A Bridge from EUDAT's B2DROP cloud service to CLARIN's Language Resource Switchboard

Claus Zinn
Seminar für Sprachwissenschaft
Universität Tübingen
`claus.zinn@uni-tuebingen.de`

Abstract

The Language Resource Switchboard is becoming a central pillar in the CLARIN infrastructure as it helps researchers to connect resources with tools that can process them in one way or another. Languages resources can be found in different places, and ideally, the switchboard is available nearby. Resources located at users' desktop computers can simply be uploaded to the switchboard, and resources found in CLARIN's Virtual Language Observatory can simply be sent to the switchboard by a simple click. Until now, the switchboard was only indirectly accessible for resources stored in the cloud. Here, users had to download a resource from their cloud storage to their desktop device before uploading it again to the switchboard to find applicable tools, which is tedious. In this paper, we describe how we linked EUDAT's B2DROP cloud service to the switchboard, giving users the capability to directly launch the switchboard with a resource from their B2DROP account. Also, we describe the usage of B2DROP to support the switchboard's back-end for intermediate file storage. The reported work makes a link to another infrastructure, and hence, facilitates and promotes the provision of complementary services to CLARIN members. We believe the cooperation between CLARIN and EUDAT to be of mutual benefit. On the one hand, our bridge makes the use of the generic cloud storage service from EUDAT more attractive to CLARIN members so that they are encouraged to use B2DROP rather than another cloud provider. On the other hand, it encourages EUDAT users to try out and profit from the CLARIN tool space, which in turn will challenge the tool providers to cope with an increased demand, and potentially new user requirements.

1 Introduction

The CLARIN Language Resource Switchboard (LRS) aims at bridging the gap between language-related resources and tools that can deal with these resources in one way or another. For a given resource, the LRS identifies all tools that can process the resource; users can then select and invoke the tool of their choosing. By invoking the tool, all relevant information about the resource is passed onto the tool, and the tool opens with most information gathered by the switchboard. This makes it easy for users to identify the right tools for their resource, but also to use the chosen tool in the most effective way possible.

The EUDAT Collaborative Data Infrastructure aims at providing services that seek to address the full life-cycle of research data. EUDAT's services include, among others, B2DROP (sync and exchange of research data), B2SHARE (store and share research data), B2FIND (find research data), and B2HANDLE (register your research data). B2DROP is directed at scientists to store and exchange data easily and to facilitate data synchronisation between cloud storage and desktop computers. EUDAT services are designed, built and implemented based on user community requirements. The CLARIN consortium contributes to EUDAT as one of the main communities in the Social Sciences and Humanities.

In this paper, we describe the use of B2DROP in the CLARIN Language Resource Switchboard. In the main use case, we anticipate an individual researcher or a small team of researchers to use B2DROP as cloud storage for language-related resources. The researcher(s) will want to work with and analyse the
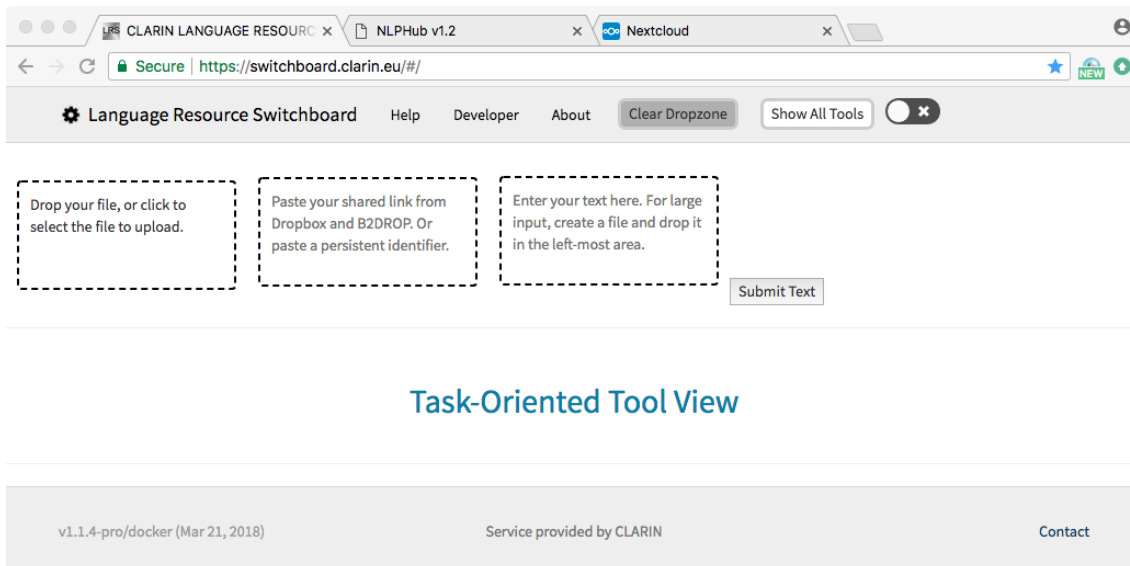
Figure 1: The CLARIN Switchboard's Main Page

resources using community-specific tools of the CLARIN tool space. From the B2DROP user interface, the researcher(s) will want to easily transfer a given resource to the LRS, which in turn suggests tools to process the resource. In a second use case, we describe the use of B2DROP as a technical vehicle for intermediate cloud storage, supporting a crucial aspect of the LRS' back-end implementation.

