

RDM: A library perspective of versioning, curating and archiving research data from diverse domains

Tuesday, 9 October 2018 16:30 (15 minutes)

Libraries are the vanguards for RDM and digital curation. However, beyond archival preservation, versioning and digital curation of research data adds value to knowledge assets insofar that these can be extended across domains to create services that are useful to the research community. At Bielefeld University, the DFG-funded Conquaire project, a collaboration between CITEC and the Bielefeld University library, has created a generic RDM framework that ensures research data quality using continuous integration (CI) principles in order to ease the process of publishing research data to PUB, our institutional repository which is based on the free and open-source LibreCat software.

The Conquaire RDM system (RDMS) automates the analytical reproducibility process by unobtrusively monitoring their research data stored within a GitLab repository to validate its data quality for CSV files. Researchers receive automated quality assessments via email whenever they upload research data into their repository that is automatically monitored using the inbuilt GitLab CI.

Furthermore, the continuous integration principle standardizes technology (platforms and tools) which enhances the cross-domain data interoperability in an RDM service. A curated digital dataset that validates standardized formats will mitigate digital obsolescence, thereby making the research data accessible, reusable, and archivable for users indefinitely.

Among research artifacts, the software source code used for the analysis being an integral part of a research project can be considered to be a form of data –research publications without the code used to process and visualise the research data cannot be analytically reproduced. The source code also needs to be properly versioned, curated and archived in order to fulfill the FAIR (Findable, Accessible, Interoperable and Reusable) data principles. Currently, in addition to the data quality framework, we are in the process of implementing a generic CI system that automates and aids the data validation system based on the technical stack used by the partner groups.

In order to understand the nine research partner groups' software toolkits and data analysis process, we undertook independent reproducibility experiments (ReX) that entailed analytically reproducing one result from a paper already published by these groups.

Our research experience during the ongoing collaboration with the case study partners has highlighted the technical challenges that diverse research projects throw up during the process of creating a generic data quality framework. These range from finding common document formats to analyse tools used among the various research groups partnering in the Conquaire project. Finding a balance between this diversity (both technical and data-wise) without disturbing the existing workflow of each research group has thrown up cross-domain challenges that need to be addressed.

Type of abstract

Presentation

Summary

We present Conquaire, a generic RDM framework that ensures research data quality using continuous integration (CI) principles.

Primary author: Mrs AYER, Vidya (Bielefeld University)

Co-authors: Mr PIETSCH, Christian (Bielefeld University); Mr WILJES, Cord (Bielefeld University); Mr SCHIRRWAGEN, Jochen (Bielefeld University); Prof. CIMIANO, Philipp (Bielefeld University); Mr PEIL, Vitali (Bielefeld University)

Presenter: Mrs AYER, Vidya (Bielefeld University)

Session Classification: Data Management Services

Track Classification: Area 1. Cross-Domain challenges / Data exchange across domains: researchers, technologist and policy makers perspectives