Opening up cultural content in non-standard language data through cross-disciplinary collaboration: insights on methods, process and learnings on the example of exploreAT!

Amelie Dorn[1][0000-0002-0848-8149], Yalemisew Abgaz[2][0000-0002-3887-5342] and Eveline Wandl-Vogt[1][0000-0002-0802-0255]

[1] Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Vienna, Austria
[2] ADAPT Centre, Dublin City University, Dublin, Ireland
amelie.dorn@oeaw.ac.at; yalemisew.abgaz@adaptcentre.ie; eveline
wandl-vogt@oeaw.ac.at

Abstract. Understanding collaboration between researchers of different disciplines requires an ability to embrace multiple views and perspectives, and communicative efforts. This paper thus provides insights on methods, processes and results of a cooperation in Humanities research supported by semantic technologies with the aim of accessing and opening up cultural knowledge contained in a non-standard language resource. The collaborative undertaking is carried out within a Digital Humanities project and an Open Innovation framework. Meta-disciplinary learnings offer insights on factors fostering mutual understanding, knowledge translation and mutual benefits.

Keywords: Cross-disciplinary cooperation, cross-organizational collaboration, Digital Humanities.

1   Introduction & Background

Culture is a complex phenomenon that offers grounds for analysis in academia, society, arts, etc. from various perspectives [1]. It encompasses several aspects of a society and has been widely expressed and conveyed over the centuries by words, stories, songs, poems, paintings, writings and several other methods most typically through the medium of language. Culture and language are thus tightly interwoven concepts that transcend several societies in time. In recent times, there has thus been a trend in the Humanities in preserving cultural content, mostly contained in written texts, taking language as a first access point. With the support of modern technological tools and the ever growing capacities of digital methods and devices, also otherwise hidden or implicit cultural knowledge contained in Humanities data can be made visible and accessible. Language data thus needs to be available digitally and in technologically enhanced and systematic formats to be accessed and used by the wider community of the modern era, for it to be ultimately preserved through re-use and connectability.

In this paper, we address the collaboration between Humanities scholars and semantic technologists in a Digital Humanities context (the exploreAT! project) [2] on the example of a historic language resource (DBÖ) [3, 4]. We discuss the opening up and exposition of this traditional non-standard German language collection using semantic modelling which exploits existing semantic web standards to represent and facilitate a common representation and interpretation of these cultural resources. Ontologies from different domains and developed by our team are integrated and used to represent the traditional resources to enhance their discoverability and usability in both independent manner and integrated with other similar standardized resource.

We here report on our collaboration results, the humanities background to our research question, the technical methods and implementation, but also on another important yet often unmentioned aspect of language in such cross-disciplinary collaborations, namely the translation of knowledge and expertise across disciplines. Openness to learnings, mutual understanding and communication are key elements in a foundation of bringing about successful results

2   Opening up cultural contents of a traditional language resource: the exploreAT! project

exploreAT! is a current DH project which aims to unveil cultural information contained in a non-standard language resource (DBÖ) [Database of Bavarian dialects in Austria; [3]] by drawing on and combining digital methods and tools from different disciplines (semantic technologies, visualisation prototyping, crowd science) (cf. [5]). At the heart of the project lies the fundamental research question originating from the Humanities background, which asks how to enable access to a non-standard language resource through a cultural lens, giving insights on the conceptualisation of the world and the local society at the time. In this context, the DBÖ resource offers a wealth of not only valuable language data, but also rich cultural content. The database counts around a total of 3.5 million entries, including original data collection questionnaires, answers as well as other digitized excerpts of folklore literature. Originally collected in the area of the former Austro-Hungarian empire with the aim of capturing the speech of the local population, the former collection and following digital preparation was already a huge collaborative effort across persons of various professions, backgrounds and functions, offering detailed documented cultural and societal insights on topics of everyday life (e.g., festivities, professions, nature, food, etc). In particular, our current efforts concentrate around the topic of food, which offers rich grounds for analyses, connectability as well as scientific and societal relevance. Through the support and application of semantic tools, this implicit cultural knowledge can be accessed and connected to other sources and resources for multilingual and multicultural comparison.

