



Common Language Resources and Technology Infrastructure

EUDAT interviews with research communities about uptake plans: CLARIN

This month EUDAT interviewed Dieter Van Uytvanck, the Technical Director for [CLARIN ERIC](#), to find out about CLARIN's ongoing work integrating its infrastructure with the EUDAT common data services. Dieter is based at Utrecht University where he is responsible for the construction and maintenance of [CLARIN's technical infrastructure](#). Dieter and the teams at the 31 CLARIN centres are working together with a group of central developers to connect language resources and tools so researchers in the humanities and social science can easily access and make use of these valuable resources.

Good morning, Dieter. For any readers who may not be familiar with CLARIN, would you please give a brief recap on what CLARIN is and does?

Certainly - [CLARIN](#) is the Common Language Resources and Technology Infrastructure in Europe. Its main aim is EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.



to provide scholars in the humanities and social sciences with easy access to digital language data (in written, spoken, video or multimodal form) and to advanced tools for finding, exploring, using, annotating, analysing or combining such data, no matter where it is located.

To this end, CLARIN is building a **networked federation of language data repositories, service centres and centres of expertise**, with single sign-on access for all members of the academic community in all participating countries. Tools and data from the different centres will be interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work. This CLARIN infrastructure is still under construction, but a growing number of participating centres are already offering access services to data, tools and expertise.

Oh, another term that is useful to know is “CLARIN ERIC” as you will hear that in relation to CLARIN quite often. A European Research Infrastructure Consortium (or ERIC) is a new type of international legal entity, established by the European Commission in 2009. Its members are governments or intergovernmental organisations. So the CLARIN Governance and Coordination body at the European level is CLARIN ERIC, which currently consists of fourteen Members and two Observers.

Thanks for the explanation about “ERIC”s, Dieter. CLARIN has been involved with EUDAT since early on, hasn’t it?

Yes, we’ve been part of EUDAT since the project began in 2011, so we have already started working with some of the EUDAT services.

That’s right - you mentioned that some of the CLARIN centres are already offering data services. So which EUDAT services has CLARIN adopted?

At the moment, the service that we are most involved with is **B2SAFE**. Various CLARIN centres are using it to perform safe replication of the language data they are hosting. **The University of Tübingen** (Eberhard Karls Universität Tübingen), the **LINDAT/CLARIN Centre for Language Research Infrastructure in the Czech Republic** (usually known as **LINDAT/CLARIN**) and the **Max Planck Institute for Psycholinguistics in the Netherlands** have all been using **B2SAFE** for a while now.

In addition, we are also **test-driving B2DROP** for internal data exchange and sharing. And **our community metadata has been integrated into B2FIND**.

And are there plans to incorporate more EUDAT services or to broaden the scope of your usage of these services?

Yes. After a call for interest among the CLARIN centres, **we have now about eight additional centres that would like to use B2SAFE**. In the next couple of weeks we will analyse the situation at each of these centres and create uptake plans tailored to the needs of each of the centres.

We also want to **harvest the B2SHARE metadata** that is related to language material and make it accessible via our search portal, the **Virtual Language Observatory**.

Finally we have further plans to look into the possibilities for using **B2DROP** in combination with linguistic analysis workflows.

So, if we look back, why did CLARIN engage with EUDAT? What did EUDAT offer that was so interesting for the CLARIN community and why was it useful?

EUDAT receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 654065.



Well, **we are glad that EUDAT is there because it provides a layer of infrastructure services on which we can build our own infrastructure.** For instance, **B2SHARE** makes it possible for “homeless” researchers and citizen scientists to deposit their resources into a place that is a good data repository with long-term preservation. There are other services – for example, the metadata integration into the infrastructure, or safe replication – that are all the kinds of services on which CLARIN and other research communities can build. Of course it makes sense to pool these resources and make sure there are good, reliable and stable services out there that can be used by everyone as, in the end, that will benefit all the research communities.

If you take a moment to imagine what things would have been like without EUDAT... did the collaboration solve problems or overcome barriers that CLARIN would not have been able to face on its own, or that would have been harder alone?

Being able to rely on the stability and scalability of the data centres in EUDAT really makes our life easier. Of course it would probably be possible to host such services ourselves but it would not be efficient and also we would not be able to take advantage of the high amount of existing expertise.

So, what practical improvements has the collaboration with EUDAT made for the CLARIN community? How has it helped the researchers on a daily basis? And what is the daily tangible impact that other CLARIN community members (such as the support and management staff) experience?

As with any infrastructure, the direct benefits might not always be directly visible to the end users. But the EUDAT services do ensure that, for example, **the data stored in repositories is well-preserved in the long term, that safe backups are in place, and that large research data sets can be safely and conveniently shared. Not having to worry about these issues definitely makes a huge difference for the CLARIN community.**

Thanks very much, Dieter. It is good to see CLARIN forging ahead so robustly in the process of taking up and using the EUDAT services.