

DataCite and FAIR Workflows

The world of research and academic publishing is complex and challenging. Researchers, institutions, funders, and publishers are all spending an enormous amount of time on administrative matters, managing compliance, navigating new business models, and seeking greater transparency. There is a need for collective efforts to enhance the discoverability and usage of research outputs.

DataCite is a leading non-profit organization that provides DOIs for research data and other research outputs – an active participant in the research community that promotes data sharing through different community building efforts and outreach activities. DataCite is the project organization that leads the “Implementing FAIR Workflows” project, funded by the Templeton World Charity Foundation. Prompted to address the complexities of scholarly information, DataCite, along with the project partners, aims to develop an exemplar FAIR and Open Research workflow based on the reality of the research lifecycle. The aim is to implement such a workflow through the integration of existing and well-accepted global open infrastructure, particularly the different identifier systems, thus enabling research evaluation and accelerated output reuse. It will also encourage researchers in the field of consciousness and beyond to adopt workflows that ensure the FAIRness of all research outputs by design, instead of as an afterthought.



ChronosHub is one of the project partners and acts as an intermediary to facilitate the information flow between the different stakeholders in the research ecosystem. By applying the FAIR principles through ChronosHub, all the different entities of research information get related to each other upon inception, i.e., upon grant approval, initial submission of articles or deposits of datasets.

More specifically, ChronosHub links the author with the manuscript when they submit it to a publisher, links that article submission to a grant, and the grant is related to a funder. In short, ChronosHub functions as a catalyst in creating a FAIR workflow that enables authors to get all their data for the whole research lifecycle with all the different entities connected. This also means that we become less dependent on peer-review systems, publishing platforms, and other systems downstream to also follow the FAIR principles. It is great if they do, but we already have the data FAIR which enables us to fully automate the processes and reuse the information in other systems. In more elaborative terms, this allows for a knowledge graph to automatically grow for each awarded granted, submitted article, and deposited dataset.

The knowledge graph in ChronosHub is fed into the PID Graph – an output from the EU funded project FREYA, preceded by project THOR and ODIN, that addresses scholarly communication infrastructure beyond traditional paper publishing. Persistent identifiers (PIDs) are not only important to uniquely identify research outputs, but the metadata stored with the identifiers can provide unambiguous linking between researchers, institutes, funding agencies, research activities, and outputs. The PID Graph is designed to connect related scholarly entities to each other by nesting their PIDs in the metadata in a standardized way. The implementation enables a wide range of use cases across stakeholders in the scholarly communities. The content of the PID Graph is publicly available over DataCite's GraphQL API. This offers, for instance, Institutional Repositories and CRIS-systems great opportunities to take advantages afforded by the PID infrastructure by contributing to and drawing from the wealth of publicly available metadata of scholarly records and their relationships, which not only increase the FAIRness of the research outputs, but also provide users with the means to tap into the valuable insights regarding the usage and impact of their work. Through the FAIR Workflows project we will explore ways for researchers to further actualize the potential of the PID Graph by actively engaging in the planning and management of research outputs in the beginning of a project to ensure FAIR practice throughout the research lifecycle, while implement necessary integrations among partnering research information systems and platforms to facilitate metadata exchange and enrichment, and finally develop an interface that demonstrate the connections established between research outputs and usage data captured by the PID Graph, thus allowing users to evaluate the research project as a whole.

Based on the above, we propose a session that will invite a discussion of how all of us in the research ecosystem can contribute with collaborative actions to build bridges and remove enormous amounts of administrative work with data collection, structuring, and quality assurance. As indicated by studies (MoreBrains on PIDs for Australia), this has the potential to save hundreds of millions of euros and make the data available for innovation in line with the EOSC ambitions worth upwards of 15 billion euros per year. We will present suggested best-practices and hands-on recommendations, applying a

“lessons learned” approach based on this initiative for establishing FAIR workflows throughout the research lifecycle and how this ties in with CRIS systems.