

CLARIN: Towards FAIR and Responsible Data Science Using Language Resources

Franciska de Jong¹, Bente Maegaard², Koenraad De Smedt³, Darja Fišer⁴, Dieter Van Uytvanck⁵

^{1,5}CLARIN ERIC, ²University of Copenhagen, ³University of Bergen, ⁴University of Ljubljana and Jožef Stefan Institute
^{1,5}Utrecht, The Netherlands, ²Copenhagen, Denmark, ³Bergen, Norway, ⁴Ljubljana, Slovenia

¹f.m.g.dejong@uu.nl, ²bmaegaard@hum.ku.dk, ³desmedt@uib.no, ⁴darja.fiser@ff.uni-lj.si, ⁵dieter@clarin.eu

Abstract

CLARIN is a European Research Infrastructure providing access to language resources and technologies for researchers in the humanities and social sciences. It supports the study of language data in general and aims to increase the potential for comparative research of cultural and societal phenomena across the boundaries of languages. This paper outlines the CLARIN vision and strategy, and it explains how the design and implementation of CLARIN are compliant with the FAIR principles: findability, accessibility, interoperability and reusability of data. The paper also explains the approach of CLARIN towards the enabling of responsible data science. Attention is paid to (i) the development of measures for increasing the transparency and explainability of the results from applying CLARIN technologies, in particular in the context of multidisciplinary research, and (ii) stimulating the uptake of its resources, tools and services by the various communities of use, all in accordance with the principles for Open Science.

Keywords: CLARIN, FAIR, research infrastructure, Open Science policies, multilinguality, digital humanities, responsible data science

1. Introduction: CLARIN Vision and Strategy

CLARIN¹ is a European Research Infrastructure providing access to language resources and tools. It focuses on the widely acknowledged role of language as social and cultural data and the increased potential for comparative research of cultural and societal phenomena across the boundaries of languages. Access to language data is crucial for scholars in the social sciences and the (digital) humanities (De Smedt et al., 2018). Language is a carrier of cultural content and information, both synchronically and diachronically. Language also plays a role as the reflection of scientific and societal knowledge, as an instrument for human communication, as one of the central components of the identity of individuals, groups, cultures or nations, as an instrument for human cognition and expression, as a formal system, and as historical records in need of documentation and preservation. These insights underline the vision and the mission of CLARIN.

CLARIN vision: All digital language resources and tools from all over Europe and beyond are accessible through a single sign-on on-line environment for the support of researchers in the humanities and social sciences.

CLARIN mission: Create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences. The vision is obviously very ambitious and the goal will not be fully reached in a foreseeable future, but it is the driving force for those who participate in the consortium and an encouragement for researchers to use and contribute to the infrastructure.

CLARIN has been implemented as a distributed research infrastructure with data centres (nodes) across Europe and beyond. The activities in CLARIN basically take place at two levels. One is the central level: the Board of Directors and the technical, communication, and administrative staff. The other is the national level: in each member country, the

national activities have by far the largest volume. Previous CLARIN activities have been described in multiple publications that have been disseminated through many channels. At LREC 2014 an overview of language resources, tools, and services on offer through CLARIN was presented (Hinrichs and Krauwer, 2014). The current paper focuses on the overall vision and in particular the adherence to the principles of FAIR and Responsible Data Science.

1.1. A Bit of History

In the preparatory phase 2008–2011 CLARIN was funded by the European Commission. It was established as a European Research Infrastructure Consortium (ERIC)² in 2012; the basic funding comes from the member countries. When it was established in 2012 CLARIN had nine members, and in 2017 it has grown to 19 members and two observers. Additionally, CLARIN has a special agreement with Carnegie Mellon University in the USA. This growth, which is still on-going, shows that the basic idea of CLARIN was well chosen and timely, cf. the map in Figure 1.