2 Background

2.1 The Language Resource Switchboard

The LR Switchboard (LRS) has been developed within the CLARIN-PLUS project (Zinn, 2016). The development of the LRS started as a browser-based stand-alone version.[1] Here, users simply upload their resource from their desktop machine to the browser, which is then temporarily stored on a file server at the Max Planck Computing and Data Facility (MPCDF)[2]. With the help of the Apache Tika library[3], the LRS then detects the resource's language and media type, and it uses this information to identify all tools registered with the LRS that can process the resource. The list of applicable tools is sorted along typical processing tasks (*e.g.*, tokenization, dependency parsing, named entity recognition) and shown to the user. When the user selects a tool from the list, the LRS constructs a URL that points to the tool's web location and also encodes the tool's parameters such as a reference to the storage location of the resource as well as the resource's language or an analysis id. The LRS then directs the browser to open the URL in a new browser tab. For a tool to be connected to the LRS, it must be reachable under the given base URL and capable of interpreting and processing all URL-encoded parameters passed during tool invocation. In particular, the tool will need to download the resource from the storage location that is encoded in the URL. The tool is then updating its internal model and graphical view accordingly.[4] Many tools leave users with no configuration options; here users simply press the start button to invoke the tool. Other tools have a richer interface where users can choose from many options before starting the tool.

The LRS has also been connected to the Virtual Language Observatory[5] (VLO), the main CLARIN site for searching language-related resources via CMDI-based metadata (Uytvanck et al., 2012). When users find a resource of interest in the VLO, they can start the LRS directly from VLO's resource viewer.
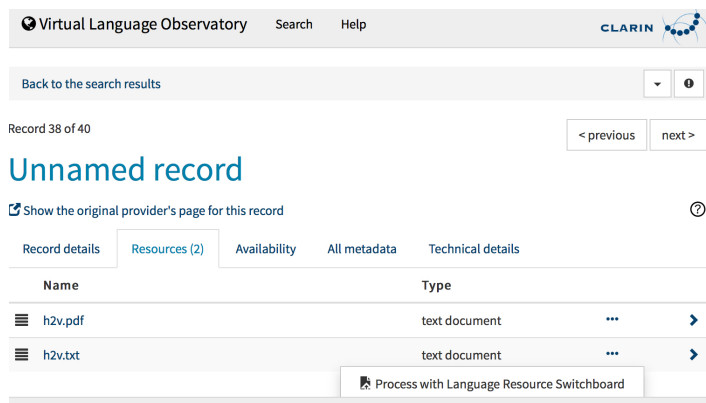
---

[1]See `http://switchboard.clarin.eu`.

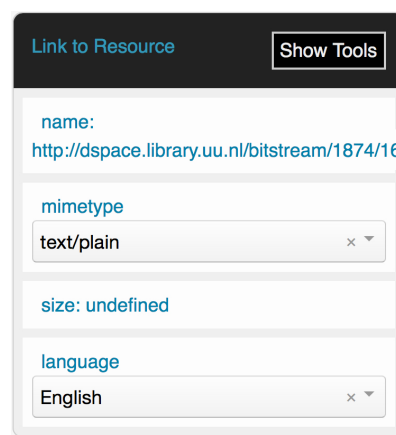[2]See `http://www.mpcdf.mpg.de`.

[3]See `https://tika.apache.org`.

[4]Many tools are capable of displaying the resource's content in a text-area, which reassures users that the resource has been successfully passed to the tool.
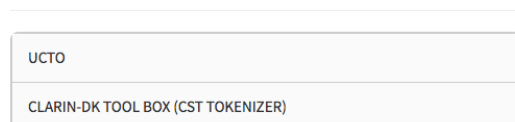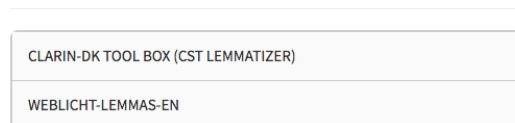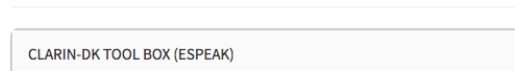
[5]See `http://vlo.clarin.eu`.

(a) The VLO – LRS Interface.

