LingHub as a possible Type of Portal for finding Information about Dictionary Resources

Thierry Declerck1,2

1DFKI GmbH, Language Technology Lab

Stuhlsatzenhausweg, 3

D-66123 Saarbrücken

2Austrian Center for Digital Humanities

Austrian Academy of Sciences
Sonnenfelsgasse, 19
A-1010 Vienna
E-mail: declerck@dfki.de

1. Linghub

Linghub[[1]](#footnote-1) is an initiative towards the reconciliation of heterogeneous metadata about language resources. This work has been pursued within the past European project “Lider”[[2]](#footnote-2). A portal has been established (http://linghub.lider-project.eu/) in which users can browse and query for language resources.

Linghub concentrated on the harmonization of metadata describing the presence of language resources in four different repositories: CLARIN-VLO[[3]](#footnote-3), LRE-Map[[4]](#footnote-4), META-SHARE[[5]](#footnote-5) and Datahub[[6]](#footnote-6). As stated in (McCrae & Cimiano, 2015), the four initiatives and repositories named just above make use of quite different types of metadata sets: from curated by professionals to the more or less free generation of metadata in collaborative environments. The metadata sets can be very large, generic and complex, or rather simple, specific and flexible.

The approach followed by Linghub to harmonize those heterogeneous metadata sets was twofold: 1) Move all the metadata descriptions, if not done already, to the RDF syntax, and 2) checking for and correcting modeling errors, and establish an accurate URL schema. In this second step, efforts have been deployed for aligning the used descriptors to established vocabularies in the Semantic Web framework. Dublin Core[[7]](#footnote-7) FOAF (Friend of a Friend)[[8]](#footnote-8), SWRC (Semantic Web for Research Communities)[[9]](#footnote-9) and the Media Ontology[[10]](#footnote-10) turned out to be the most relevant vocabularies to be aligned with.

1. Is the Linghub Approach relevant to the Field of e-Lexicography?

Linghub was developed in the context of language resources that should be made available for Natural Language Processing (NLP) applications. And the question we would like to raise and discuss is if a similar approach can be beneficial to the field of e-Lexicography. A first question to this aspect concerns the type of metadata used in e-Lexicography and about their accessibility. Bothma (2011, p. 91) states that “To the best of my knowledge, such complex metadata markup of data does not yet occur in e-dictionaries”. If Bothma was right, and if the status of use of metadata in e-dictionaries didn’t change too much since the date of his statement, we could think in re-using as much as possible the harmonization work on metadata that has been done in the field of language resources, while specifying a set of metadata that are specific to the fields of the Working Groups of ENeL[[11]](#footnote-11), considering more specifically one of the objectives of WG1of ENeL[[12]](#footnote-12).

1. Acknowledgements

The presented work has been made possible by our participation to the European Concerted Research Action IS1305 “European Network of e-Lexicography (ENeL)”. Part of the work was supported by the H2020 Freme project (Grant agreement number 644771). Special thanks to John McCrae, The Insight Centre for Data Analytics, Galway, Ireland, for providing with the permission to re-use his slides and for all the discussions on Linghub and related topics.

1. References

Bothma, TJD. (2011). Filtering and adapting data and information in the online environment in response to user needs. In: Fuertes-Olivera, PA and Bergenholtz, H. (Eds.) 2011. e-Lexicography: The Internet, Digital Initiatives and Lexicography. London & New York: Continuum. Pp. 71 - 102.

McCrae, J.P., Cimiano, P., Doncel, V. R., Vila-Suero, D., Gracia, J., Matteis, L., Navigli, R., Abele, A., Vulcu, G. & and Buitelaar, P. (2015). Reconciling Heterogeneous Descriptions of Language Resources. *Proceedings of the 4th Workshop on Linked Data in Linguistics*, pp 39-48.

McCrae, J.P. & Cimiano, P. (2015). Linghub: a Linked Data based portal supporting the discovery of language resources. *Proceedings of the 11th International Conference on Semantic Systems*.

McCrae, J. P., Labrapoulou, P., Gracia, J., Villegas, M., Doncel, V. R. & Cimiano, P. (2015). One ontology to bind them all: The META-SHARE OWL ontology for the interoperability of linguistic datasets on the Web. In *Proceedings of the 4th Workshop on the Multilingual Semantic Web*.

Data catalog vocabulary (DCAT). W3C recommendation. The World Wide Web Consortium. https://www.w3.org/TR/vocab-dcat/ (accessed 22.03.2016)

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>



1. http://linghub.lider-project.eu/ [↑](#footnote-ref-1)
2. http://www.lider-project.eu/ [↑](#footnote-ref-2)
3. https://vlo.clarin.eu/;jsessionid=09B85365EF5C7DDE4158BAF6C25AAEA5?0 [↑](#footnote-ref-3)
4. http://www.resourcebook.eu/searchll.php [↑](#footnote-ref-4)
5. http://www.meta-net.eu/meta-share [↑](#footnote-ref-5)
6. https://datahub.io/ [↑](#footnote-ref-6)
7. http://dublincore.org/ [↑](#footnote-ref-7)
8. http://www.foaf-project.org/ [↑](#footnote-ref-8)
9. http://ontoware.org/swrc/ [↑](#footnote-ref-9)
10. http://www.w3.org/TR/mediaont-10/ [↑](#footnote-ref-10)
11. http://www.elexicography.eu/ [↑](#footnote-ref-11)
12. http://www.elexicography.eu/working-groups/working-group-1/wg1-objectives/ [↑](#footnote-ref-12)