FRIS:
Exploring and introducing the CERIF-object *Research Infrastructures*
and its golden record-view

Pascale Dengis and Ils De Bal, team FRIS, Dept. of Economy, Science and Innovation (EWI), Flemish government, Belgium

Introduction:

This contribution gives an overview of the newly added entity "Research Infrastructures" in FRIS, the regional CRIS of Flanders. With this contribution we want to share the process and model we used to capture the metadata. Furthermore we will focus on the visualization of the golden record for research infrastructures: agreements with stakeholders and final view.

FRIS, the regional CRIS of Flanders

FRIS, the Flanders Research Information Space, is a regional CRIS, capturing information on publicly funded research performed in Flanders. The FRIS-platform is connected with the CRIS-systems of all Flemish universities and many other research institutes. These connections enable research information to flow automatically from the research institutions to FRIS. As soon as a research institute changes something in its CRIS-system, an automated update is sent to FRIS. The exchange format of the data is CERIF, the Common European Standard for Research Information[1], version 1.5.

Until 2022, FRIS covered research information on these objects: researchers, research organizations, publications and projects. Recently FRIS includes also information on additional objects like patents, datasets and research infrastructure.

The (meta)data supplied are made accessible via the FRIS portal (www.researchportal.be/en) and through open API's.

Capturing Research Infrastructures

Since 2022 FRIS contains also information on the research infrastructures of the different research institutes in Flanders. A new legislation for funding instruments in 2019, and new agreements with the Research Foundation – Flanders (FWO) made that there was an obligation to gather information of research infrastructures and that this information needed to be sent to FRIS. The scope is well-defined: international infrastructures, large-scale infrastructures, medium-scale infrastructures, infrastructures ad hoc subsidized by the Flemish government and those infrastructures that are financed by the instruments BOF (Special Research fund) and IOF (Industrial Research fund).

The goal to capture this information is threefold: reporting of budget spent on (international) research infrastructures, making research infrastructures visible which stimulates cooperation and

---

[1] https://eurocris.org/eurocris_archive/cerifsupport.org/index.html

knowledge exchange, and in the future also having a track on which projects use which infrastructures and which publications are linked with which research infrastructures.

To capture the information in FRIS, a new datamodel needed to be developed and definitions to be made. Based on Cerif, research infrastructure can be an equipment or a facility, and in FRIS both – equipment and facility- can further be specified as an e-resource. Equipment is defined as an instrument for scientific research – typically something that is purchased, located on one site, often bought off-the-shelf). Facility is a (virtual and/or distributed) space for scientific research usually with one or more equipment's and/or resources, and functions as a service. An e-resource can then be ICT based resources (computers, storage devices, networks, platforms, software etc.), analysis tools and data(bases) to support scientific research.

Merging infrastructures into a golden record

International, large-scale and medium-scaled research infrastructures are in almost every case infrastructures in consortia[2]. This means that there will be a coordinator and partners in 1 project related to the infrastructure.

In FRIS we hold on tight to the silo-principle where each research institute owns their own data and guards the quality of its data. This means that, in case of infrastructures and consortia, different dataproviders send (different) data to FRIS about the same infrastructure. Together with the dataproviders, we agreed that each partner delivers info to FRIS, but the coordinator has to deliver more information than other partners. On some attributes of the entity "Infrastructure" only the name delivered by the coordinator is shown (e.g. the name of the infrastructure), on other attributes we will combine the information of the coordinator and partners, e.g. keywords.

Of course, the only way we can merge this information is when a persistent unique identifier (PID) is given. Without a pid we cannot link data and knowing that the information is about the same infrastructure object. In case of infrastructures financed by the FWO and the ad hoc infrastructures, the Department of Economy, Science and Innovation makes a list of infrastructures and attributes unique ID's. We need to do this as – to our knowledge - there is no worldwide PID existing for research infrastructures. Obviously, each partner has to use this same PID to succeed in merging these data. Data quality by deliverance is therefore essential.

In FRIS we have been developing this merging scenario and are ready to expose a golden record of research infrastructures. In our database we will keep the different records separated so we know which information has been delivered by each partner, but on the research portal[3] (new portal >2024) and the webservices we will expose this golden record.

We believe the quality of the golden record will be high as there is a consensus made by the stakeholders, the coordinator has an important role, and as information is merged, lacking information becomes apparent due to missing PID's.

---

[2] A consortium is a collaboration between organizations that have joined together to work on a particular project, in this case: on research infrastructures. A consortium consists of 1 coordinator and a various number of partners.
[3] www.researchportal.be

Golden record via aliasing via pids

In FRIS there are many entities which would benefit from having a golden record generated in the portal. This would avoid showing duplicates delivered by different providers and it would gain completeness by adding information from different perspectives. To build a golden record, we work in 2 phases: first we alias objects through PID's and then we decide which attribute has priority to another and on which base: role, last modified date, original date,…etc.

In the first phase we have identified for each entity some PID's which are used for the aliasing system, e.g. ROR for organizations, ORCID and ISNI for persons, DOI, Handle a.o. for publications,…etc.

In the second phase, the actual merging is based on scenarios agreed upon with the stakeholders. To do that we need to answer a questions like: Do we merge information of a reference provider with the information of regular data providers, and which information is dominant over other information. The logic needs to be hardcoded developed and might make it complex if we have multiple scenario's for one entity. To keep it as simple as possible, we aim to have one merging scenario per entity. At least for research infrastructures, we have succeeded to determine the PID and the scenario.

Future Steps

Research Infrastructures is a use case where we can see how to agree and develop the visualization of a golden record, but of course, we will go further and build golden records for almost all entities in FRIS.

These golden record visualization will be visible on the new portal, planned for 2024.

For research infrastructures it would be helpful if a worldwide PID would be developed and managed so that a true PID for infrastructure can be used by researchers worldwide.