

Addressing sustainable long-term preservation wall for scientific data: the European Trusted Digital Repository (ETDR) service

Tuesday, 9 October 2018 16:45 (15 minutes)

Digital data preservation should be a key feature of all research projects. Some research data are unique and cannot be replaced if lost or destroyed; scientific results can be considered as trustworthy only they refer to verifiable data.

In addition to a bit-stream preservation service that ensures data integrity technically, Trusted Digital Repositories (TDRs) are providing a quality of services that preserve information over a long period of time. This requires extra and certified capabilities in the area of curation, metadata, file formats, long-term preservation, diverse data access levels, data quality assessment based on the FAIR principles, etc.

During EUDAT and EUDAT2020, TDRs ever used to preserve research data have been assessed and a European generic, innovative and large added-value service has been developed: the European Trusted Digital Repository (ETDR). This constellation of TDRs and other service providers can offer to scientific communities some important securities on data reuse enabling. The three main guarantees taken by the ETDR are on:

- data integrity (i.e. bit-stream preservation),
- hardware and software readability (i.e. file formats, emulation...),
- and understandability of the information over time (i.e. metadata, information classification ...).

The ETDR front-office will offer access to EUDAT, EGI and OpenAIRE distributed data storage services. Data that needs to be preserved for the long-term will be automatically ingested into the distributed ETDR back-office infrastructure. In addition, front-ends can be featured in discipline-specific research infrastructures or researcher deposit platforms that do not yet have access to certified TDR service. The ETDR provides also customer support on data management including data management planning and requirements for long-term preservation.

Within EUDAT2020, Herbadrop and ICEDIG, the ETDR has ever been used by about ten national institutions that belong to DiSSCo, the e-infrastructure for natural sciences. During the EOSC-Pilot and EUDAT2020 projects, a three partners association (CERN, CINECA, CINES) has demonstrated genericity, scalability and accessibility of the ETDR architecture.

The next years would be at the convergence of increasing the research community's number and the ETDR network expansion.

Type of abstract

Presentation

Summary

Sharing and enabling reuse of scientific data is one of the great challenges for the next decade and the EOSC. Within the EUDAT and EUDAT2020 projects, Trusted Digital Repositories (TDRs) were assessed and a new approach was used to develop a sustainable certified long-term data preservation service: the European Trusted Digital Repository (ETDR). In EOSC-Hub, this service –which the back-office requires expertise in data and metadata stewardship and curation that goes beyond simple data management services –will be expanded.

Primary author: MASSOL, Marion (CINES)

Presenter: MASSOL, Marion (CINES)

Session Classification: Data Management Services

Track Classification: Area 5. Digital Infrastructures for EOSC and/or EDI