(b) The LRS Resource Pane.
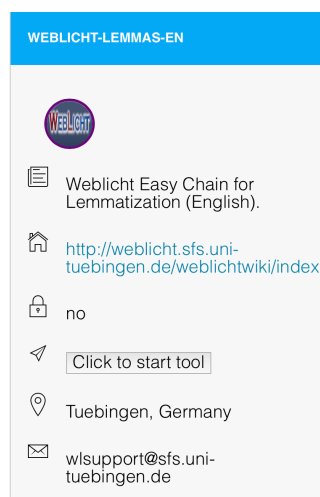
(c) The LRS Task Oriented View.

(d) The LRS Tool Detail View.

Figure 2: The LRS in Action.

Here, the VLO passes to the LRS data that is read from the CMDI metadata record of the resource: a URL pointing to the resource as well as information about the resource's language and media type. Given these pieces of information, there is no need for the switchboard itself do access the resource; the switchboard trusts the information given by the VLO, and does not derive media type and language itself.

Of course, all tools connected to the switchboard need access to the resource in order to process it. Here, resource providers must trust tool providers to handle their data with care. Switchboard users should be aware of data privacy issues, and when in doubt, they should not share sensitive data with the switchboard and its associated tools. Fig. 1 displays the main user interface of the CLARIN switchboard when it is called in stand-alone mode. At the centre, users have access to three input devices. A file drag & drop mechanism (left), a link drop mechanism (middle), and a text area, where users can enter or paste

textual data (right). In the first and third input method, the input is uploaded to a file storage server so that the tools connected to the LRS can access the data from this server. In the second approach, the data is already located in (Internet-accessible) cloud storage, so tools access the data from there.

As we have said, the switchboard can also be invoked from another software application, the CLARIN Virtual Language Observatory[6]. Consider the scenario where a linguist uses the VLO to find an English text which she then would like to investigate further. On the VLO search results page, the user can now click on the · · · area to invoke the LRS with this resource, see Fig. 2(a).[7] In a new browser tab, the LRS opens and shows a resource pane that depicts all relevant information about the resource, see Fig. 2(b). The user is free to correct this metadata[8], before clicking on 'Show Tools' to get to the task-oriented view, shown in Fig. 2(c). If the user, say is interested in the lemmatization task, she may wish to get more information about the two tools offered, in which case more detailed information about the chosen tool is given, see Fig. 2(d). When the user then clicks on 'Click to start tool', the chosen tool, here WebLicht, opens in a new browser tab. WebLicht obtains from the LRS a reference to the resource, the resource's mimetype and language as well as the chosen task. WebLicht opens with the predefined easy chain for lemmatization, loads itself the resource, and sets all relevant parameters so that the user is left to click on WebLicht's RUN command to start the processing chain. No further user action is required to parameterize WebLicht for this.

Status. At the time of writing (January 2018), a total of 60 browser-based applications and a dozen of web services have been connected to the switchboard. The tools are sorted along the tasks they achieve. Tools include: a chunker for Polish, constituent parsers for English and German, dependency parsers for Polish, German, Dutch, English, Slovenian, Croatian, and Serbian, named entity recognizers for German, English, Polish and Slovenian, shallow parsers for Polish, and tools for word sense disambiguation and sentiment analysis. There are also web services for the analysis of audio data such as runASR for the transcription of speech signals, and runMinni for the segmentation of speech data into phonetic segments. So far there has been an emphasis on tools for the processing of German, English and Polish texts but we strive to integrate tools that offer NLP tasks for other European languages. Note however, that the switchboard has a rather generic nature: given the media type and language of the resource, it suggests applicable tools that can process the resource. Once the language characteristics is set to generic, only the media type become a discriminating factor for tool selection.

2.2 The EUDAT service B2Drop

B2DROP is one of the main data services offered by the EUDAT Collaborative Data Infrastructure. The service is advertised as "a secure and trusted data exchange service for researchers and scientists to keep their research data synchronized and up-to-date and to exchange with other researchers"[9] (van de Sanden et al., 2015). B2DROP's base functionality competes with commercial services such as Dropbox[10], OneDrive[11], Google Drive[12], and many others. Standard functionality includes some free amount of cloud storage, cross-platform synchronization support, file versioning, and the ability to share files with other users. B2DROP's added value stems from its embedding in the EUDAT infrastructure. B2DROP is targeted at European researchers and guarantees that all research data stays on European servers.

Fig. 3 shows the role of B2DROP in the context of the other EUDAT Services. While B2DROP is meant to help researchers managing volatile research data (*e.g.*, draft research papers, experimental setups), it offers a bridge to B2SHARE[13] to publish such data once it has reached a final state. For this,

---

[6]It is planned to link the switchboard to CLARIN's Virtual Collection registry (https://clarin.ids-mannheim.de/vcr and Federated Content Search (https://spraakbanken.gu.se/ws/fcs/2.0/aggregator).

