WAGENINGEN
UNIVERSITY & RESEARCH

In-situ data

The *Wageningen Common Data Solutions* (WDCS) 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.

The project *In-situ data (management and opening of field agronomy data) persued three aims*. First: to identify current practices and datasets of WUR on in-situ crop agronomy. Second: to further extend our global crop repository AGROSTAC with relevant data. And third: to develop a roadmap forward for crop data at WUR.

Technical results

To ensure the repository meets the needs of its users, we began with a user-centric co-design approach: an online survey and in-depth interviews with six WUR scientists to understand how they use and manage in-situ agronomy data. Findings revealed that data are scattered over many different sources (within and outside WUR). Reuse is hampered by policy, incomplete metadata, and malfunctioning API interfaces. Overall, our community acknowledged the need for standardisation and sharing of data, for example by means of a global crop repository (AGROSTAC).

In this project we further improved AGROSTAC by:

- adding data sets (for example Crop_Data_Ludemann_et_al_2022 and C2D2_V200);
- adding a user-friendly import;
- extended the API interface (see this registration link)
- registered AGROSTAC in the GEO portal in support of global crop production monitoring.

Additional developments and implementations planned

In summary, most of these improvements were foreseen in the proposal. We explored the iRods framework and the WUR implementation of Yoda. Yoda is not actively used to manage in-situ agronomy data and it seems not useful to manage data sets during project execution. Nevertheless, we do see Yoda as a good solution to properly archive, annotate, and publish data sets at the end of project. For example, it could be used to publish data sets documented in the Open Data Journal for Agricultural Research (ODJAR).

We also recommend adding an optional metadata item which users can use to indicate if data is useful for further standardization by AGROSTAC. An important new feature would be a supervised semi-automated upload procedure to upload and publish new user data sets, such as data collected by mobile apps. For example, such procedure could harvest metadata from selected data sets in Yoda and guide the user in selection and preparation of data to be loaded into AGROSTAC. Next, the user could combine own data with all other available data within AGROSTAC to train phenology detection algorithms.

Key outcomes

Improved the AGROSTAC infrastructure

- We improved software environment – separate test and production instances;
- We extended the API interface with more requests, extra options, and extra response output (see register link);
- We improved the import procedure supporting different formats;
- We updated the technical documentation (API, import).

Inventoried, harmonised and added data sets

Three examples:

- Crop_Data_Ludemann_et_al_2022
- C2D2_V200
- Open Data Journal for Agricultural Research (ODJAR)

Use cases

- We improved the monitoring of the potato crop within EC-JRC MARS Crop Yield Forecasting System (MCYFS). For this, we used public-observed phenology data from AGROSTAC to train planting date algorithms.
- We explored the added value of AGROSTAC for the GYGA initiative to update yield gaps. We did so by updating the crop model parameterization using recent in-situ phenology data from AGROSTAC. See this link for Jupyter notebook.

Outreach

- We contributed to webinars. Examples are the Open Data and Open Knowledge (ODOK) workshop in Hangzhou, China and the crop map validation workshop in Washington, USA.
- We registered AGROSTAC in the GEO portal (see catalogs in beta release – link).
- We organised the online survey, in-depth interviews, and a final workshop to stimulate collaboration on FAIR data sharing.

Reflections

The WCDS call has been a very welcome opportunity to further strengthen our work on standardisation of in-situ agronomy data. The timing was good as there is more and more attention for this subject recently (see for instance Top et al. 2022, Boogaard et al, 2023). During project execution, we aligned with other in-situ data initiatives in which we are active. Examples are:

- the RVO AMS project (Area Monitoring System)
- WorldCereal (in-situ data for training of crop map algorithms)
- Geoportal (dashboard to explore in-situ data)
- GEOGLAM (in-situ data working group).

The improved and enriched AGROSTAC infrastructure is being used in several projects in our business unit WENR (such as MARSOP-MCYFS, EFSA and GYGA).

It was helpful to learn about the iRods framework and the WUR implementation of Yoda and how these technical solutions relate to our work. There is a large variety of dedicated data management solutions within our organisation with varying levels of FAIR data standards. It is important to support these solutions in improving their FAIR standards. So, in addition to offering and implementing new solutions, it will be helpful to support these existing initiatives to improve and evaluate if migration to new FAIR-supportive solutions is feasible and needed in the longer term.
The support, provided by other work packages of WCDS, was useful especially the monthly progress talks and workshops to explore the sustainability of our initiative.
The administration was challenging. The financial approval came very late. This complicated planning and created extensive delays.

Project takeaway

To sustain AGROSTAC as a reliable and valuable resource, dedicated funding is required. In short, AGROSTAC requires common, joint investment in harmonisation ‘hubs’ beyond projects. On the long term this pays off as it leads to higher efficiency and higher quality.

Recommendations

- Continue to further coordinate and enhance FAIR data management of in-situ crop agronomy data by engaging the WUR scientists and acknowledging the variety of solutions.
- Publish data sets, documented in open data journals such as ODJAR, in Yoda to both archive and publish data in one go.
- Focus future investments on technical infrastructure and strong policy and support systems to maximise data value in an AI-driven world. Effective policies can guide investments in data collection apps, management platforms, harmonisation, and valorisation tools. Appointing department or team data stewards as intermediaries can further enhance data curation and alignment with analysis needs, increasing its impact and relevance.
- Not all applications and initiatives around FAIR data need an iRods/Yoda solution but rather support on further improving their applications and initiatives. In addition, clearly explain the purpose of Yoda as a means to annotate, archive, and publish data.

Contact
Wageningen Environmental Research
Ir. Hendrik Boogaard
E hendrik.boogaard@wur.nl