

DaMaLOS 2021 - Second Workshop on Data and Research Objects Management for Linked Open Science - Editorial Note

Co-located with the International Semantic Web Conference
ISWC 2021

Leyla Jael Castro ^{1*}[0000-0003-3986-0510], Markus Stocker ²[0000-0003-3986-0510], Dietrich Rebolz-Schuhmann ^{1,3} [0000-0002-1018-0370]

¹ ZB MED Information Centre for Life Sciences, Gleueler Str 60, 50931 Cologne, Germany

² Leibniz Information Center for Science and Technology, Welfengarten 1 B, 30167 Hannover, Germany

³ University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
* ljgarcia@zbmed.de

Abstract. Research data is the mirror of experimental work. Data, together with the software used to produce and analyze it, complements scientific publications and is core input to data- and knowledge-driven research. Most research activities follow the research data cycle, where data is continuously produced, transformed and (re)used, transitioning from one research to another. For this cycle to prosper, we require Research Data and Research Objects Management (RDM and ROM) plans supporting the findable, accessible, interoperable and reusable (FAIR) principles. Despite playing an important role, data on its own is not sufficient to establish Open Science nor Linked Open Science, i.e., Open Science plus Linked Open Data (LOD) principles. LOD principles, aka LOD 5 stars, follow objectives that overlap with FAIR principles and Open Science (e.g., LOD 5 stars include “openness” and the use of “non-proprietary open formats”). DaMaLOS 2021 workshop explores what is required for RDM to effectively instantiate Linked Open Science, including effective support for LOD, automation by, e.g., machine/deep learning approaches, FAIR and Data Spaces/Ecosystems. Furthermore, we are interested in innovations to also support other Research Objects such as software and workflows, in order to get an integrated layer supporting all the edges of Linked Open Science.

Keywords: DaMaLOS, Data Management Plan, Research Objects, Linked Open Science, Scholarly outcomes, FAIR

Editorial note

This second edition of the workshop on Data and Research Objects Management for Linked Open Science ([DaMaLOS 2021](#)) took place in the virtual space on the 24th of October as part of the [International Semantic Web Conference \(ISWC\) 2021](#). It was organized by by Leyla Jael Garcia Castro and Dietrich Rebholz-Schuhmann, both from [ZB MED - Information Centre for Life Sciences](#) (Cologne, Germany), and Markus Stocker from the [Leibniz Information Centre for Science and Technology, University Library](#) (Hanover, Germany).

In the past years we have seen a shift in the research cycle where results were commonly published as a scientific paper only. Nowadays, it is common to share and publish other research outcomes, aka research (digital) objects, including, for instance, software and workflows. Furthermore, the scientific community also recognizes the need of becoming open and thus, we see researchers sharing presentations, blogs, videos, tutorials. Although all these efforts contribute to realize the [FAIR principles](#) (findable, accessible, interoperable and reusable), improve reproducibility in science, and take us one step further towards Open Science; we still need more initiatives putting all the different elements in a coherent way so we move closer towards Linked Open Science. Our DaMaLOS 2021 workshop provided a space for researchers to present efforts in this regard and discuss further how to keep going as there are still many open questions on how to get there.

Keynote

The workshop delivered an excellent overview about the scope of existing solutions. Already the keynote talk (Bianca Kramer: “Open metadata - where do we stand?”, [DOI:10.4126/FRL01-006429411](#)) gave an excellent introduction over what types of data are nowadays typically available in their digital form, i.e. scientific publications and the related data and metadata, but also about the benefits from linking and sharing the data and its interoperability. The metadata has to be FAIR and Open, even if the data or the publications are not open. The metadata is public information (i.e., cannot be copyrighted), however collections of metadata are different. Also, abstracts have a creative leverage, i.e. they can be protected. The talk explored different metadata providers, i.e. Microsoft Academic vs. CrossRef, and their biases. In the long run, the research community has to provide solutions for sustainability.

First session: Applications and practical cases

The first session after the keynote was concerned with applications and practical cases. In the first talk (“The Survey Ontology: Packaging Survey Research as Research Objects”, [DOI:10.4126/FRL01-006429412](#) by Mario Scrocca) the requirements for the ontology have been presented: (1) capture the structure of the

survey, (2) semantically organise the response options, (3) provide a structure to collect the input, (4) make all collected answers available, (5) keep provenance, and (6) capture the survey methodology. The final outcomes are delivered as research objects, which capture the scientific workflow and the provenance traces.

In the second talk (“Environmental Observations in Knowledge Graphs”, [DOI:10.4126/FRL01-006429414](https://doi.org/10.4126/FRL01-006429414), by Rui Zhu) presented the KnowWhereGraph, which is a knowledge graph for observations about environmental problems. The covered events have complete space and time tags and cover events such as fires, biodiversity issues, and air quality. One key problem resides around the question: how to differentiate between individuals (e.g., single datum) and collections (e.g., entire datasets)? Time series can be collected as well, but this generates limitations (detail, storage). The talk gave a nice evaluations between different types of knowledge representations and the impact on the amount of data.

The third talk (“Analysis of Scientific Literature of LDOW Workshops: A Scientometric and NLP approach”, [DOI:10.4126/FRL01-006429416](https://doi.org/10.4126/FRL01-006429416), by Enayat Rajabi) described a analysis of about 160 articles from the LDOW workshops. The use of ontologies, such as BIBO for bibliographic data, SWRC for research entities, as well as FOAF, DC, DCTerms, and GeoName, enabled the structured representation of the data as well as the analysis of the content for topics that trend and for the scientometric analysis. The dataset is available over Zenodo (in RDF).

Second session: FAIR and Data Management Plans

The second session comprised two talks around machine-actionable Data Management Plans (ma-DMPs) and FAIR principles for Machine Learning. The first talk (“Automating Evaluation of Machine-Actionable Data Management Plans with Semantic Web Technologies”, [DOI:10.4126/FRL01-006429413](https://doi.org/10.4126/FRL01-006429413), by Tomasz Miksa) presented recent advancements regarding ma-DMPs, in particular an alternative to automatize the evaluation of ma-DMPs so the application profile defined by the [DMP Common Standards Working Group](#) at the [Research Data Alliance](#) (RDA) can be tailored to, for instance, different funders’ needs. The evaluation is done via SPARQL queries and applied to the use case of Science Europe using six categories and a satisfaction value ranging from zero to five. The approach facilitates the evaluation process so that a final assessment can be manually done by a person.

The second talk (“Working Towards Understanding the Role of FAIR for Machine Learning”, [DOI:10.4126/FRL01-006429415](https://doi.org/10.4126/FRL01-006429415) by Daniel S. Katz) of this session, and last one of the DaMaLOS workshop this year, discussed some aspects to be taken into account for the application of the FAIR principles to the Machine Learning (ML) scenario, a scenario that brings together data, software and workflows. There are already some platforms making it easier to publish, share, run and compare ML approaches, e.g., DLHub, Kipoi and OpenML, as well as some communities working on the subject, e.g., Pistoia Alliance, ELIXIR Europe and CLAIRE. The presentation posted some queries that are being tackled at different venues including webinars and RDA plenaries. In addition to aspects around FAIR, other topics such as reproducibility (and other *ilities) were discussed.

Final words

DaMaLOS 2021 workshop was attended by a total of 31 participants. The keynote, some of the talks and the final discussion raised some open questions regarding the inclusion and connection of research outcomes beyond the traditional scientific paper, the implementation of the FAIR principles beyond data, and how DMPs connect to it and bring together different stakeholders. We expect to see more developments in those regards, as, despite being topics with multiple initiatives around, more work is still needed to make them a common practice in research.