An important factor for the success and sustainability of a research infrastructure like CLARIN is its scope and size: since CLARIN is collecting data for languages from all regions and periods that are of interest to the European research era, CLARIN is multilingual. A large part of the collected language resources come from the member countries, so the number of members matters for realizing wide coverage. Additionally, CLARIN is open to registering resources in all languages, and CLARIN is truly multilingual by providing metadata for at least five resources for each of more than 1,500 languages.

CLARIN has established the Virtual Language Observatory (VLO)³, a registry of Language Resources (LRs) based on the CMDI metadata standard (cf. Section 2.). The VLO contains information about all LRs provided by the member countries, plus information from other registries that

²According to regulations, the consortium needs to have a majority of European countries, but it is not restricted to Europe.

³<http://vlo.clarin.eu>

¹<http://www.clarin.eu>

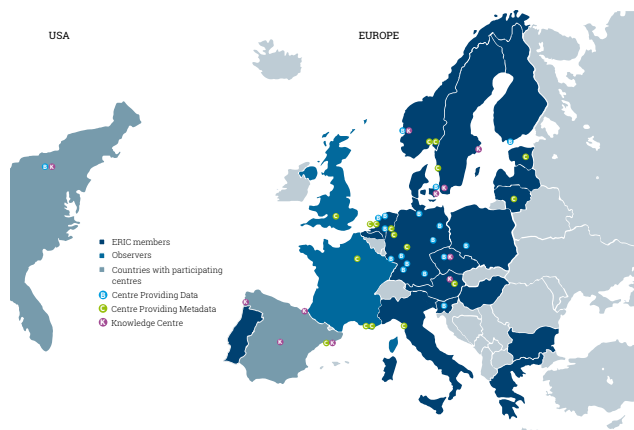


Figure 1: Map of CLARIN members, observers, and participating centres at the start of 2018.

want to be present in the VLO. Recently more than 700,000 records have been added from Europeana,⁴ the EU digital platform for cultural heritage, which is of particular interest to humanities researchers.⁵ With more than 1,600,000 entries in the VLO CLARIN supports access to language resources including cultural heritage in many languages.

1.2. Strategy

After the first 5–6 years, CLARIN is currently shaping its strategy for the next three years, 2018–2020. The strategy builds on the vision given above and further expressed in the Value Proposition⁶ (Maegaard et al., 2017). The strategy covers four areas: uptake by researchers, technical infrastructure, knowledge sharing infrastructure and sustainability.⁷

The strategy concerning the latter two points will be addressed in the following subsections. Technical infrastructure will be treated below in Section 2., where also the strong connection to FAIR is described, while responsible data science is addressed in Section 3. and uptake is described in Section 4..

1.2.1. Knowledge Sharing Infrastructure

CLARIN's Knowledge Sharing Infrastructure (KSI) is the necessary 'glue' between the Technical Infrastructure and the researchers. Several instruments are used to share knowledge. Here we mention only a few of them. *Knowledge centres* provide expertise on data and methods to researchers. Knowledge centres may offer expertise and support for certain languages (e.g. the languages of Spain), or a technology or type of data (e.g. treebanks). CLARIN has nine knowledge centres at the time of writing and more are being established. *Workshops* are key to share knowledge and to develop new ideas. Workshops may be used to support new user communities, to make the needs of

a particular community visible and to discuss known approaches across borders (country borders or discipline borders). Workshops are normally organised by members and financially supported by CLARIN ERIC.

1.2.2. Sustainability

Membership is an essential part of the sustainability strategy. Countries and international organizations can join the ERIC as members or observers. They contribute to the governance of CLARIN and the membership fees form the core of its business model.⁸ The strategy is to maintain the current membership and to grow in Europe and beyond. An important aspect of the strategy is collaboration across borders: CLARIN aims at cross-institutional and cross-sectorial collaboration, e.g., with the GLAM sector (Galleries, Libraries, Archives, Museums) and with the industry. Collaboration with other research infrastructures, be it in the humanities and social sciences area or with eScience, is also pursued in order to foster multidisciplinary and the inherent need for innovation of methodological frameworks. Through cross-border collaboration, as well as through the focus on training and education, on centres of expertise etc. CLARIN will increase its societal impact and will contribute to the development of methodologies for measuring impact of research infrastructures in the social sciences and humanities. Finally, an inherent goal of all these activities is to integrate and contribute to Europe's Open Science policies (De Smedt et al., 2018).

