



## RDM at GTB

The *Wageningen Common Data Solutions (WCDS)* programme is striving to connect use cases from Wageningen University & Research (WUR) research institutes to the research data management infrastructure for FAIR Data. Open source tools such as iRODS and Yoda are used for this purpose. WCDS is funded by the Ministry of LVNV for two years with a budget of 2 million euro.

In this fact sheet, we outline the progress with the project *RDM at GTB*. The aim of this project is to bring research data management at the WUR greenhouse horticulture business unit (GTB) one step closer to practice.

### Technical results

- We invested part of the development capacity in testing and improving iRODS. As feared, the developments are going slower than expected. We managed to run some trials and to start writing the data (currently 40TB) from the NPEC facility to tape. This is partly funded by NPEC. Much work remained to be done before iRODS can be scaled up.
- As a possible alternative, we examined SCALITY. This is faster than iRODS and we have already managed to set the rights and build some simple pilot applications.
- We also developed a new two-page RDM flowchart for our group and shared this with the coordinating data stewards. With this, we offer researchers a solution to the amount of paper and links they encounter in the RDM process.

- Lastly, we discussed possibilities to automate part of the work of data stewards and researchers. We discovered options for this in the new Myprojects. We recommend the WDCC to follow up on these possibilities and automate as much as possible.

### **iRODS not ready for roll-out yet**

With iRODS, we took big steps this year. However, iRODS is not ready to be rolled out yet. Currently, Azure and SCALITY are more mature and support our use cases better. Towards the future, a fully functional iRODS may be a good alternative.

### **Current limitations for further improvement**

In our quest for further improvement, we face two limitations:

- At the start of the project there was a 3 man team at FBIT, unfortunately one colleague left after a few months and another got sick for quite some time. This slowed the developments. Which blocked us from making the progress we hoped.
- Some building blocks must be properly tested before you can continue. It took some creative thinking to find out why the WUR iRODS servers threw errors on all uploads over 100GB and files got corrupted. Added to other issues to solve, this required a lot of time. Nevertheless, the original project goals are still within reach. We would like to keep improving iRODS until it can be rolled out within WUR and all researchers can start using it.

### **Additional developments and implementations planned**

- Although the iBridges shell implementation is stable,
-

---

the user interface still requires work. In addition, we want to integrate additional functionalities, such as the tape archive, so users can move files with a single click. The automated ingest script could use a similar polish, with a good amount of testing in different operating systems.

- Furthermore, there are a lot of small features to be added. For instance, SRAM was enabled for our iRODS server, but the current implementation is not user friendly. After that we could connect the metadata-platform through ISATAB jsons and DOIs and work out some small features like zipping instead of tarring when uploading to tape (data compression saves 50% of the space).
- Scalality and *Starfish* are interesting alternatives or additions to be explored. Next to this, we would like to help shape and test the new Myprojects to simplify RDM as much as possible and look at procedures.

## Key outcomes

- IRODS ingest
- Tested and debugged
- Most important outcome: experience. Thanks to the experience gained, we can now upload data to the tape archive: a major milestone

Furthermore, the project contributed to cooperation and interdisciplinarity within WUR. For example, our note to directors has triggered a discussion on how to move forward. We also organised discussions with the team for RDM protocols and had some good talks within the FAIR-iRODS-SEEK teams group.

We keep an eye on iRODS. Once the FAIR tooling becomes easier to use and, for instance, access management is implemented and tested, we will move forward with it. Until then, it is mainly the protocol that will be used.

## Reflections

In this project, keeping the focus has been a challenge. It is better to fix one thing than to find bugs in ten things. Furthermore, lack of technical support and guidance slowed us down.

After presenting our RDM note to the PSG directors, the discussion has moved to the Business Unit Managers' platform of Wageningen Research. As a result, RDM becomes a political process, and it may take years before progress is made.

## Embedding in Business Unit

A roll-out of iRODS in the current state would result in many questions and problems. Our advice is to let more pilot users move ahead on behalf of PSG and scale-up applications that work throughout PSG, while building up a support team.

As for the data management procedure, we have similar advice: follow the central discussions before moving forward.

## Project takeaway

RDM is broader than first imagined. The real issues are not technical. In the end, we gain more through the central discussions and talks at the director's level and the brainstorm sessions with FBIT. If this project continues, we prefer to shift our focus to helping other early adapters get to our skill level. This way, they can build on the work that is already done. Together, we can take the next steps.

## Recommendations

- Central coordination is key. Use the central position of the (coordinating) data steward to coordinate efforts and secure the knowledge after the project has finished.
- Automate and simplify the technical infrastructure: make things simple, so that most people can do it themselves. Also make sure good documentation, examples and training material are in place.
- Make sure to have at least one operational iRODS server at the start of a similar project, along with basic tooling, expertise and workshops how to use it.
- At the level of the WDCC programme: narrow the focus, so project teams work together instead of separate.

---

## Contact

WR Researcher | Wageningen Plant Research  
Ing. J.T. (Tim) van Daalen  
E tim.vandaalen@wur.nl

