FAIRness assessment of datasets and ontologies at Eurac Research

IMPLEMENTATION STORY
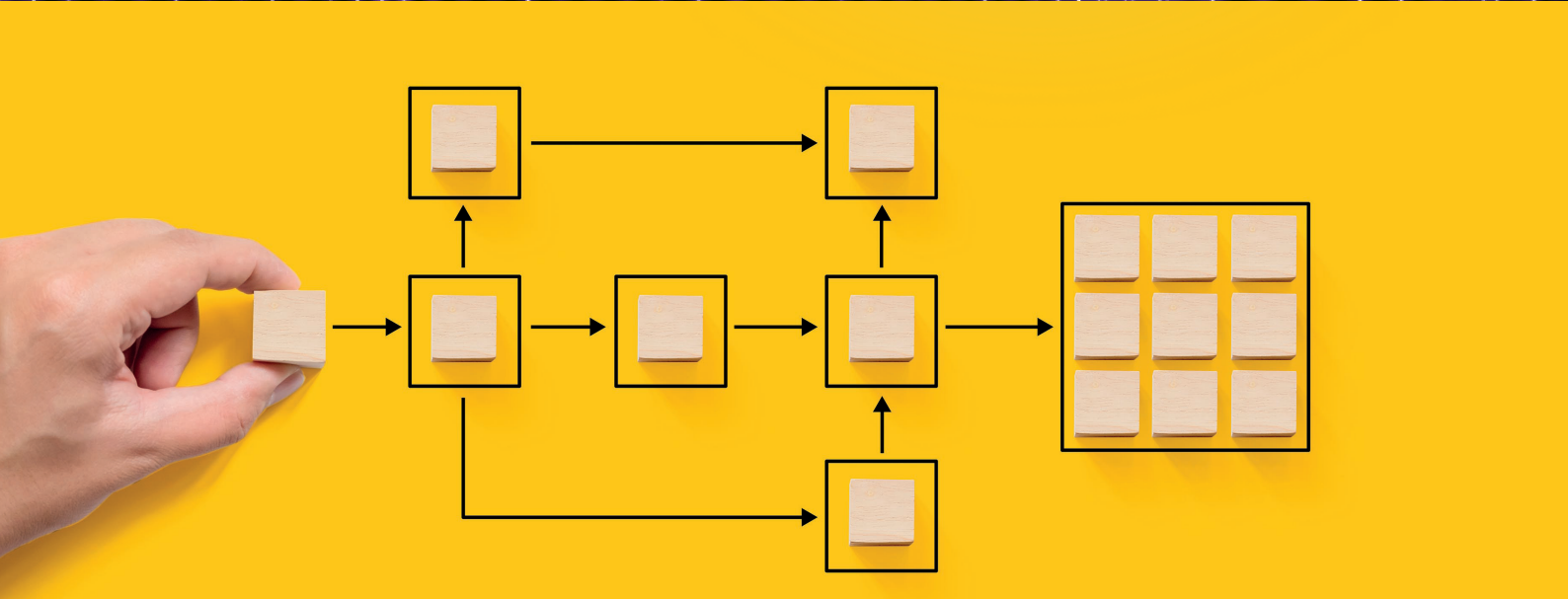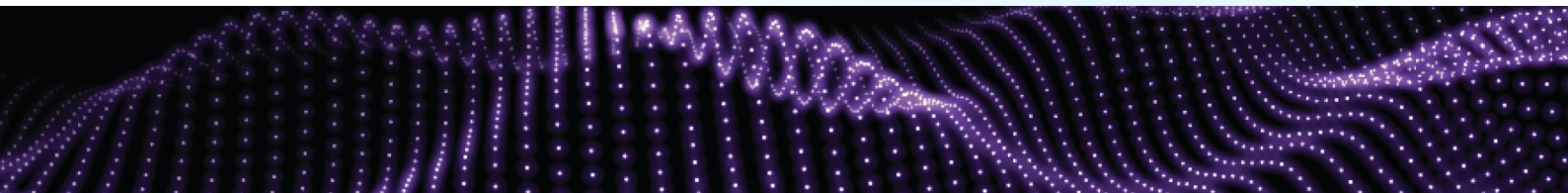
Authors names, affiliations and ORCIDs:

☐ Andrea Vianello, Eurac Research,
https://orcid.org/0000-0002-3017-6045

☐ Ekaterina Chuprikova, Eurac Research,
https://orcid.org/0000-0002-8981-0442

☐ Clara Linés, Digital Curation Centre (DCC),
https://orcid.org/0000-0003-3437-5145

Summary:

A team from Eurac Research assessed the FAIRness of a dataset hosted in their Environmental Data Platform using F-UJI and an ontology using O'FAIRe. After an initial low score in F-UJI, they implemented signposting to improve the machine interoperability of their metadata catalogue.

Introduction

At the Centre for Sensing Solutions (CSS) of Eurac Research, we aim to help our researchers share their data in a FAIR way through our Environmental Data Platform (EDP). We share about 450 datasets of different types in our web platform, all of them with metadata that we harvest from different data repositories into a central catalogue for metadata. We joined the FAIRness assessment challenge support action to understand better how to assess and improve the FAIRness of our repository.

Our team recently started exploring semantic artefacts, a new area for us. We saw this as a chance to learn about managing semantic metadata and tools for doing so and connecting with experts in ontology engineering and FAIR principles (making data Findable, Accessible, Interoperable, and Reusable). The Apple Ontology, our first published ontology, came out of a workshop. We aimed to understand these tools and principles early on to make our work better and more aligned with FAIR standards right from the start. This approach helps us integrate semantic artefacts into our research effectively.

