fact, 52 people (≈90%) confirmed that they do use metadata to describe their data. On the other hand, 6 (≈10%) participants do not use metadata.

The most shared reasons for not using metadata for describing the data are summarized in Figure 5. They are: (i) task consuming task; (ii) manual effort is required; and (iii) there is a lack of harmonization among metadata models. Only one user mentioned that the reason why they do not use metadata is that they have difficulties finding the right model.

Understanding and interpreting LLD is difficult as information about the context of the data is often missing (di Buono et al., 2022). Still, even for the available metadata, there are issues. Searching through or browsing LOD is not straightforward because the metadata is often not structured and not machine-readable (Zuiderwijk et al., 2012). However, the majority of the participants (30, ≈58%) have declared that they provide metadata in machine-readable format (see Figure 6). Participants who declared that they do not provide the metadata in a machine-readable format have the following backgrounds: two are computational linguists, and one is a linguist. Most of the participants who have declared that they provide metadata in a machine-readable format are computer scientists.

Regarding the type of metadata that participants use (Figure 7), it seems that descriptive metadata is the most used. 52 (≈98%) participants use such metadata to describe the content of the data. Among such metadata, we can find the title, keywords, abstract, etc. Moreover, the descriptive elements that fall into this type support also the discovery, and the locating of such resources and they are also used to track the origin of the data. Then, the types provenance (26, 50%) and technical (25, 48%) metadata were the second and the third most used types of metadata declared. Provenance metadata provides information about the digital resource's history helping track its lifecycle,

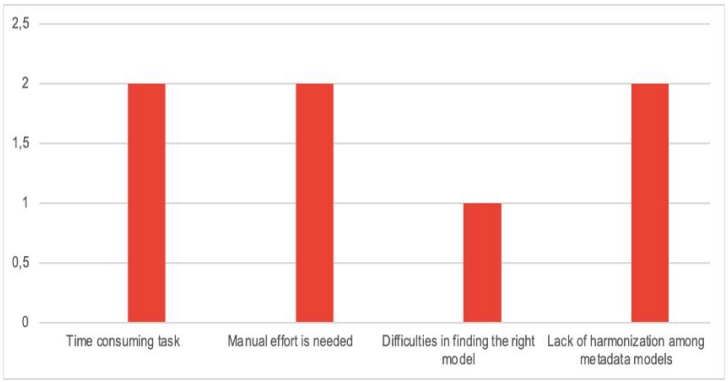

Figure 5: Reasons invoked for not using metadata to describe the developed LD resources.

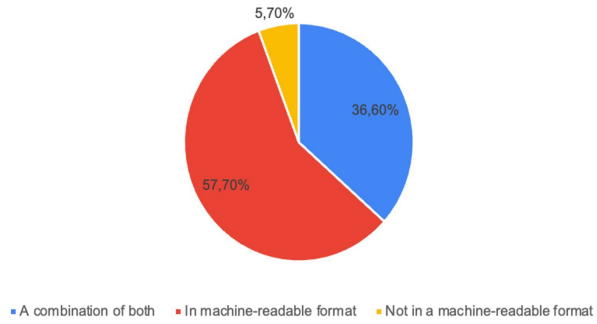

Figure 6: The distribution of the types of metadata reported as used for describing LD resources: machine-readable, not machine-readable and both.

while technical metadata provides information related to how a system functions or metadata behaves.
Administrative metadata, which aims at providing information about managing and administering collections and information resources, is the fourth most used type with declared by 24 participants (46%). The second less-used type of metadata is the Use metadata (19, ≈37%) which provides information related to the level and type of use of collections and information resources. Finally, Preservation metadata (12, ≈23%) are the ones that provide information about the preservation management of the resource.

Vocabularies are means of sharing information and documenting definitions that should be clear, thus reducing the ambiguity of terms used in the data. In order to describe the data, data producers use existing vocabularies or ad-hoc developed ones. When creating a vocabulary, it is a common practice to use or extend pre-existing ontologies and vocabularies, which favors communication between people and computer applications. However, most of the participants (27, ≈52%) declare that they develop their own vocabulary, while 25 (≈48%) use external vocabularies.

We asked all participants if they would be interested in a service or tool that supports them in the process of metadata creation or conversion. Figure 8 shows that 40 (69%) participants declared that they would be interested in such a service, 15 (26%) said that they might be interested, while 3 (5%) said that they have no interest. In fact, looking at the answers, 4 (66%) of the participants who did not use metadata to describe the data are interested in such a service, and 2 (34%) said that they might be interested. However, all the participants that do not have an interest in such a service do provide metadata about their data. This might be related to the fact that such users have already set the process of metadata creation and have no interest in a new service.