3 Cross-cultural Team Communication and Knowledge Exchange: Methods & Tools

The exploreAT! project is all the more interesting as it not only combines cross-disciplinary expertise, but also collaborators of very different cultural and linguistic backgrounds, located across Austria, Spain and Ireland. Methods and tools used for team communication and knowledge exchange are thus key in harmonising and leveraging results and communicating tasks, but also addressing challenges or uncertainties in the workflow. In the wider context of exploreAT!, a combination of digital and analogue methods and tools are employed for ideation (e.g. agile and design thinking tool kits), communication across team members (e.g. web-based project management and communication technologies) or for capturing project ideas and development.

In this paper we concentrate on the description of the specific collaboration scenario which focuses on the creation of the semantic data model. This collaboration arises out of the humanities research question on how to make cultural knowledge in a language resource accessible, discoverable and connectable. In this particular context, current digital tools for communication and task management (Slack, Trello, Skype) were employed, as well as regular face-to-face meetings. While online tools were used for frequent exchange, face-to-face meetings served more specifically for discussions on major project goals, creating work plans or for joint team meetings including also project members. In order to implement collaborative writing, editing or brainstorming a free web-based software office suite was used that could be accessed from any computer with an internet connection.

Drawing on these tools, in what follows we elaborate on the methods, collaborative processes and learnings on the example of the composition of the semantic model [6,7] based on the Humanities research question and resource.

4 Cross-disciplinary Collaboration: the example of creating a Cultural Semantic Data Model

4.1 First processes towards joint collaboration for Semantic Modelling

The aim of the semantic modelling in the context of exploreAT! was to enable the discovery of cultural content in our language collection and connect it to other multilingual and multicultural resources using LOD [8]. The data collection questionnaires and related questions served as the initial access point to the remainder of the collection and to enable connectability to other resources. The modelling further served to understand the semantics of the core entities as defined by the humanists and as contained in the collection, and to represent them and their relationships using existing up-to-date semantic web technologies and standards. With the language collection being focused on a specific domain (non-standard language), and the overall method used to collect the data dating back to the beginning of the 20th century, it was crucial for the semantic

technologists to collaborate in direct exchange with the humanists. In our case, the collaboration involved three major teams. The first team (humanists) consisted of the domain knowledge experts who were involved in or had in-depth knowledge about all steps of the original data collection, organisation and utilisation. The second team (linguists, lexicographers) are researchers in the area of socio-cultural linguistics and related fields, and the third team (technical experts) comprised ontology engineers and semantic web experts responsible for developing the semantic model and uplifting the collection using a linked open data (LOD) platform. The collaboration example we report on here, evolved in three steps.

1. The first joint work laid the foundation for understanding the overall area of expertise and the fundamentals of the dataset.
2. The next step involved collaborating for modelling the core entities of the collection using current semantic web technologies.
3. Finally, search, visualisation and exploitation of the results is presented.

Each of the three steps is described in the following sections.

4.2   Methods and interactions enabling access to implicit data knowledge

Understanding and identifying the implicit knowledge contained in the language collection in general and the detailed meaning and interpretation of the cultural and linguistic entities, in particular, was among the challenges largely faced by both technical experts and linguists. As soon as the semantic modelling process started, the gap became visible in that much of the knowledge which is useful to understand the collection is not self-contained in the data. Thus, it became necessary to gain a deeper knowledge of the data from sources other than the collection itself. Especially for the technical experts, this became a challenge as their objective was to semantically organise and describe the content. Initially, all available information was shared among the teams on the cloud platforms used in the project. This included several resources such as publications describing the collection, notes and change logs. Although the information helped the technical experts to better understand the collection, it generated new questions to the humanists, given the complex structuring of the materials, resulting in less productive weekly meetings and only partially satisfactory advancement. The process became time-consuming as technical experts were remotely located from the humanists, and knowledge experts could not provide the necessary information at the same pace as the technological advancement proceeded.