Approach taken:

We selected a dataset formed by a collection of processed remote sensing images to assess with the F-UJI tool and our Apple Trait ontology, which was already published in the AgroPortal, to be tested with O'FAIRe.

During the first workshop, we understood that we could easily improve the FAIRness of the dataset because we manage our web interface for the GUI of the metadata catalogue. To enhance the FAIRness of the Apple Trait Ontology hosted on AgroPortal, our strategy involved a focused engagement with the ontology's metadata, leveraging AgroPortal's integrated O'FAIR tool. This tool facilitates an assessment of the ontology's adherence to FAIR principles—Findability, Accessibility, Interoperability, and Reusability—providing a foundational understanding of our ontology's current FAIRness status.

We also understood that we had much to do to improve the FAIRness of our service. We have previously done a lot of work to provide good information for human readability but not so much for the machine usability of the metadata, which is the focus of the assessment performed by F-UJI. Because of that, our initial score was relatively low. We realised there is so much more we could do to improve machine interoperability regarding the metadata.

Knowing this, we started working to improve the metadata catalogue interoperability using signposting to evolve our metadata catalogue. We were lucky to have another Eurac team following the support action for signposting and RO-crate, who could support us with this implementation. Without this, it would have been challenging to reach the good score that we finally achieved.

Although all our datasets have associated metadata, not all the metadata records are complete enough. We try to automate the process of completing the metadata records as much as possible, as it would be challenging for us as a data provider to do this manually for all the datasets. To implement the signposting, we need to provide different information fields as links (e.g., contact ORCID, link to the resources, link to ID). If we manage to get this information automatically, we will be able to reach a very high score for all our datasets.

Challenges encountered and addressed:

The main challenge was understanding how to use signposting in our metadata catalogue in a way that can be read by the F-UJI tool and therefore facilitates machine interoperability.

For us, a limitation of the assessment done by F-UJI is that it does not consider metadata quality. It can give you a high FAIRness score if you provide all the required links to metadata, even if this information is unclear. If the metadata is wrong, that high score would be meaningless. Maintaining the information up to date over time is a big effort with an associated cost, which can make the cost of keeping the FAIRness of the service relatively high.

We learned that it is not enough to have good metadata if they are not shared in a way that facilitates machine interoperability and we understood that the CSW standard is old-fashioned but still the official standard for metadata catalogues.

We identified some limitations of the F-UJI tool; it doesn't consider CSW for machine interoperability, that is a recognised OGC standard to share metadata. Another limitation of the F-UJI tool is that it cannot evaluate new metadata specification in json format such as the STAC specification. Web applications are adopting json format to exchange information and, from our perspective, this should be integrated in the F-UJI tool, otherwise the score will always be very low even if the information is linked to the data in an interoperable format. Finally it is not clear which applications are actually using/reading signposting, this aspect was not clarified in the workshop and it would be really important to motivate its adoption.

For this reason, we had an internal discussion to understand if signposting is the only way to get machine interoperability and to be FAIR. We think it may be the future technology, but not the only way to achieve data FAIRness.

Advancing the FAIRness of the ontology posed significant challenges, primarily due to our limited efforts to enhance the ontology's metadata and core structure—tasks that require considerable time. The ontology's prior publication on AgroPortal and an additional platform represented a significant aspect of our initiative. Fortunately, our engagement with the O'FAIRe tool, seamlessly integrated into AgroPortal, proceeded without complications. The straightforward design of the tool and the simplicity of updating metadata entries aided our endeavours in improving the ontology's FAIRness, enabling us to proceed without encountering any procedural hurdles. A two-pronged approach is deemed necessary for further advancement in the ontology's FAIRness. First, modifications within the ontology itself are essential. This involves thoroughly reviewing and enhancing the ontology's structure and content to ensure greater alignment with FAIR principles. Second, to elevate the ontology's interoperability and reusability, a strategic incorporation of external ontologies and classes is crucial. By aligning with and reusing established external ontologies, we can foster a more interconnected and seamless ecosystem of semantic resources.

Impact:

We have several people involved in helping the researchers to increase FAIRness inside our organisation, so we were already looking into these topics as an organisation. However, participating in the support action has improved our understanding and competencies to assess FAIRness, and it has triggered an internal discussion on the publishing practices of ontologies within our organisation. We expect to discuss how we can manage metadata information for ontologies in our platform. Currently we publish the ontology metadata in external portals, while the datasets instead can all be managed in our web portal. We learned about possible solutions for ontologies from seeing what other teams are doing, and we will try to implement this in our services. This experience will also help us have a better involvement in the FAIR community for further improvement of our data and semantic artefacts practices.

The support action was a very good experience, we tried to do as much as possible during its duration, but we will

continue afterwards, and we will try to have a look at all the other tools that were shared with us during the support action too.

We hope to increase visibility of our datasets and to obtain, in the near future, a certification of quality for our data portal. We currently try to follow the guidelines regarding data sharing and FAIRness in general, but we don't have any certification. We first need to find the right certification for our portal, which is something we are currently discussing.

Key message:

FAIRness should be assessed as automatically as possible to reduce maintenance time.

Metadata quality has a high relevance to describe your datasets following FAIR data principles. If your metadata information is wrong, FAIRness score could be incorrectly high.

Working on ontology engineering should initially include the thought of FAIR data principles, as it eases the process of making the ontology interoperable and thus reusable.