When it comes to the improvement of the metadata creation process, Figure 9 shows that 44 (76%) participants declared that they would be interested in a service that supports them in improving the metadata creation process; 13 (22%) said that they could be interested, and only 1 (2%) does not have any interest in such a service. The latter participant further declared that they use metadata to describe the data.

Table 4 contains the list of vocabularies and the number of times they were mentioned by the participants. The most used vocabulary is DublinCore[5], which is a set of fifteen "core" elements (properties) for describing resources. These properties are: Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, and Type. In fact, all 8 participants who use DublinCore use

[5] https://www.dublincore.org/

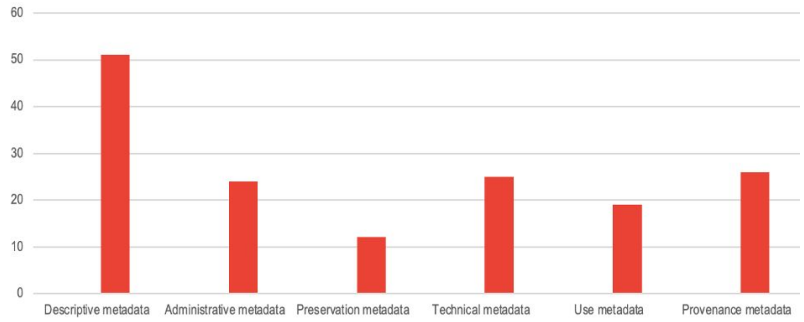

Figure 7: Kinds of metadata reported as used to describe the developed LD resources.

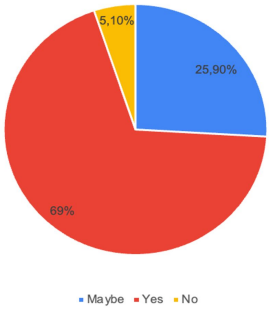

Figure 8: Distribution of participants who declared themselves interested in (red), not interested in (orange) and hesitant (blue) about a service/tool that would help provide metadata for LD resources.

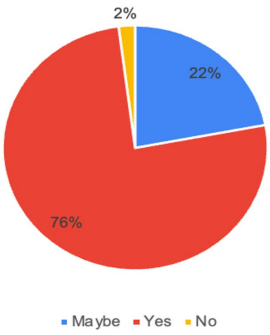

Figure 9: Distribution of participants who declared themselves interested in (red), not interested in (orange) and hesitant (blue) about a service/tool that would help improve creating/converting metadata?

Descriptive Metadata for their data. The second most used vocabulary is META-SHARE[6], which is used to describe language resources (corpora, lexical/conceptual resources, models, grammar, etc., and language processing tools and services) for Language Technology needs. DCAT[7] (Data Catalog Vocabulary) and OntoLex[8] are the third most used vocabularies. While DCAT is used with the aim of facilitating interoperability between data catalogs published on the Web, OntoLex is used to take care of the representation of lexica relative to ontologies. The less used vocabularies are used for specific purposes and include DRMJ[9], Preservica[10], etc.

Table 4: List of used vocabularies and times mentioned.

Vocabulary	**#**
DublinCore	8
METAShare	4
DCAT	3
Ontolex	3
CLARIN	2
LIME	2
Lexinfo	1
Wiki Vocabularies	1
IMDI	1
Preservica	1
EDM	1
Eurovoc	1
Prov Ontology	1
DDML	1
DataID	1
VoID	1
DPV	1
http://drmj.eu/	1

[6] http://www.meta-share.org/
[7] https://www.w3.org/TR/vocab-dcat-3/
[8] https://www.w3.org/2016/05/ontolex/
[9] http://drmj.eu/
[10] https://preservica.com/

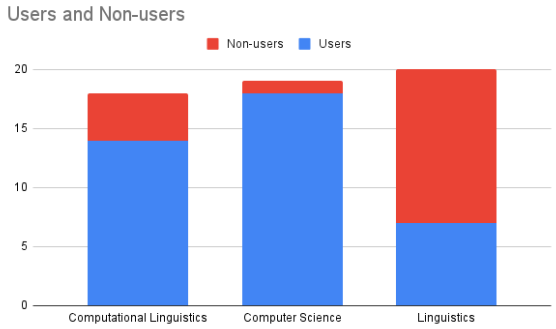

Figure 10: Participants as users of LLD resources (in blue) and non-users of LLD resources (in red), according to their declared background.