2. CLARIN and FAIR

With the wide emergence of digitally available language data (be it digitally native or retrodigitized), the possibilities beyond the mere archiving and viewing of such data sets have grown significantly. The idea of transitioning towards an ecosystem of registered, updatable and actionable data has been proposed in the context of the so-called Live Archives (Trilsbeek et al., 2008). These insights were guiding foundations during the establishment of the CLARIN infrastructure (Broeder et al., 2008).

More recently, similar principles have been put forward in the broader data science community, specifically in the context of the Research Data Alliance's Data Fabric working group.⁹ The most visible embodiment of these principles are the so-called FAIR principles (Wilkinson et al., 2016) that are now also widely promoted as part of the Open Science paradigm. In the paragraphs below we describe the data architecture of CLARIN in a nutshell and demonstrate afterwards how it is largely FAIR-compliant.

2.1. CLARIN Data Architecture

As a distributed infrastructure, CLARIN exists of a network of technical centres (nodes). A CLARIN centre typically provides a data repository where language resources are stored and made available for researchers. Additionally, many centres also provide tools (web applications, web services or stand-alone applications) to process language data.

⁴<https://pro.europeana.eu/>

⁵<https://www.clarin.eu/blog/>

bridging-europeana-and-clarin-infrastructures

⁶<https://www.clarin.eu/value-proposition>

⁷<https://www.clarin.eu/content/vision-and-strategy>

⁸See (OECD, 2017) on business models for research data infrastructures

⁹<https://www.rd-alliance.org/group/data-fabric-ig.html>

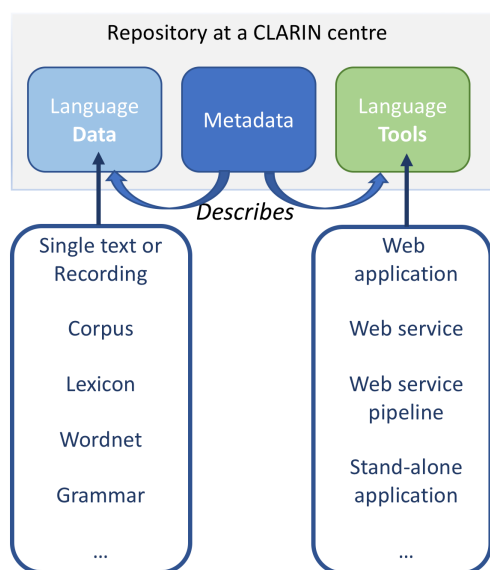


Figure 2: Schematic overview of what CLARIN technical centres have on offer.

Figure 2 is a schematic description of what type of content a repository incorporates. Figure 3 illustrates how metadata from CLARIN repositories are harvested in order to make items searchable and how described datasets can be processed with CLARIN tools.

There are two important principles behind the establishment of CLARIN centres. The first one is the organic nature: CLARIN centres typically evolve from hotspots of know-how, e.g. academic research groups or language data archives. Such fields of specialization could be based on a specific language (e.g. Hungarian), a modality (e.g. speech data), data types (e.g. lexicons) or a processing paradigm (e.g. rule-based morphological analysis). Secondly, there are strict interoperability requirements, in particular for the technical centres. As free as the process of establishing a CLARIN centre can be, a reasonable degree of integration and compatibility is possible only if at the basis there is a clear technical and organisational framework. This is why each CLARIN centre is periodically assessed on various levels, to check if it is following the technical guidelines and if there is a sound organisational backing for the repository. One should note that these requirements only pertain to the protocol level (e.g., OAI-PMH for metadata exchange), leaving complete freedom to the individual centre to choose its preferred software stack.