As communication by digital means only didn't prove optimal, resorting to a different form of knowledge exchange, namely face-to-face meetings, became inevitable. The first face-to-face meeting on the topic of semantic modelling brought the different members involved (humanists, domain experts and technical experts) together in a workshop setting with the aim of building a common understanding of the collection, the methods, resources and techniques used for the original data collection process and to investigate other possible sources of information. This collaboration workshop took

place at exploration space @ ACDH-OeAW in Vienna. The workshop provided valuable insights for both humanists and technical experts as it initiated discussions on topics, such as the identification of cultural content indicators, identification of relevant data fields, or task distribution and enabled the humanists to create new structures for cultural content discovery supporting and enhancing the semantic modelling process. The workshop paved the way for opportunities on planning and proposing a concrete way forward in terms of tasks and workflows, and gave team members a solid understanding of the challenges and complexities involved, and made its contribution to elicit the requirements of each team. Since the initial meeting, a number of similar workshops were conducted in Dublin, Salamanca, Vienna and CERN by incorporating different stakeholders to discuss new opportunities.

Oftentimes a unilateral attempt to model a non-standard language resource can result in an ill-representation, potentially leading to less usability. This face-to-face interaction enabled the discovery of key aspects which would have been challenging, time-consuming or even more complex to communicated by digital or written means only. The semantic modelling exercise resulted in the identification of cultural and linguistic indicators from the side of the humanists and a conceptual model of the collection and its representation using an ontology in owl language, from the technical experts. The resulting ontology and its representation is discussed in detail in [7,9].

A key takeaway for collaboration, is that face-to-face meetings and direct exchange may foster team spirit among collaborators, potentially fuelling further collaboration beyond the current project. In addition, it allows for cross disciplinary collaboration of seemingly far apart areas and benefits members in terms of understanding potential complexities involved in other areas of expertise.

4.3 Synthesizing Humanities and technical expertise towards a first prototype

A next step in the joint collaboration included establishing individual workflows for each team and working towards first common results, a cultural data model for non-standard data questionnaires [6]. Through weekly exchanges and updates using digital communication channels, advancements from both humanists and technical experts were consolidated. Particularly in the joint creation of a data model, the consolidation of views from a semantic web expert and a Digital Humanities are key, as naming conventions or details of representations may vary significantly. Bringing these differences together and narrowing the gap on the representation is crucial, often triggering further revisions, where trade-offs need to be made.

Finally, a first prototype of the data model was presented and discussed with other members of the exploreAT! project in a second workshop. There opportunities arose for the technical expert to engage other project members in a constructive discussion by demonstrating the solution and the application areas. This further face-to-face meeting enabled the technical expert to perform several refinements of the model, including in cleaning noisy data, and it also paved the way for further discussion of the architecture of the implementation. As a result of the direct interaction, several key decisions

could be taken and implemented by all experts involved. Any follow-up communication could thus be continued in online meetings and standups via Skype and Slack channels in regular intervals.

4.4 Creating exploration paths for mutual understanding: facilitating search, visualisation and exploitation of the results

After collaborating in smaller groups for the purpose of elaborating the data model, the next step involved the consolidation and communication of results to the other project members and areas of expertise, such as visual prototyping.

Translating the queries provided by the humanists into a high level technical query language proved challenging. The purpose of the semantic modelling and annotation of the collection was to enable the users to discover cultural content in a non-standard language collection and explore their semantic relationships discovering new insights and support for their research hypotheses. However, providing the resulting semantic research collection with a query user interface often fails in serving the purpose. To address this gap, the initial queries of the humanists were translated to exploration paths in order to elicit the exact requirements. This process involved navigating through the data collection step-by-step, building navigation paths of one or two steps at a time to include further requirements after identifying an initial pivotal query. The exploration paths laid a foundation for the semantic web and visualisation experts to understand the requirements of the users in their own perspectives and to interpret the queries of the target users. At the same time, it enabled the humanists to understand how the semantic data could be efficiently exploited to support their research questions. This was a significant step in the collaboration to understand how the semantic modelling process enhanced the requirements of the users and to provide additional customisable user interfaces to enable the users to pose their own questions.

5 Insights & Conclusion: metadisciplinary learnings

Our collaboration of Humanities research supported by semantic technologies has brought about valuable insights and learnings regarding the knowledge exchange process in terms of creating scientific results, but also in terms of team composition that can prove helpful for training purposes. From our experience, we can report that embracing team diversity brings wealth in both expertise and perspectives. Bringing together researchers of various roles enables a more complete picture and analysis of various perspectives in terms of addressing a particular research question, ultimately consolidating results. What is a key prerequisite, however, is the individual ability to bringing openness and flexibility to a team, which, if lacking, may pose difficulties to the collaboration process. In addition, fostering mutual understanding for involved disciplines can be assured by taking part in training courses in order to obtain basic knowledge in, for example semantic technologies, which also proved beneficial in terms of communication and translation of knowledge. Finally, experimenting and re-

flecting on novel methods of communication and idea-finding may additionally contribute to bringing together different perspectives and enable better mutual understanding. The team applies and analyzes novel approaches towards collaboration in an Open Innovation [10] framework, for example working together with designers [9] to increase the learning curve and potential mutual benefits.

