Enabling research data management: iRODS/Yoda
V0.6 22-08-2019

In a data driven age, research data management (RDM) is becoming an integral part of any research project. This can be from either a practical point of view - you have to deal with large amounts of data, ethical - you need to guarantee the integrity of the data you work with or simply good data management - you need to write a data management plan (DMP) and realize it in practice.

Increasingly researchers are faced with one or more of the following needs:
- Collaborate with other researchers in a shared project environment.
- Have full control over who accesses your project data and how they interact with it.
- Conditionally lock your data folders so that they cannot be accidentally altered.
- Add DataCite\(^1\) compatible metadata, making data findable and reusable by others (as well as future you).
- Easily freeze your annotated data in a vault for a specified number of years.
- Publish your data and make it available via a data catalogue directly from the vault.
- Develop procedures that enhance your data integrity.

It appears that few researchers have the knowledge, tools and protocols in place to address the above needs. Writing a DMP and implementing all requirements can be a significant hurdle and there is no easily accessible RDM technology available at the VU that facilitates this. To address this gap the VU is participating in a national program, initiated by SURF, which will pilot the iRODS/Yoda software as a potential RDM platform. This technology consists of two parts iRODS and Yoda. For more detailed information on any of the following sections please follow the links provided as footnotes.

iRODS\(^2\) is an open source, community driven, software infrastructure that has been developed to facilitate the management, sharing and annotation of very large amounts of data. It is being used in a number of Dutch universities, Utrecht, Groningen and Maastricht as well as EU projects such as EUDAT\(^3\). Internationally iRODS is supported by a diverse range of organisations including the Sanger Institute, UCL, NIEHS and SUSE.

While iRODS has a number of clear advantages for IT departments and “power users” it is the second component, Yoda, which will be of interest to most researchers wanting to engage in good RDM practices. Yoda\(^4\) stands for “Your Data” and is open source software developed and used at Utrecht University. It is designed to give researchers the capability to efficiently manage project data though the data management cycle\(^5\). Yoda is already being investigated as a potential RDM technology at SURF and the VU AIMMS institute.

\(^1\) https://schema.datacite.org/
\(^2\) https://irods.org/
\(^3\) https://eudat.eu/services/userdoc/irods-deployment
\(^4\) https://yoda.sites.uu.nl/home/introduction-to-yoda-2/
\(^5\) https://yoda.sites.uu.nl/home/introduction-to-yoda-2/data-lifecycle-in-yoda/
SURF is coordinating the national program for piloting the iRODS/Yoda software as a potential RDM platform. Several universities (i.e. UvA, TU Delft, WUR, TU Eindhoven, Erasmus University, Leiden University) have already defined the pilot projects. All participating pilot projects follow the suggested planning and evaluation criteria. Such coordination improves the transfer of lessons and experiences among the pilot project participants and provides insights for iRODS RDM technology implementation for other research projects starting from 2020.

Would you like to improve your RDM practices and do you need help addressing the common challenges with regard to research data? Please contact Alexey.Pristupa@vu.nl, Brett Olivier b.g.olivier@vu.nl or Peter Vos p.j.m.vos@vu.nl for more information about how to enrol for the VU iORDS/Yoda pilots.