MASTER di I LIVELLO IN

ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE

ANNO ACCADEMICO 2023/2024

Studente

Filippo Vasone

Titolo

Towards FAIR Research Data Management in Experimental Materials Science: A Metadata Pipeline

Abstract

In contemporary academic culture, the Open Science movement promotes a series of practices to increase the transparency and, ultimately, the quality and reproducibility of scientific research. One prominent practice is research data management according to the FAIR principles (Findable, Accessible, Interoperable, Reusable). Despite their widespread institutional uptake, applying the FAIR principles, i.e., FAIRification, is challenging. The thesis describes an exploratory FAIRification project undertaken during my internship at the National Metrology Institute of Italy (INRiM). The project focused mostly on metadata management, interoperability, and the use of electronic laboratory notebooks (ELNs). The main result is a metadata pipeline for a specific experiment in materials science, namely the Spin-Orbit Torque (SOT) measurement in the field of spintronics. The metadata pipeline transfers metadata from the experimental context of the data production (controlled by a Python program) to their publication in a domain-specific repository in a FAIR way. We adopted a contextual methodology for FAIRification, adapting the FAIRification process to the scientific context at hand. To realize the metadata pipeline, we used three tools. The first tool is CHADA, a conceptual metadata standard for material characterization. The second tool is an ELN software, eLabFTW. The third tool is NOMAD, an open-source repository and archive for materials science data. The pipeline starts from the definition of a conceptual metadata schema, written according to CHADA. This metadata schema is made machine-actionable by building an experiment template in eLabFTW. Successively, we gather metadata thanks to an API integration between the computational environment of the laboratory and eLabFTW. The user's entry also contributes to the metadata. Finally, the metadata is prepared for publication in a FAIR and interoperable way using NOMAD, which allows dataset publication with DOIs and ensures interoperability through its schema language, NOMAD Metainfo. To create a suitable schema and realize (partial) interoperability, the keys from the conceptual metadata schema are matched with the NOMAD Metainfo. Then, API integration enables automatic data transfer from eLabFTW to NOMAD. In sum, the metadata are transformed from an unstructured, non-FAIR, format to an organized and FAIR format. A FAIR infrastructure, like this metadata pipeline, enables future data analysis and scientific discoveries with computational means, e.g., machine learning.