Without going into details,¹⁰ the main pillars for interoperability with CLARIN are based on HTTP-accessible data and metadata, persistent identifiers and single sign-on whenever resources are not openly available. In the subsequent sections we will relate these pillars to the FAIR principles.

¹⁰The full requirements are available at <http://hdl.handle.net/11372/DOC-78>.

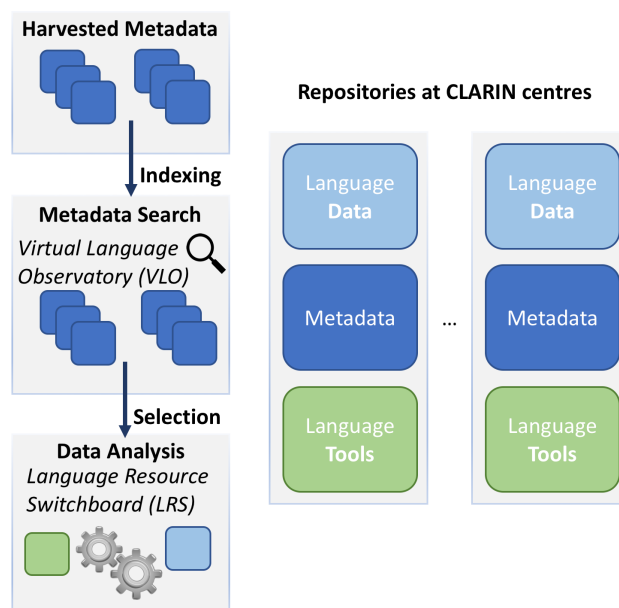


Figure 3: The metadata for both data and tools from CLARIN repositories are harvested and indexed by the Virtual Language Observatory. Selected data can then be processed with relevant CLARIN data analysis tools.

2.2. Findable

Use and re-use of language resources is only possible when they can be easily found by researchers. FAIR therefore requires persistent identifiers, rich metadata which are indexed and searchable and links from the metadata to the data identifiers. As a metadata standard, CLARIN requires CMDI (Goosen et al., 2015). *Handles* are used to refer to all metadata records (and to data where this makes sense). In addition, the Virtual Language Observatory (VLO) (Van Uytvanck et al., 2012) provides a powerful search portal for all the collected metadata records, probably the largest catalogue of its kind for language data. Figure 3 depicts how the VLO architecture supports the discovery of data and tools. One can search through all metadata descriptions to find a relevant data set. The VLO offers free text search as well as faceted search. An example is given in Figure 4.

Making a searchable catalogue for over a million resources is obviously a daunting challenge. Usability tests have been performed and have triggered a series of improvements to the user interface. Also, issues with harvested metadata are always to be expected; these are being addressed through continuous adaptation and curation of the metadata.

2.3. Accessible

Once language data are found, they should be easy to access. FAIR translates this requirement into the need for a standardized communication protocol, with the option for easy authentication and authorisation if needed. Metadata should remain accessible, even when the data are no longer available. CLARIN relies on the widely used HTTP protocol and SAML for federated single-sign-on for its implementation of these policies.

The screenshot shows the VLO search interface with the following elements:

- Header:** Virtual Language Observatory, Search, Help, CLARIN logo.
- Breadcrumbs:** VLO / Faceted search / Search results.
- Search Bar:** newspaper
- Filters:** Showing 1 to 10 of 24214 results within selection for newspaper x Finnish x. Results per page: 10.
- Faceted Search:**
 - Language:** Finnish x
 - Collection:** (empty)
 - Resource type:** (Menu open)
 - TEXT (24207)
 - Analytic serial (24196)
 - Newspaper Issue (24196)
 - Newspaper Title (11)
 - Serial (11)
 - Corpus (4)
 - Written Corpus (1)
 - Modality:** (empty)
 - Format:** (empty)
 - Availability:** (empty)
- Search Results:**
 - Finnish newspaper subcorpus from 2012 (fin_news_2012_300K)** (Part of Leipzig Corpora Collection)
 - 300.000 sentences of a Finnish newspaper corpus based on material from 2012
 - The Karelian Finnish Newspaper Corpus** (Part of CLARIN Centres)
 - The corpus contains issues of the Karjalan Sanomat newspaper published in 2012-2014. The corpus is available in Kielipankki - the Language Bank of Finland (http://urn.fi/urn:nbn:fi:lb-2016112501). In case you are not a member of an academic institution please read the access rights instructions at https://www.kielipa...
 - The Karjalainen Corpus** (Part of LRT + Open Submissions Data & Tools)
 - computer corpus of Finnish newspaper texts of the 1990s (newspaper Karjalainen, Joensuu)
 - The Karjalainen Corpus** (Part of CLARIN Centres)
 - Computer corpus of Finnish newspaper texts of the 1990s (newspaper Karjalainen, Joensuu). The purpose of the resource use must be outlined in a research plan.
 - The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771-1874)** (Part of CLARIN Centres)