[7]For this, the VLO constructs a URL that points the switchboard, and which encodes (i) a reference to the resource (often a handle), the resource's media type and the resource's language.

[8]The Apache Tika Library usually yields good results, but sometimes the detection of a resource's media type or language is incorrect, for example, when the resource contains too little or obscure data.

[9]See https://eudat.eu/services/b2drop.

[10]See https://www.dropbox.com.

[11]See https://onedrive.live.com/about/en-us/.

[12]See https://www.google.com/drive/.
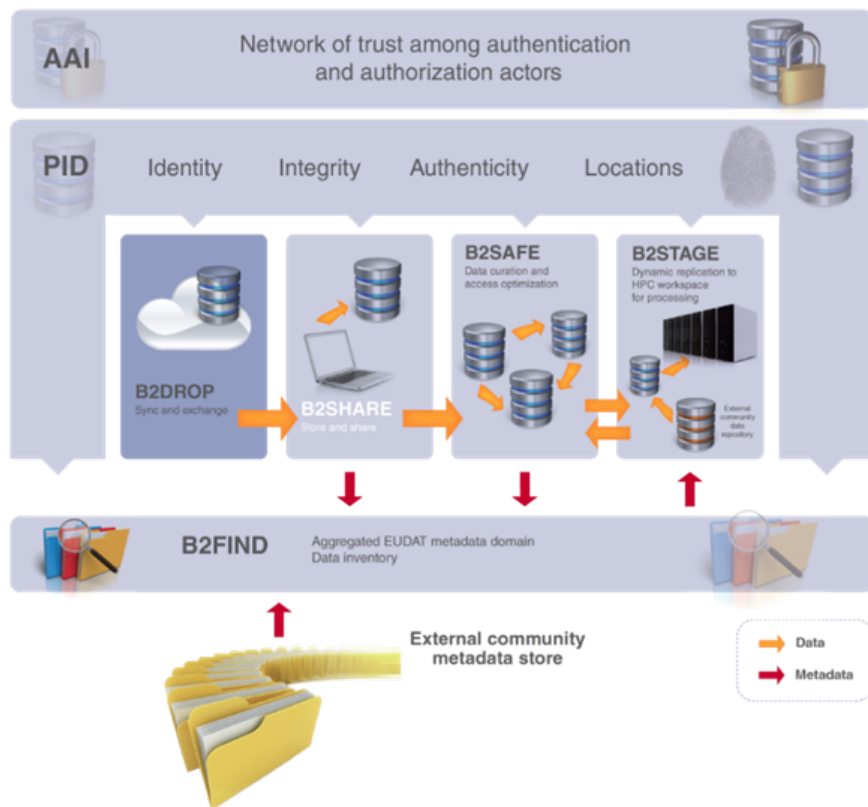
[13]See https://b2share.eudat.eu.

Figure 3: B2DROP in the B2 Service Suite

research data needs to be described with metadata and propagated to B2SHARE, where it also receives a persistent identifier. For public research data, the B2FIND service[14] can be used to search for the data using its metadata descriptions. EUDAT's services constitute a network of trust where users can be authenticated via B2ACCESS; here users can log in with an identity from a research organization they work for, or alternatively with their social identity such as their Google or ORCID[15] account.

Now, reconsider B2DROP is more detail. B2DROP allows individual users to store 20G of research data in the cloud, and to exchange such data with selected colleagues, over a given amount of time.[16] B2DROP is built upon Nextcloud, a fork of ownCloud[17], which is written in the PHP programming language.[18] B2DROP's major contribution to Nextcloud is the provision of a common EUDAT look-and-feel of the cloud's interface. Also, EUDAT developers have provided a Nextcloud plug-in that helps researchers to transfer resources from their personal B2DROP account to B2SHARE, where research data can be stored and preserved for the longer term. The official B2DROP service at `https://b2drop.eudat.eu` is hosted by the Forschungszentrum Juelich. With Nextcloud's software and B2DROP's extension being open-source, it is however possible to easily install, configure and operate a B2DROP server at a local host. For the following use cases, we have set up such a local B2DROP instance using a departmental server.

3 Integration Use Cases

We will discuss two scenarios where the use of B2DROP is beneficial for the LRS and its users. In the stand-alone version of the LRS, we propose replacing the existing file storage server with B2DROP. We

---

[14] See `http://b2find.eudat.eu`.
[15] See `https://orcid.org`.
[16] See `https://eudat.eu/services/userdoc/b2drop#UserDocumentation-B2DROPUsage-Documentdata`.
[17] See `https://nextcloud.com/` and `https:owncloud.org`.
[18] See `https://owncloud.org/blog/owncloud-and-php/`.

also suggest complementing the existing usage of the LRS (its use in stand-alone mode or via invocation from the VLO) with a cloud-based usage. While the first integration is of a purely technical nature (it changes the switchboard under the hood), the second integration offers a more visible usability benefit for B2DROP and switchboard users. We have implemented prototypes for both scenarios.