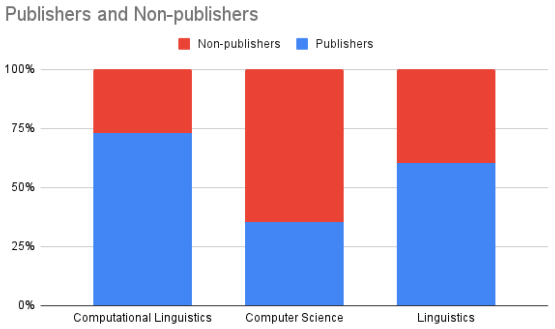

Figure 12: The correlation between the background and the tendency to publish LL(O)D resources.

5 Analysis and Discussion

In this section we try to correlate the responses to the different parts of the survey, with the aim of better understanding the conditions that prevent the wider adoption of LLD principles in the language resources community.

The **use** or non-use of LLD resources is highly correlated with their declared background: as shown in Figure 10, most (95%) computer scientists, many (77%) computational linguists, but only a third (33%) of the linguists used LD resources before.

We notice the same tendency when correlating the involvement of the participants in LLD resources **development** with their background: many (74%) computer scientists, a little more than half (56%) of the computational linguists, but only almost a third (30%) of the linguists were involved in the development of LLD resources. This distribution is rendered in Figure 11.

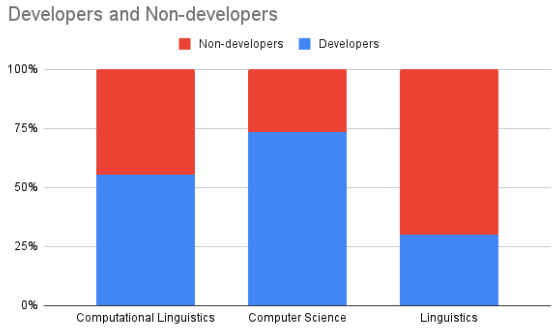

Figure 11: The involvement (in blue) and lack of involvement (in red) in the development of LD resources of participants according to their declared background.

Representing data in LLD format requires programming skills (Marshall and Shipman, 2003), which linguists rarely have. Thus, when asked under what conditions they would develop or convert LD resurces more frequently, 71% of the participants mentioned the existence of user-friendly tools to help them do this. The creation of such tools, however, might come with a cost: while easing the job of those less skilled in programming, such tools may work only for some domains or contexts, given the different nature of the data to be represented in various fields (Marshall and Shipman, 2003).

Looking at the background of those who publish or do not publish resources, we notice that computational linguists tend to publish the resources they develop more than linguists, while computer scientists tend not to do so (see Figure 12).

With respect to the relation between the background of the 28 non-developers of LLD resources and the reasons for not developing such resources, we find the data in Table 5, where we show the distribution of participants according to their declared background[11]. We can see that incompatibility between LD resources and other resources is a problem, especially for linguists (theoretical or computational), while rarely do computer scientists have it. The other reasons are invoked by members of all communities.

6 Conclusion

In this paper, we presented the results of a survey we conducted within the NL CA – WG1 to collect information useful to support the penetration of LLD, the identification of causes preventing

[11] One of the 14 participants mentioning incompatibility with other tools/resources as reason declared cognitive science as his/her background and this is not rendered in the table.

Reason	Total #	#CS	#CL	#Ling
Incompatibility with other tools/resources used	14	1	5	7
Lacking knowledge about adequate model/vocabulary	6	–	2	4
Models not totally appropriate for representing data	5	2	3	–
Unclear example or guidelines	8	2	2	3
Unclear advantages that LD has over other formats	3	1	2	–
Other	4	2	–	2

Table 5: Reasons for not developing LD resources correlated with participants' background. CS = computer science, CL = computational linguistics, Ling = linguistics.

such penetration, and possible limitations of such resources/technologies.

What emerged is that some bounds have to be overstepped in order to spread LLD principles to a wider community and fully exploit their potential. This survey is a confirmation of what the LD community is already aware of, and thus, it shows the need to take action.

We need to promote knowledge and skill transfer to support linguists in acquiring the necessary competencies for adopting LLD principles and technologies to their resources. On the other hand, the engagement of computer scientists in sharing knowledge and data as early as possible in the research process in open collaboration with all relevant knowledge actors (Von Schomberg, 2019) could contribute to support *open scholarship*[12].

At the same time, easing the (re)use, the creation, and the exposure of such resources could spread the adoption of LLD. This goal can be achieved through the development of specific adaptive tools, able to support different domains and languages, as well as formats to facilitate resource exchange and integration.

Furthermore, existing resources suffer from not being easily accessible, both in terms of findability, mostly due to the lack of harmonised and full-informative metadata descriptions, and usability, as LLD documentation is reported as scarce and inadequate.