Figure 4: Screenshot of a VLO search for the keyword “newspaper”, restricting the “language” facet to “Finnish”. Further restrictions can be imposed with other facets, such as “Resource type” for which a menu has been opened.

2.4. Interoperable

To attain interoperability, FAIR demands a formal, shared and broadly applicable language for knowledge representation, using FAIR vocabularies and links between metadata and data. CLARIN relies on the CMDI framework as common metadata language, including links to standardized OpenSKOS¹¹ vocabularies and standardized ways of linking to datasets and landing pages. There are also recommendations for the use of standard data formats.

2.5. Re-usable

Researchers need information on how and under which conditions they can re-use existing resources. FAIR states that re-usable data require clear license and provenance information and adherence to community standards. CLARIN has clear recommendations on license disclosing and user-friendly ways of categorizing these (Arppe et al., 2011). Provenance is an important part of the metadata in CLARIN. While community standards are hard to define, the bottom-up structure of the centres definitely brings

along close ties with such good practices.

3. CLARIN and Responsible Data Science

In order to serve the research community, the adherence to the principles of Open Science not only demands clarity about the data services offered, but also about what is to be expected from the analysis tools offered. In particular researchers that do know how to couple the data they aim to use but that have not developed tools themselves, are in need of guidance and explanation of the various aspects of quality. Given that CLARIN offers support for multiple data science paradigms (e.g., rule-based vs. data-driven), performance expectations need to be carefully managed.

3.1. FACT

Across almost all domains of research there is growing concern for how data, especially ‘big data’, are put to use. Any use of data which is biased, violates privacy or confidentiality, or lacks transparency, may distort conclusions or break trust relations. These concerns also pertain to language data and the analysis tools integrated in the CLARIN infrastructure.

¹¹<http://openskos.org>

A recent expression of these concerns is by the Responsible Data Science consortium (RDS)¹² which aims to tackle ethical and legal challenges, promote data science techniques, infrastructures and approaches that are responsible in the sense that data and data use should be fair, accurate, confidential and transparent (FACT) (Van der Aalst et al., 2017). These aims complement the FAIR principles, especially in a context where the use of data-driven methods typically applied to larger datasets is on the rise.

CLARIN intends to contribute to responsible data science by the design as well as the governance of its infrastructure and to achieve an appropriate and transparent division of responsibilities between data providers, technical centres, and end users. License agreements, which are established between a data provider and a data centre (repository), regulate the terms under which some well-described data is made available. These terms include an end-user license agreement which, together with the terms of service at the data centre, may place some restrictions and responsibilities on the end user, particularly in the case of privacy concerns. The requirement of provenance data in CLARIN metadata makes data traceable and the use of PIDs makes data citable and their use replicable. Furthermore, CLARIN has started to provide guidance on which tool is recommended for which data through a service called the Language Resource Switchboard (Zinn, 2016).