3.1 Using B2DROP as Alternative to the MPCDF server

When users of the stand-alone version of the LRS upload a resource, this is temporarily stored at an external file storage server at MPCDF. This is necessary as all tools connected to the switchboard need web-based access to the resource. The existing server has two drawbacks: the amount of available disk space is limited, and there is little access control in place permitting users aware of the server address to view and access all uploads. To address privacy concerns, it is necessary to better restrict access to file uploads. For this, we have replaced LRS' usage of the MPCDF file storage server with B2DROP:

1. an instance of B2DROP has been installed on a departmental server at the University of Tübingen;

2. a designated B2DROP user 'switchboard' has been registered;

3. when a user uploads a resource to the LRS, the resource is transferred to the B2DROP account of the designated user;

4. using B2DROP's API, the 'switchboard' user creates a shared link for the resource with a link expiration date set to 24 hours;

5. any tool invoked from the switchboard is given access to the shared link to access the resource.

Note that the entire content of the switchboard's B2DROP account is only visible to the 'switchboard' user. A shared link gives only access to the resource that is associated with the link; moreover, the link expires within a short time frame. This is a vast improvement with regard to the MPCDF solution.

A future version of the LRS may allow users with an existing account at `https://b2drop.eudat.eu` to use their own B2DROP cloud storage rather than the generic designated 'switchboard' account. In the meantime, we have also developed an input facility (see middle box in Fig. 1) where users can paste their shared links from their B2DROP or Dropbox account into the LRS.

3.2 Creating a Bridge between B2DROP and the Language Resource Switchboard

We have also created a GUI-based bridge from B2DROP to the Language Resource Switchboard. The inverse direction aims at supporting researchers who manage (part of) their language resources in the cloud, in part, because they need to easily share resources with other researchers (using, for instance, shared links). Here, using the switchboard's drag & drop mechanism would feel rather clumsy: users would need to copy the resource from the cloud to their local desktop, and then open the file explorer to drop the resource into the switchboard (the left-most dotted area in Fig. 1). To improve the usability aspect, we have built a switchboard plugin for B2DROP, which is depicted in Fig. 4. The "Files" view (see circled 1) shows all the files (including directories) stored by the user. Files can be shared with other researchers in which case a "Shared" tag is associated with the resource, see (2), together with a URL pointing to the resource, *e.g.*, `http://weblicht.sfs.uni-tuebingen.de/nextcloud/s/0qeeLnfsj3urgik`. Researchers can give this URL to other researchers so that they get access to the resource as well. Note that each file or directory is associated with a triple dot icon, see (3). When users click on the dots, a menu with actions connected to the resource opens, see (4). This menu has been extended with the action "Switchboard". When users select this option for a resource, the LR switchboard opens in a new browser tab, capable of processing the shared link created by the user.

