FAIRsFAIR *in Action*
Fostering Fair Data Practices in Europe

EOSC-NORDIC FAIRification study testing F-UJI

The aim of the FAIRification initiative for repositories run by the EOSC-Nordic project (http://eosc-nordic.eu) is to integrate datasets, software, publications, and other research outputs for researchers consistently and orderly. The way to do that is to adhere to the FAIR principles to make research data more reusable and science more transparent, efficient, and trustworthy . By doing so, the F-UJI tool developed by FAIRsFAIR was tested.

EOSC NORDIC

www.eosc-nordic.eu

The Challenge

The EOSC-Nordic FAIRification initiative team defined a sample size of about 100 data repositories, for which they evaluated the FAIRness of their (meta)data, with the aim to guide and coach the repositories, over time, to a higher level of FAIRness. The guiding and coaching were done via a number of webinars , whereby they selected specific topics that contributed to a lower FAIRness score for a substantial part of the sample and where they felt improvements could be achieved. Not all repositories host datasets that were easily findable via a unique, global identifier or even a persistent, resolvable identifier: 24 % of the sample could simply not be evaluated due to lack of identifiers to the datasets.

Software to perform automated evaluations proved to be a good tool to evaluate the FAIRness of the findable metadata and data of repositories. For this purpose the team decided to embrace the F-UJI evaluator designed by the University of Bremen and Pangaea as part of one of the work-packages of FAIRsFAIR as the standard assessment tool for the EOSC-Nordic FAIRification initiative.

Machine actionable metadata is the key to interoperability and reusability. It is in fact crucial that the metadata is **machine actionable**, so that a "machine-a-gent" or algorithm can find, interpret and process the data, based upon the metadata found and harvested on the landing page of the repository. FAIR is meant for both human and machine interactions and is therefore all about automated findability and machine-actionability of data and metadata.

The implementation

Thanks to this study the EOSC-Nordic team found that many Nordic and Baltic repositories were struggling to define the right level and content of "generic metadata" (metadata related to the overall aspects of the research, independent of the domain). Defining the right level and content of the domain specific metadata was a major challenge. It is essential that domain communities work towards a clearly defined meta-data-schema that repositories should adopt.

@FAIRsFAIR_eu
/company/fairsfair
www.fairsfair.eu

EOSC-NORDIC FAIRification study testing F-UJI

Main findings:

- 30 % of the repositories had no support for machine-actionable metadata what so ever, a few repositories supported some degree of machine-actionable metadata or had some metadata standards in place. Only one handful of repositories scored more than 50 % on machine actionable metadata. There is still a lot of work to do!

- Data and metadata should be separated, metadata should point to the data and data should point to the metadata that holds its description. The concept of "FAIR DIGITAL OBJECT" plays an important role here in terms of permanently and intelligently linking the metadata to the related data sets and vice versa.

- A noticeably higher FAIR scoring is verified among the repositories being run on established platforms (Dataverse, Figshare, and others), and somewhat higher scoring among CoreTrustSeal certified repositories.

- The importance of "controlled vocabularies" is crucial in reducing the amount of free text and strongly increases the interoperability and reusability of data. Combining domain expertise and FAIR metadata-expertise in a Metadata-4-Machines (M4M) Workshop will result in a (domain specific) metadata schema that can be used, reused and shared by a particular community.

- A large percentage of the data repositories struggled to provide the right object types to datasets (is the file a graph, a report, a picture, or an excel file). Controlled vocabularies with mandatory input could help to improve the situation.

- Existing standard structures like DCAT and Dublin Core may assist in defining metadata-schemas for a community.The 4 known domain agnostic data schema services (Dublin Core, Schema.org, DCAT and Datacite) can all be used for generic meta data creation, whereby the user needs to consider that they can all be used for embedding in the landing page.

Thanks to this study, the EOSC-Nordic project was able to make repositories in the Nordic and Baltic countries more aware of the necessity of sharing data and they have been able to guide the repositories through the technical challenges of making their data more FAIR, through a step-by-step process.

Adopting Organisation/Body:

EOSC-Nordic project

Which stakeholder category do you best represent?

- **Repositories**
- **Research funding organisations & national agencies**
- **H2020 or other funded Projects**

Contact person

Bert Meerman - Member of WP 4
b.meerman@gofairfoundation.org

@FAIRsFAIR_eu
/company/fairsfair

www.fairsfair.eu