Based on the learnings from exploreAT!, a virtual and physical space for experimentation and innovation has been funded, namely exploration space, currently a best practice example of the Open Innovation platform of the Austrian government (http://open-innovation.gv.at/portfolio/oeaw-exploration-space/).

Acknowledgements

References

1. Longhurst, B. & Baldwin, E. (Eds) Introducing Cultural Studies. Routledge (2008).
2. Wandl-Vogt, E., Kieslinger, B., O'Connor, A. & Theron, R. exploreAT! Perspektiven einer Transformation am Beispiel eines lexikographischen Jahrhundertprojekts. In: DHd2015. Von Daten zu Erkenntnissen. 23. bis 27. Februar 2015, Graz. Book of Abstracts. (2015)
3. [DBÖ] Österreichische Akademie der Wissenschaften. (1993–). Datenbank der bairischen Mundarten in Österreich [Database of Bavarian Dialects in Austria] (DBÖ). Wien. [Processing status: 2018/01]
4. Wandl-Vogt, E. …wie man ein Jahrhundertprojekt zeitgemäß hält: Datenbankgestützte Dialektlexikographie am Institut für Österreichische Dialekt- und Namenlexika (I DINAMLEX) (mit 10 Abbildungen). In P. Ernst (Ed.), Bausteine zur Wissenschaftsgeschichte von Dialektologie / Germanistischer Sprachwissenschaft im 19. und 20. Jahrhundert. Beiträge zum 2. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen, Wien, 20. - 23. September 2006. Wien: Praesens, pp. 93–112 (2008).
5. Dorn, Amelie, Eveline Wandl-Vogt, Yalemisew Abgaz, Alejandro Benito Santos, and Roberto Therón. Unlocking Cultural Conceptualisation in Indigenous Language Resources: Collaborative Computing Methodologies. In: Claudia Soria, Besacier, Laurent, and Pretorius, Laurette. (eds.) Proceedings of the LREC 2018 Workshop "CCURL 2018 – Sustaining Knowledge Diversity in the Digital Age", 12 May 2018, Miyazaki, Japan, pp. 19-22 (2018).
6. Abgaz, Yalemisew, Amelie Dorn, Barbara Piringer, Eveline Wandl-Vogt, and Andy Way. Semantic Modelling and Publishing of Traditional Data Collection Questionnaires and Answers. Information 9: 297-320 (2018). doi:10.3390/info9120297
7. Abgaz, Yalemisew, Amelie Dorn, Barbara Piringer, Eveline Wandl-Vogt, and Andy Way. A Semantic Model for Traditional Data Collection Questionnaires Enabling Cultural Analysis. In: John P. McCrae, Chiarcos, Christian, Declerck, Thierry, Gracia, Jorge, and Klimek, Bettina. Proceedings of the LREC 2018 Workshop "6th Workshop on Linked Data in Linguistics (LDL-2018)". Miyazaki (2018).
8. De Wilde, Max, and Simon Hengchen. Semantic Enrichment of a Multilingual Archive with Linked Open Data. *Digital Humanities Quarterly 11:* 1938 – 4122, (2017).

9. Goikhman, Alisa, Roberto Therón, and Eveline Wandl-Vogt. Designing collaborations: could design probes contribute to better communication between collaborators?. In: Francisco José Garcia-Peñalvo. TEEM '16. Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality. Salamanca, Spain – November 02 - 04, 2016. New York: ACM. (2016) doi:10.1145/3012430.3012431.

10. Open Innovation Strategy for Austria. Goals, Measures & Methods. *Federal Ministry of Science, Research & Economy (bmwfw) and Federal Ministry of Transport, Innovation and Technology (bmvit).* (2015) http://openinnovation.gv.at/wp-content/uploads/2015/08/OI_Barrierefrei_Englisch.pdf, last accessed 2019/01/08