Implementation and Installation Details

The developers of Nextcloud praise its open architecture; Nextcloud's functionality is extensible via a simple but powerful API for applications and plug-ins ("apps"). One such app is "b2sharebridge", which allows B2DROP users to share their resources via EUDAT's B2SHARE service. We have taken the
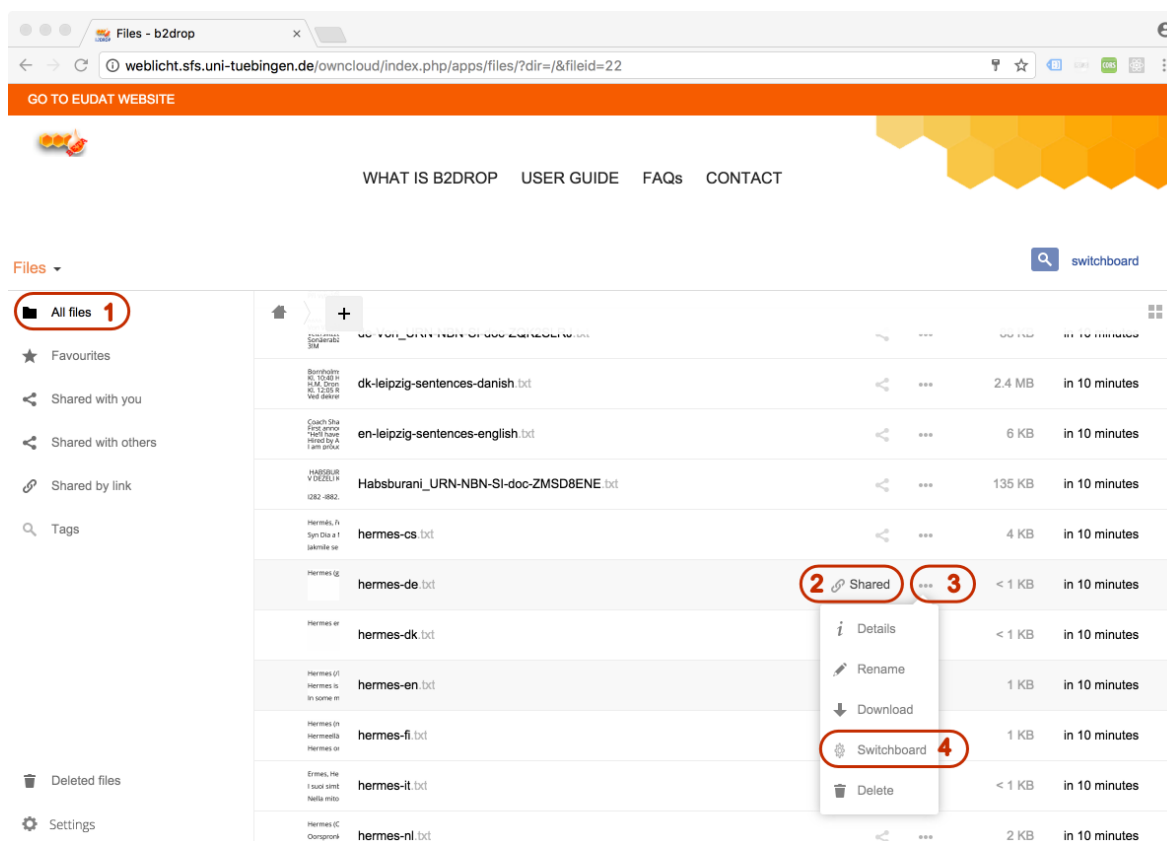
Figure 4: Bridge between B2DROP and the LRS

"b2sharebridge" code as example for the "lrswitchboardBridge" and followed the Nextcloud developer manual.[19] Most of the work required the coding of Javascript code that (i) adds the new item "Switchboard" to the pop-up menu that associates file actions with a given resource (*e.g.*, "Details", "Rename"); and (ii) implements an action handler for the new action item. The handler creates a new XML HTTP request; here, a URL is constructed that encodes the web location of the LR switchboard, information about its caller, and the shared link to the resource in question. The plug-in then opens the URL[20] in a new browser tab. On the LRS side, we have added code that detects from the invocation URL the caller ("b2drop"), downloads the resource from the shared link, and determines the media type and language of the resource. Subsequently, the LRS proposes applicable tools to process the resource. The tool selected for invocation downloads the resource using B2DROP's shared link.

The installation of the "lrswitchboardBridge" plug-in must be performed by the administrators of the B2DROP/Nextcloud server, following the standard procedure for plug-in installs.

4   Related Work

CLARIN has started to make use of EUDAT's infrastructure in several ways.[21] To give CLARIN users easy access to the EUDAT infrastructure, the CLARIN identity provider (IdP) has been integrated with B2ACCESS. Now, users can use their CLARIN account to access EUDAT services. On the repository level, there is an increasing uptake of B2SAFE, EUDATs infrastructure for data replication and backup. At the time of writing, the repository systems of five CLARIN centres (CLARIN-AT, Meertens, Tübingen, TLA, and Språkbanken) have been been directly integrated with B2SAFE. The integration is ongoing at three other CLARIN centres. So far, more than 100TB of data is managed with B2SAFE.

---

[19]See `https://doc.owncloud.org/server/9.0/developer_manual/app/`.
[20]For   example,   `https://switchboard.clarin.eu/#/b2drop/http://weblicht.sfs.uni-tuebingen.de/nextcloud/s/0qeeLnfsj3urgik/download`.
[21]For a full account, see the CLARIN-PLUS Deliverable D4.2 (Zinn et al., 2017).

B2STAGE is advertised as "a reliable, efficient, light-weight and easy-to-use service to transfer research data sets between EUDAT storage resources and high-performance computing (HPC) workspaces. To the author's knowledge, the functionality has not yet been taken up by the CLARIN community. There has been considerable progress, however, in integrating the WebLicht workflow engine (Hinrichs et al., 2010) with EUDAT's experimental Generic Execution Framework service (Dima et al., 2015). The integration allows users to bring the language processing tools integrated into WebLicht to an execution environment that also hosts the data, hence allowing language processing close to the data. The new development enables researchers to use WebLicht for both sensitive and big data (Zinn et al., 2018).

There is ample potential for CLARIN to profit from EUDAT and its generic cloud storage and computing services. An increasing number of users make use of cloud computing[22], and many members of the CLARIN community use Dropbox, Microsoft's OneDrive, Google Drive, or another commercial provider to manage their research data. B2DROP offers a non-commercial alternative: it is based on non-proprietary, open-source software, all data is stored on European-based servers, and EUDAT's Terms of Use are user-friendlier than those of the commercial providers.