With reference to the use of metadata, the current scenario could be improved by the availability of (semi)automatic solutions to reduce the time and effort for enriching resources manually, providing useful and consistent descriptions.

The documentation limits also affect the creation of new resources, preventing the adoption of LLD vocabularies/models to formalise linguistic data. This issue could be addressed by ensuring updated and maintained guidelines, enhanced by different examples and use cases and tailored to different backgrounds and levels of expertise, to support also less expert contributors/users through the whole cycle of linguistic linked datafication of their resources.

In future work, we intend to provide our contribution to defining some of the requirements to meet in order to ensure a large adoption of LL(O)D principles and promote a collaborative evolution of such resources.

[12]We adopt the term *open scholarship* instead of *open science* to adhere to the European policies, directed toward "open scholarship", as "open scholarship" reflects the inclusion of the humanities in the equation as well as emphasising the open input side to science in the form of open collaboration and active data and knowledge sharing prior to publishing and other scientific open outputs (Burgelman et al., 2019).

Acknowledgment

This work has been carried out within the COST Action CA 18209 European network for Web-centred linguistic data science (Nexus Linguarum). Maria Pia di Buono has been supported by Fondo FSE/REACT-EU - Progetti DM 1062 del 10/08/2021 "Ricercatori a Tempo Determinato di tipo A) (RTDA)". Azione IV.4 - Dottorati e contratti di ricerca su tematiche dell'innovazione/Azione IV.6 - Contratti di ricerca su tematiche Green.
The authors thank Julia Bosque Gil, Liudmila Rychkova and Max Ionov for their contribution to the survey drafting.

References

Julia Bosque-Gil, Verginica Barbu Mititelu, Hugo Gonçalo-Oliveira, Max Ionov, Jorge Gracia, Liudmila Rychkova, Giedre Valunaite-Oleskeviciene, Christian Chiarcos, Thierry Declerck, and Milan Dojchinovski. 2022. Balancing the digital presence of languages in and for technological development. A Policy Brief on the Inclusion of Data of Under-resourced Languages into the Linked Data Cloud. DOI: 10.5281/zenodo.7142513.

Jean-Claude Burgelman, Corina Pascu, Katarzyna Szkuta, Rene Von Schomberg, Athanasios Karalopoulos, Konstantinos Repanas, and Michel

Schouppe. 2019. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data*, 2:43.

Christian Chiarcos, Bettina Klimek, Christian Fäth, Thierry Declerck, and John Philip McCrae. 2020. On the Linguistic Linked Open Data infrastructure. In *Proceedings of the 1st International Workshop on Language Technology Platforms*, pages 8–15.

Maria Pia di Buono, Hugo Gonçalo Oliveira, Verginica Barbu Mititelu, Blerina Spahiu, and Gennaro Nolano. 2022. Paving the way for Enriched Metadata of Linguistic Linked Data. *Semantic Web Journal*, 13(6):1133–1157.

Margaret R Geddes. 2019. Strategies to support wider adoption of Linked Open Data in smaller museums. Johns Hopkins University.

Fahad Khan, Christian Chiarcos, Thierry Declerck, Maria Pia Di Buono, Milan Dojchinovski, Jorge Gracia, Giedre Valunaite Oleskeviciene, and Daniela Gifu. 2022. A survey of guidelines and Best Practices for the Generation, Interlinking, Publication, and Validation of Linguistic Linked Data. In *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference*, pages 69–77, Marseille, France. European Language Resources Association.

Leonardo Lezcano, Salvador Sánchez-Alonso, and Antonio J Roa-Valverde. 2013. A survey on the exchange of linguistic resources: Publishing Linguistic Linked Open Data on the Web. *Program*, 47(3).

Catherine C. Marshall and Frank M. Shipman. 2003. Which Semantic Web? In *Proceedings of the 14th ACM conference on Hypertext and Hypermedia*, pages 57–66.

Max Schmachtenberg, Christian Bizer, and Heiko Paulheim. 2014. Adoption of the Linked Data best practices in different topical domains. In *The Semantic Web–ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I 13*, pages 245–260. Springer.

Blerina Spahiu, Andrea Maurino, and Robert Meusel. 2019. Topic profiling benchmarks in the linked open data cloud: Issues and lessons learned. *Semantic Web*, 10(2):329–348.

Rene Von Schomberg. 2019. Why responsible innovation? In *International handbook on responsible innovation*, pages 12–32. Edward Elgar Publishing.

Anneke Zuiderwijk, Keith Jeffery, and Marijn Janssen. 2012. The potential of metadata for Linked Open Data and its value for users and publishers. *JeDEM-eJournal of eDemocracy and Open Government*, 4(2):222–244.