3.2. From Big Data to Big Conclusions and Decisions?

Additional steps are foreseen which relate to the need to document and explain the performance levels that can be expected from the analysis tools, and thereby of the suitability of certain tools for specific scenarios of use. This is particularly relevant for the uptake of CLARIN functionality in the context of multidisciplinary collaboration where methodological frameworks rooted in multiple scholarly traditions have to be combined. This is not always straightforward, for example because quite diverse approaches exist towards the annotation of resources, ranging from part-of-speech tagging to the conceptual tagging of content.

The step from big data to big conclusions and decisions requires a certain level of transparency from the algorithms applied, since black box applications are not likely to be accepted as the basis for conclusions. Merely focussing on the effectiveness of tools on prediction tasks is only relevant within the narrow field of NLP. Sustainable scientific and societal impact of the tools offered can only be expected if the validity of analysis results can be assessed and explained to the communities that CLARIN aims to serve (Manovich, 2016; Nguyen et al., 2016). Clearly it may require the renewal of the quality evaluation frameworks proposed thus far if the CLARIN contribution to data science is to live up to the emerging standards for responsible data science.

4. Uptake

CLARIN's efforts towards uptake reflect and address the vision with two strands of core activities. First, a series of

surveys have been designed to evaluate the comprehensiveness and usability of CLARIN services and prioritise future development efforts. In parallel, training models have been developed that stimulate the uptake of CLARIN resources, tools and services by researchers from a wide range of disciplines in the humanities and social sciences where in addition to the teaching of how the technologies and services work, an important goal is also to stimulate methodological and paradigm shifts towards quantitative approaches, interdisciplinary research design, open science policies and transnational collaboration.

4.1. Surveys

4.1.1. User Base

Language in textual, spoken and multimodal form is a rich source of social and cultural data. CLARIN's language resources are therefore highly relevant for a wide range of disciplines and have a big potential for their reuse and repurposing. Especially the combination of multiple resources and the analytic tools available for multiple languages make CLARIN an enabler of comparative studies across regions, periods and languages. In addition to supporting research excellence, CLARIN is therefore also an infrastructure that is likely to have societal impact along various dimensions. To be able to cater for the various scenarios of use successfully, we are monitoring user experience with the infrastructure and identifying obstacles that still need to be addressed through focus groups.

Apart from continued efforts to support existing users, systematic analyses of relevant active research networks are being conducted to increase user engagement, uptake and dissemination for CLARIN.

4.1.2. Resource Tracking

Use and re-use of language resources is only possible when they can be easily found and understood by researchers. The most prominent types of resources that are offered by most consortia, such as corpora of parliamentary records, newspaper corpora, social media corpora, parallel corpora and oral history data that have a big potential for interdisciplinary and comparative research are therefore evaluated in terms of findability through the VLO, comprehensiveness of the metadata records, and usability in concordancing or text analytics environments. While we have observed a good coverage of the surveyed resources in the CLARIN infrastructure overall, there is still room for improvement. The level of inclusion into the CLARIN infrastructure differs across the data types we have analysed (i.e., while many corpora have been added to national repositories, they still cannot be identified through VLO directly due to lacking, idiosyncratic or vernacular names, keywords or description fields. Similarly, uneven granularity of the deposited resources, which ranges from complete archives to single-file records, makes navigation and use of the resources more difficult). The second group of major issues is the incomplete documentation (metadata). While information on corpus size is available for 67 (83%) of the identified corpora, information on the time period covered by the data, linguistic annotation and license information is less readily available. Of all the four resource types we have

¹²<http://www.responsible-datascience.org>

analysed so far, the metadata on the parliament corpora are generally of the best quality (i.e., the only information that was lacking was related to annotation in the case of two out of the 11 corpora in the infrastructure). By contrast, the metadata on the parallel corpora is much poorer: information on the level of textual alignment is available for 43 corpora (52%). For more information on the overviews of the CLARIN key resource ‘families’ see (Fišer et al., 2018).