The sharing of services across infrastructures is of mutual benefit for both EUDAT and CLARIN. EUDAT increases its user base as any new "customer" strengthens the role of EUDAT as central infrastructure service provider. Also, CLARIN avoids to duplicate and maintain infrastructure that is available elsewhere. There is, however, a natural tension between generic infrastructure providers such as EUDAT and community-specific infrastructure providers such as CLARIN. For this, reconsider the switchboard plugin, which is currently being deployed and tested on CLARIN-hosted development servers, and which EUDAT has started testing on a EUDAT-hosted development server. Upon successful testing (and pending an agreement with CLARIN to ensure a long-term software support), EUDAT is likely to offer the plugin for all its B2DROP users. In this respect, note that the switchboard's basic functionality is quite generic: it helps users to connect their resources with tools that can process them. So it might well be that the switchboard's current tool range expands beyond language-related tools, especially when B2DROP users from other communities expect their community-specific tools to be connected to the switchboard.

From the CLARIN perspective, CLARIN-based B2DROP users would profit from a number of community-specific adaptations. The easiest change would be the inclusion of graphical elements of CLARIN's corporate identity in B2DROP's branding (its Nextcloud theme). This would give B2DROP's GUI a more CLARIN-like look & feel. Also, EUDAT should support the connection of B2DROP with OpenCloudMesh[23], a framework for federated cloud sharing. This would enable users to seamlessly share files with each other, no matter whether they reside on the same (within B2DROP), or on a different cloud server.

A more complex issue is the provision of user delegation services for (trusted) CLARIN applications. Here, we believe that tools connected to the switchboard should be allowed to read from and write to a user's B2DROP cloud space (with the permission of an authenticated user). In fact, CLARIN-D has tested a prototype implementation based on the UnityIDM authentication service[24] and its SAML/OAuth2 bridge with success, see (Blumtritt et al., 2014).

In any case, switchboard users should be aware that data travels through the network. Tools invoked via the switchboard need access to the resources to process them. Including the EUDAT network into CLARIN's network of trust would certainly help the case. Another step is to make all tools connected to the switchboard part of the trusted network, but this requires considerable future work, especially with regard to user delegation issues.

---

[22]A eurostat report claims that 21 % of EU enterprises used cloud computing in 2016, mostly for hosting their e-mail systems and storing files in electronic form (Giannakouris and Smihily, 2016). In a related statistics, it is reported that in 2014, one in five EU citizens aged 16-74 saved files on internet storage space. Most cloud users appreciated the ease of accessing files from several devices or locations (Seybert and Reinecke, 2014). Quite likely, such numbers will be higher for academic institutions and individual researchers.

[23]See `https://oc.owncloud.com/opencloudmesh.html`.

[24]See `http://www.unity-idm.eu`.

5 Discussion and Conclusion

In this paper, we have sketched two uses of the EUDAT infrastructure service B2DROP for the CLARIN Language Resource Switchboard. The first use of B2DROP improves the back-end of the LRS with the provision of a file storage server that strengthens the privacy aspect of file uploads. File uploads are only accessible for users with access to the shared link, and such links expire after a short time frame.

We consider the second use case more important. So far, the services of the LRS have been at the users' fingertips for personal resources (the stand-alone version of the LRS with file uploads) and for resources advertised in the CLARIN Virtual Language Observatory. With the latest addition, the LRS is now easily accessible for teams of researchers sharing a cloud storage. Resources are uploaded to a Nextcloud-based server, and when a resource is marked as shared, a user can invoke the switchboard with a single click. Once directed at the LRS, users then invoke the tool of their choice also with a single click. We believe that the Nextcloud-based access to the LRS is a feature many users will want to have.

The author is in contact with the administrators of the B2DROP service at `https://b2drop.eudat.eu` to get the switchboard plug-in installed for all B2DROP users. While the technical installation is itself simple, issues regarding the long-term support for the plugin need to be addressed (*e.g.*. who updates the plugin when B2DROP is updated the next version of Nextcloud?). Here, some kind of formal agreement between CLARIN and EUDAT needs to be drawn. B2DROP has been very forthcoming so far, given that the CLARIN community is only one of many communities that take part in the EUDAT project. Having the "lrswitchboardBridge" plug-in enabled by the official B2DROP administrators would bring the CLARIN and EUDAT communities closer together and contribute to service compatibility across digital research infrastructures. With the new bridging service, CLARIN researchers would get the incentive to use B2DROP (and hence, associated EUDAT services such as B2SHARE). As a consequence, commercial services with no such benefits would loose their attractiveness. If the "lrswitchboardBridge" were supported by EUDAT, then European researchers using B2DROP would get easy access to the CLARIN tool space via the LRS. This would significantly increase the usage of many tools across communities, which in turn would challenge tool developers to cope with the new demand, and probably, with new user requirements.

