M6.7 Ingest points of two TDRs integrated with EOSC-hub data transfer services

Milestone 6.7 worked towards providing two Trusted Digital Repositories to support European Research Infrastructures. This task was undertaken using two different technology stacks to ensure flexibility and enhance our ability to re-shape these services as future use-cases are encountered.

The CINES instance is now part of the EOSC-hub catalogue and is openly available for usage, the DANS instance has been delayed, due to an unforeseen B2SHARE/invenio upgrade. However, the work at DANS provides a very promising proof of concept that will help with uptake from other possible service providers.

The two models benefit not only the end users and communities but also the service providers. A great deal of flexibility is present both for users and providers. Much value has been provided by this milestone, one eTDR has been realized and a proof of concept for an alternative model has been provided which opens the way for more TDRs in the future.

A more extensive description of the two TDR deployments, their technical implementations and their strengths follows.

CINES instance of eTDR in Production:

- Is added to the EOSC-hub catalogue https://marketplace.eosc-portal.eu/services/etdr-european-trusted-digital-repository
- Validates the concept of eTDR following the Herbadrop pilot with the community of digitized herbarium specimens : https://eudat.eu/news/eudat-supports-the-analysis-and-long-term-preservation-of-herbarium-specimen-images-watch-the
- Uses existing reliable EUDAT services : B2SAFE (file transfer), B2HANDLE (PIDs) and B2FIND (access) ;
- Demonstrates the feasibility of building new services by integrating front-ends available in the EOSC-hub portfolio ;
- Ensures that digital information remains findable, accessible, interoperable and reusable for an unlimited period of time.

SUCCESS points CINES:

- Service available in EOSC-hub marketplace;
- One successful pilot, approx. 25 TB archived.

DANS instance of eTDR in development:

- Demonstrates the concept of a new front-end for an eTDR: B2SHARE

  o Proofs and stimulate the potential to increase the use of B2-services as ingest point to an eTDR;

  o DANS in contact with developers of B2SHARE; Need by B2SHARE to have a preservation functionality as intended by this milestone. Connection point to DANS

eTDR fulfills this need. The implementation will stimulate the preservation of data for reuse in further research.

- Solution built in two steps, first Proof of Concept, second Final implementation

  o Plan to reuse generic solution (XSLT metadata transformation, automatic bag generation, etc.), developed by DANS, to connect B2SHARE to DANS eTDR;

  o Proof of Concept of this idea created and finished in 2019, in cooperation with development team of B2SHARE;

  o Unfortunately by unforeseen long-time absence of the developer of B2SHARE further development was postponed till 2020;

  o Therefore aim was to use second year (2020) to implement and publish a production ready version of this solution.

- DANS developed a proof of concept during 2019. Proof of working functionality and solution can be found on Git-hub. Code published on: https://github.com/ekoi/b2share-dtap;

- Relies on SWORD2 bridge from B2SHARE to DANS instance of eTDR

  o DANS eTDR implemented a SWORD2 api and protocol for data transfer and ingestion, SWORD is a communication protocol between repositories;

  o B2SHARE is based on Invenio2 and supports SWORD2 for data transfer, this matched solution build by DANS. This was initial plan to use this interface for the transfer of data to the DANS eTDR;

  o DANS developed a generic bridge (see proof of concept), which offers an interface from any repository to the interface of the DANS eTDR using the SWORD2 protocol; this bridge should have formed the basis for the final implementation of the eTDR. This bridge handles metadata mapping, data preparation and data transfer to the DANS eTDR.

- B2SHARE to be migrated to Invenio3 in 2020 by surfSARA as provider of B2SHARE in the Netherlands

  o In 2020 an upgrade from Invenio2 to Invenio3 was on the surfSARA B2SHARE development roadmap. This upgrade was not foreseen during the start of the project and was unfortunately only communicated early 2020;

  o Impact of the B2SHARE upgrade to Invenio3 is that the SWORD2 interface will not be supported and will be replaced by SWORD3 support, thus not possible to connect directly to DANS eTDR via this protocol as initially planned. SWORD3 protocol follows different standards than SWORD2, therefore DANS needs to upgrade its service to SWORD3 to establish a connection between B2SHARE and DANS eTDR;

- Impact on development as code needs to be migrated to SWORD3

  o DANS started innovating their long term preservation service early 2020 and will take till 2022;

  o The upgrade of B2SHARE causes DANS to investigate the upgrade of the SWORD2 interface to SWORD3 in light of EOSC-hub project. During 2020 DANS aims to work on investigating SWORD3 for further implementation during the innovation upgrade of their eTDR in 2021/2022;

- o DANS focuses again on a generic solution. Therefore, the solution will not only fit the DANS infrastructure, but will also be easy to implement in other repositories. Main focus is on reusability of the solution to make more eTDRs available in the EOSC-hub service catalogue (marketplace);

- o Solution will be published as open source via Git-hub to other eTDRs outside EOSC_hub too.

SUCCESS points DANS:

- Working generic proof of concept for SWORD2, offered to the public via Git-hub. Any service can use this solution to connect to any eTDR that supports data transfer and ingest via SWORD2;

- Due to the upgrade to SWORD3 on B2SHARE side, DANS needs to investigate the possibilities to upgrade from SWORD2 to SWORD3 within its own service starting in 2020, so the connection from B2SHARE to the DANS eTDR can become available in a later stage;

- DANS focuses on a generic solution, just as with the proof of concept, which will be offered via Git-hub to other services and eTDRs. Both proof of concept and future solution fulfills in this case the need of B2SHARE for a long term preservation functionality;

- Next to this other eTDRs can use this generic solution, whereby B2SHARE data can be deposited in multiple eTDRs (not only DANS eTDR). Therefore the solution will eventually increase the use of and availability of eTDRs within the EOSC_hub service catalogue.

Overall, one instance of eTDR is fully functionnal. Another has demonstrated its great potential via a proof of concept of the working functionality, but deployment in Production has been postponed due to unforeseen illness and requests for upgrades on the B2SHARE side.