4.2. Training

It is of strategic importance for CLARIN to offer training opportunities for researchers at all stages and from various backgrounds. Live events and on-line training materials not only demonstrate the functionalities of the available tools but also showcase projects that have successfully used the CLARIN infrastructure, address issues related digitally-enhanced and data-driven research design, contribute towards the advancement of the methodological apparatus and, last but not least, actively promote cross-disciplinary and cross-border collaboration. Co-operation with related research infrastructures is part of the strategy for capacity training. The wider aim is to ensure that the knowledge and skills needed to use and maintain the infrastructure are in line with the dynamics in the overall landscape of infrastructures for research.

5. Concluding Remarks

CLARIN was first conceived more than 10 years ago and has been steadily expanding, both in terms of members and in terms of disciplines for which services can be offered. We have outlined the CLARIN strategy and have shown how from the outset CLARIN has been in line with the more recently formulated principles for FAIR data. Plans for being compliant with the Responsible Data Science framework are being developed, as well as efforts to reinforce the multidisciplinary potential and uptake of CLARIN.

Acknowledgements

The work reported in this paper has been supported by the member countries and observers in the CLARIN ERIC, and it has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 676529 for project CLARIN-PLUS, and under grant agreement No. 654119 for project PARTHENOS. The authors owe a special merci to Maria Eskevich for her insights and practical support during the writing of this paper.

Bibliographical References

Arppe, A., Bruun, S., Koskeniemi, K., Lindén, K., Oksanen, V., and Westerlund, H. (2011). A report including Model Licensing Templates and Authorization and Authentication Scheme. Technical Report D7S-2.1, CLARIN ERIC, Utrecht, The Netherlands.

Broeder, D., Nathan, D., Strömquist, S., and Van Veenendaal, R. (2008). Building a federation of language resource repositories: the DAM-LR Project and its continuation within CLARIN. In *Sixth International Conference on Language Resources and Evaluation (LREC 2008)*. ELRA.

De Smedt, K., de Jong, F., Maegaard, B., Fišer, D., and Van Uytvanck, D. (2018). Towards an Open Science Infrastructure for the Digital Humanities: The Case of CLARIN. In *Proceedings of the Third International Conference on Digital Humanities in the Nordic Countries (DHN2018)*, Helsinki, Finland.

Fišer, D., Lenardič, J., and Erjavec, T. (2018). Meet CLARIN’s Key Resource Families. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. ELRA.

Goosen, T., Windhouwer, M., Ohren, O., Herold, A., Eckart, T., Āurčo, M., and Schonefeld, O. (2015). CMDI 1. 2: Improvements in the CLARIN Component Metadata Infrastructure. In *Selected papers from the CLARIN 2014 Conference, October 24–25, Soesterberg, The Netherlands*, pages 36–53. Linköping University Electronic Press.

Hinrichs, E. and Krauwer, S. (2014). The CLARIN Research Infrastructure: Resources and Tools for e-Humanities Scholars. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pages 1525–1531.

Maegaard, B., Van Uytvanck, D., and Krauwer, S. (2017). CLARIN Value Proposition. Public Report CLARIN-CE-2017-1093-P001, CLARIN ERIC, Utrecht, The Netherlands.

Manovich, L. (2016). The science of culture? Social computing, digital humanities and cultural analytics. *Journal of Cultural Analytics*.

Nguyen, D., Dođruöz, A., Rosé, C., and de Jong, F. M. G. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, 42(3):537–593.

OECD. (2017). Business models for sustainable research data repositories. OECD Science, Technology and Innovation Policy Papers 47.

Trilsbeek, P., Broeder, D., Van Valkenhoef, T., and Wittenburg, P. (2008). A grid of regional language archives. In *Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, pages 1474–1477. ELRA.

Van der Aalst, W. M. P., Bichler, M., and Heinzl, A. (2017). Responsible Data Science. *Business & Information Systems Engineering*, 59(5):311–313.

Van Uytvanck, D., Stehouwer, H., and Lampen, L. (2012). Semantic metadata mapping in practice: the virtual language observatory. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey. ELRA.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018).

Zinn, C. (2016). The CLARIN Language Resource Switchboard. In *Abstracts of the CLARIN Annual Conference 2016*, Aix-en-Provence, France.