

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

We describe the usage of EUDAT's B2DROP cloud service to increase the usability, visibility and attractiveness of the CLARIN Language Resource Switchboard.

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 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. It aims at easily connecting users and their resources with the tools that can process them (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 library³, 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

¹See <http://weblicht.sfs.uni-tuebingen.de/clrs/>.

²See <http://www.mpcdf.mpg.de>.

³See <https://tika.apache.org/>.

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. The tool is then updating its internal model and graphical view accordingly. Ideally, the user is left with little other configuration options, and can request the tool to start processing the resource.

The LRS has also been connected to the Virtual Language Observatory (VLO)⁴, 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. 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.

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”⁵ (van de Sanden et al., 2015). 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.⁶ B2DROP is built upon Nextcloud, a fork of ownCloud⁷, which is written in the PHP programming language.⁸ 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 B2SHARE⁹, where research data can be stored and preserved for the longer term. The official B2DROP service at <https://b2drop.eudat.eu> is currently 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 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. 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, it 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;

⁴See vlo.clarin.eu.

⁵See <https://eudat.eu/services/b2drop>.

⁶See <https://eudat.eu/services/userdoc/b2drop#UserDocumentation-B2DROPUsage-Documentdata>.

⁷See <https://nextcloud.com/> and <https://owncloud.org>.

⁸See <https://owncloud.org/blog/owncloud-and-php/>.

⁹See <https://b2share.eudat.eu>.

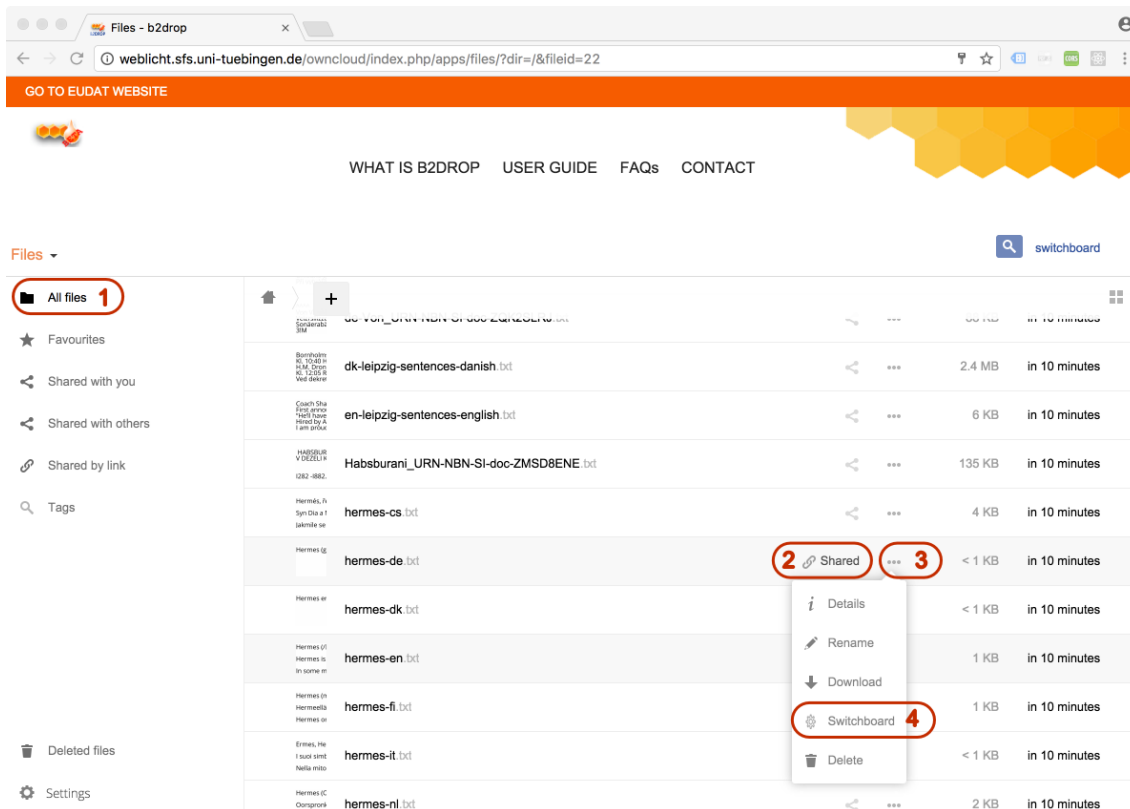


Figure 1: Bridge between B2DROP and the LRS

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 will 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, or to paste shared links from their B2DROP account into the LRS.

3.2 Creating a Bridge between B2DROP and the Language Resource Switchboard

We have also created a bridge from B2DROP to the Language Resource Switchboard. Fig. 1 depicts the user interface of B2DROP with our LRS plug-in. In the “Files” view (see circled 1), it 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 (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 "b2sharebridge" code as example for the "IrswitchboardBridge" and followed the Nextcloud developer manual.¹⁰ 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 URL¹¹ 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 installation of the "IrswitchboardBridge" plug-in must be performed by the administrators of the B2DROP/Nextcloud server, following the standard procedure for plug-in installs.

4 Discussion and Conclusion

In this abstract, 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.

It is an open question, however, whether the administrators of the B2DROP service at <https://b2drop.eudat.eu> will install the plug-in for all its users, given that the CLARIN community is only one of many communities that take part in the EUDAT project. Having the "IrswitchboardBridge" 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 lose their attractiveness. If the "IrswitchboardBridge" were supported by EUDAT, then European researchers using B2DROP will get easy access to the CLARIN tool space via the LRS. This will significantly increase the usage of many tools across communities, which in turn will challenge tool developers to cope with the new demand, and probably, with new user requirements.

As an intermediate step (also for further testing and development), the CLARIN community may well decide to administer its own CLARIN-specific, Nextcloud-based cloud with the LRS plug-in.

References

- D. Van Uytvanck, H. Stehouwer, and L. Lampen. 2012. Semantic metadata mapping in practice: the virtual language observatory. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*. European Language Resources Association (ELRA), pages 1029–1034.
- M. van de Sanden, C. Staiger, C. Cacciari, R. Mucci, C. J. Hakansson, A. Hasan, S. Coutin, H. Thiemann, B. von St. Vieth, and J. Jensen. 2015. D5.3: Final Report on EUDAT Services. Technical report, EUDAT. Available at <http://hdl.handle.net/11304/2433d23a-6079-49a6-9010-ca534f6e348d>.
- C. 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>.

¹⁰See https://doc.owncloud.org/server/9.0/developer_manual/app/.

¹¹For example, <http://weblicht.sfs.uni-tuebingen.de/clrs/#/b2drop/http://weblicht.sfs.uni-tuebingen.de/nextcloud/s/0qeeInfsj3urgik/download>.