From a wider perspective, the CLARIN community needs to reconsider and potentially adapt its infrastructure pillars. Clearly, many researchers will want to use cloud computing to store research data. Here, EUDAT's B2DROP service helps those researchers to manage and share their data in the cloud. Although B2DROP's adaptation is minimal (it rebrands Nextcloud with the provision of a GUI theme), costs for hosting the service and for user support need to be taken into account. In this paper, we hinted at a tighter integration of EUDAT services within CLARIN. We believe that cloud computing is becoming increasingly important to the CLARIN community. Rather then re-implementing and providing such services within CLARIN, it might be less expensive to use existing generic services of other infrastructures, and have them adapted to community-specific needs whenever possible. The use of components from other infrastructures provides a good opportunity to revisit the overall design rationale of the CLARIN infrastructure and to reconsider the appropriateness of certain components. This assumes, of course, that EUDAT's follow-up project, the EOSC-Hub project[25] continues developing, maintaining, and supporting B2DROP and related services.

References

[Blumtritt et al. 2014] Jonathan Blumtritt, Willem Elbers, Twan Goosen, Marie Hinrichs, Wei Qiu, Mischa Sall, and Menzo Windhouwer. 2014. User Delegation in the CLARIN Infrastructure. *Linkoping Electronic Press*, (116):14–24.

[Dima et al. 2015] Emanuel Dima, Christian Pagé, and Reinhard Budich. 2015. D7.5.2: Technology Adaptation and Development Framework (final). Technical report, EUDAT deliverable. Available at

---

[25]See `https://www.egi.eu/about/newsletters/introducing-the-eosc-hub-project/.`

    `https://b2share.eudat.eu/api/files/4cc8cf0e-99a2-4b6b-981a-0ffcd870af19/`
    `EUDAT-DEL-WP7-D7%205%202-Technology_adaptation_and_development_`
    `framework-2.pdf.`

[Giannakouris and Smihily 2016] Konstantinos Giannakouris and Maria Smihily. 2016. Cloud computing - statistics on the use by enterprises. Technical report, eurostat - Statistics Explained. ISSN 2443-8219, available at `http://ec.europa.eu/eurostat/statistics-explained/index.php/Cloud_` `computing_-_statistics_on_the_use_by_enterprises`.

[Hinrichs et al. 2010] Erhard Hinrichs, Marie Hinrichs, and Thomas Zastrow:. 2010. Weblicht: Web-Based LRT Services for German. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (System Demonstrations)*.

[Seybert and Reinecke 2014] Heidi Seybert and Petronela Reinecke. 2014. Internet and cloud services - statistics on the use by individuals. Technical report, eurostat - Statistics in focus 16/2014. SSN:2314-9647, available at `http://ec.europa.eu/eurostat/statistics-explained/index.php?title=` `Internet_and_cloud_services_-_statistics_on_the_use_by_individuals`.

[Uytvanck et al. 2012] Dieter Van Uytvanck, Herman Stehouwer, and Lari Lampen. 2012. Semantic metadata mapping in practice: the virtual language observatory. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, pages 1029–1034. European Language Resources Association (ELRA).

[van de Sanden et al. 2015] Marie van de Sanden, Christine Staiger, Claudio Cacciari, Roberto Mucci, Carl Johan Hakansson, Adil Hasan, Stephane Coutin, Hannes Thiemann, Benedikt von St. Vieth, and Jens Jensen. 2015. D5.3: Final Report on EUDAT Services. Technical report, EUDAT. Available at `http://hdl.handle.` `net/11304/2433d23a-6079-49a6-9010-ca534f6e348d`.

[Zinn et al. 2017] Claus Zinn, Twan Goosen, Marie Hinrichs, Emanuel Dima, Willem Elbers, Dieter Van Uytvanck, Dirk Goldhahn, Thorsten Trippel, and Josef Misutka. 2017. Joint infrastructure services. Technical report, CLARIN-PLUS Deliverable D4.2. Available at: `https://office.clarin.eu/v/` `CE-2017-0985-CLARINPLUS-D4_2.pdf`.

[Zinn et al. 2018] Claus Zinn, Wei Qui, Marie Hinrichs, Emanuel Dima, and Alexandr Chernov. 2018. Handling big data and sensitive data using EUDAT's Generic Execution Framework and the WebLicht workflow engine. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*. European Language Resources Association (ELRA).

[Zinn 2016] Claus Zinn. 2016. The CLARIN language resource switchboard. In *Proceedings of the CLARIN Annual Conference*. CLARIN ERIC. Available at `https://office.clarin.eu/v/` `CE-2016-0917-Proceedings-CAC-2016.pdf`.