



Advancing Fair Data Practices through integrated tools and Community engagement at WUR

The *Wageningen Common Data Solutions (WCDS)* programme is striving to connect use cases from 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.

Advancing FAIR data practices through integrated tools and community engagement is the title of the WCDS programme that aimed to create a FAIR infrastructure. This was done by combining FAIRDOME seek with iRODS and creating tools in Python and R for easy upload and access.

Technical results

- We successfully deployed an instance of FAIRDOME-seek on our Kubernetes platform. Additionally, an R package was developed to facilitate processing of the ISA-JSON file format, alongside the creation of multiple notebooks to test and validate the tools and infrastructure developed.
- We initially underestimated the demand within our institute for technological solutions beyond standards and ontologies. To address this, we established a community within WUR, primarily involving software engineers, data stewards, and data scientists. Together, we focused on FAIR data practices, ISA standards, and FAIRDOME-seek.

- Initially, we focused on using Yoda with iCommands due to the incomplete infrastructure for optimal iRODS implementation. At a later stage, we expanded our approach to include iRODS as well.

Current limitations for further improvement

The notebooks and ISA-JSON-based processing pipelines are in the early stages of development. We aim to expand these tools significantly to enhance usability and encourage broader adoption, especially given the high demand within the community. However, progress is currently limited by funding and time constraints.

Additional developments and implementations planned

Although the project has ended, we continue to pursue collaborations and funding to advance our initiatives. The concepts are currently being adopted by NPEC, PHENET, UNLOCK and PhenomUK. Additionally, we have received interest from industry partners and are establishing collaborations with Amsterdam University of Applied Sciences (Hogeschool van Amsterdam). Expanding the community will enable greater progress.

Key outcomes

The main outcomes of the project include:

- the installation of a FAIRDOM-seek instance
- Python notebooks
- an R package (to be released soon)
- an established workflow within Biointeractions
- integration with NPEC for managing Phenospex F500 data.

These achievements have largely realised the objectives outlined in the original proposal.

The outcomes of the project are being extended to benefit a broader community of WUR researchers and collaborators. The concepts are actively being adopted (see above). FAIRDOM-seek is intended for use within NPEC, where we are currently focusing on implementing iRODS with ISA-JSON integration. Additionally, PHENET will host a workshop on ISA-JSON to facilitate broader knowledge sharing and adoption.

The technical implementations from this project have enhanced the FAIRness of our data, particularly improving its reusability within the scientific community. By sharing our results and experiences with a broader audience, we are enabling more researchers to benefit from these advancements.

Fostering collaboration and interdisciplinarity at WUR around FAIR data sharing, was a key achievement of the project. By establishing a community focused on FAIR data principles, ISA standards, and FAIRDOM-seek, we brought together software engineers, data stewards, and data scientists across disciplines. Our partnerships with the aforementioned parties exemplify the project's success in building a cohesive network. Workshops, shared workflows and our technical solutions have further strengthened interdisciplinary ties and encouraged widespread adoption of FAIR data practices.

The adoption of iRODS within WUR ensures that this component of the infrastructure is secure for the future. However, like the ongoing role of data stewards, continuous development and maintenance of these tools and infrastructure require sustained time, support, and financial resources. We are actively seeking funding through other projects to support further advancements. WUR would greatly benefit from a fully funded Wageningen Data Competence Center, staffed with software engineers and data scientists, akin to the dedicated support available for IT infrastructure.

Currently, our business units lack dedicated funds to support these outcomes sustainably. We aim to incorporate these needs into the budgets of individual projects and proposals, though this approach results in inconsistent, wave-like support, which is neither efficient nor optimal for long-term data management strategy.

Funding was granted for the PHENET project, and additional proposals, including a TKI proposal, are pending. Given the critical need for structured metadata in AI developments, our approach has been incorporated into AI-focused projects.

Reflections

Challenges at the project level included limited availability of IT support staff due to understaffing, an initially incomplete iRODS setup, and the need to invest more time in community building than originally planned.

Project takeaway

The need for user-friendly tools to support FAIR data practices is a widespread issue across the organisation. While there is much discussion on FAIR principles, standardisation and ontologies, the actual availability of practical tools for researchers remains limited. If the project were to continue, we would focus more on developing general-purpose notebook and Python tools rather than domain-specific solutions, as there is strong demand for versatile tools of this kind.

Recommendations

- To achieve 'common data solutions' at WUR, we recommend significant investment in data stewards, data scientists, and software engineers, as a dedicated team is essential. With the rapid advancement of AI, metadata and tools for metadata management are increasingly critical. Expanding the WDCC with additional staff and funding would support these needs effectively.
- We recommend prioritising investment in personnel – data stewards, data scientists, and software engineers – rather than solely focusing on policy development. While technical infrastructure at WUR is strong and should be maintained, policies alone are insufficient without dedicated resources to support researchers effectively. Investing in skilled staff will prevent scientific progress from being hindered by a lack of practical support.
- Alongside institution-wide investments, domain-specific challenges – such as data volume and metadata standards – require targeted support. Programmes like this are invaluable in addressing these unique needs. We highly recommend their continuation to ensure tailored solutions across diverse research areas.

Contact

Wageningen Plant Research
Dr. S. Warris
E sven.warris@wur.nl

