H2020 NanoCommons Transnational Access Services: Data management in nanosafety research. From bench to database thus streamlining analysis and publication

A. G. Papadiamantis* and I. Lynch

School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK
Contact: A.Papadiamantis@bham.ac.uk

Introduction

The transformation of nanosafety research into a data intensive field, and the advancement of machine learning and modelling and analysis techniques, has raised the demand for large volumes of high quality findable, accessible, interoperable and reusable (FAIR) data (Figure 1). The two key aspects of data interoperability are data annotation and metadata implementation. Data annotation is the process of labelling the produced data using ontologies, making sure that all terms are defined and can be harmonised with other datasets. Metadata is the process of implementing the necessary explanatory data to make the dataset findable, easier to analyse and replicate. Despite the significance of both processes, during the entire data lifecycle, they are not widely implemented in everyday scientific practice, due to lack of training in data management. As a result, a huge amount of data remains unexploited and data curators working with already published data are not usually able to identify all the information needed to fully implement the necessary annotations and metadata.

Aims and Objectives of UoB Transnational Access

- To design, build and implement data management processes, tailored to the needs of the nanosafety community and covering the whole data lifecycle (Figure 2)
- To make metadata capturing and recording and data annotation part of everyday scientific practice (Figure 3)
- To facilitate the process through automated processes and online lab notebooks (e-notebooks)
- To promote data harmonisation, enhance data quality and increase interoperability
- To promote Open and FAIR data through provision of data management tools
- To offer a range of analytical and modelling tools to the scientific community.

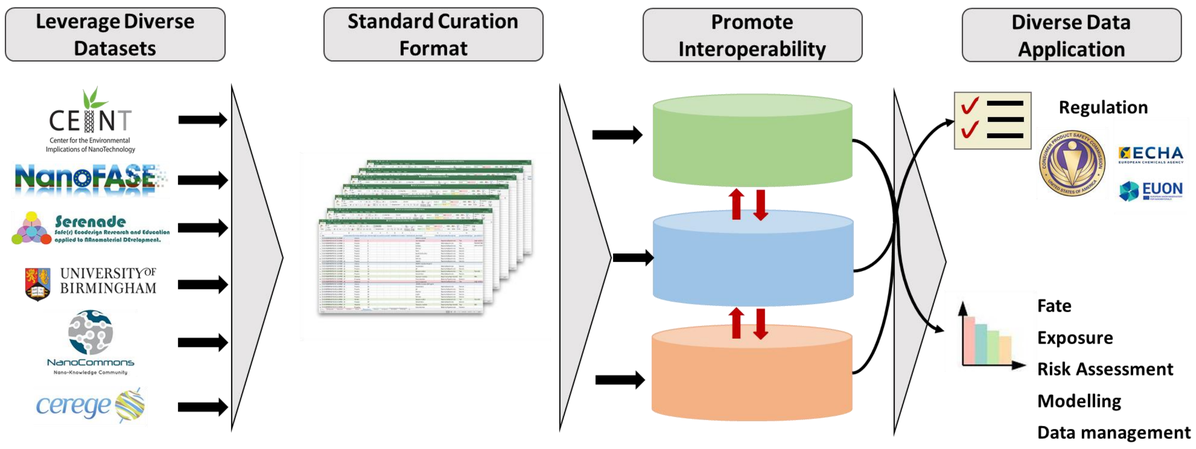

Figure 1: Data interoperability workflow

Approach and Applications

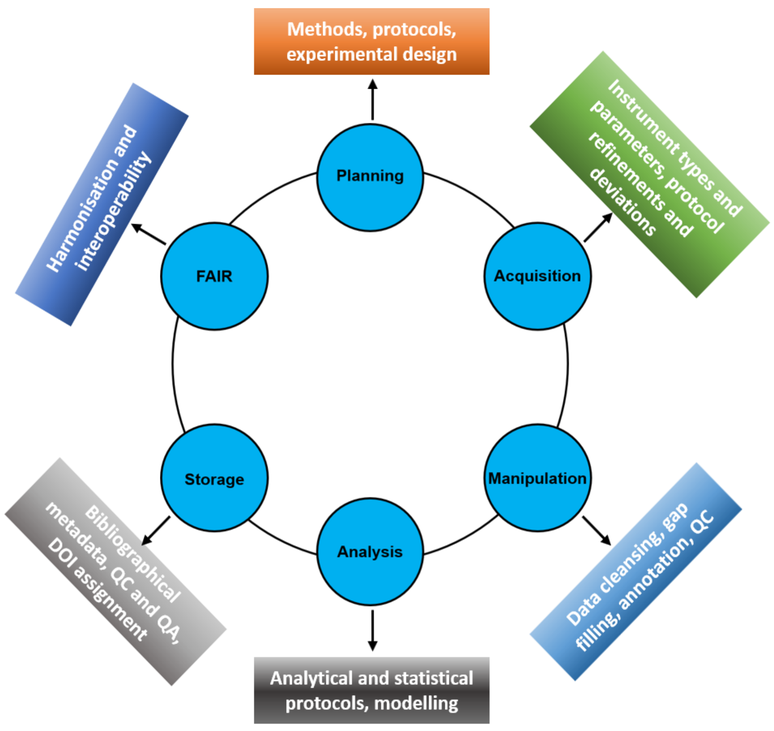

Figure 2: Data lifecycle, annotation and metadata implementation

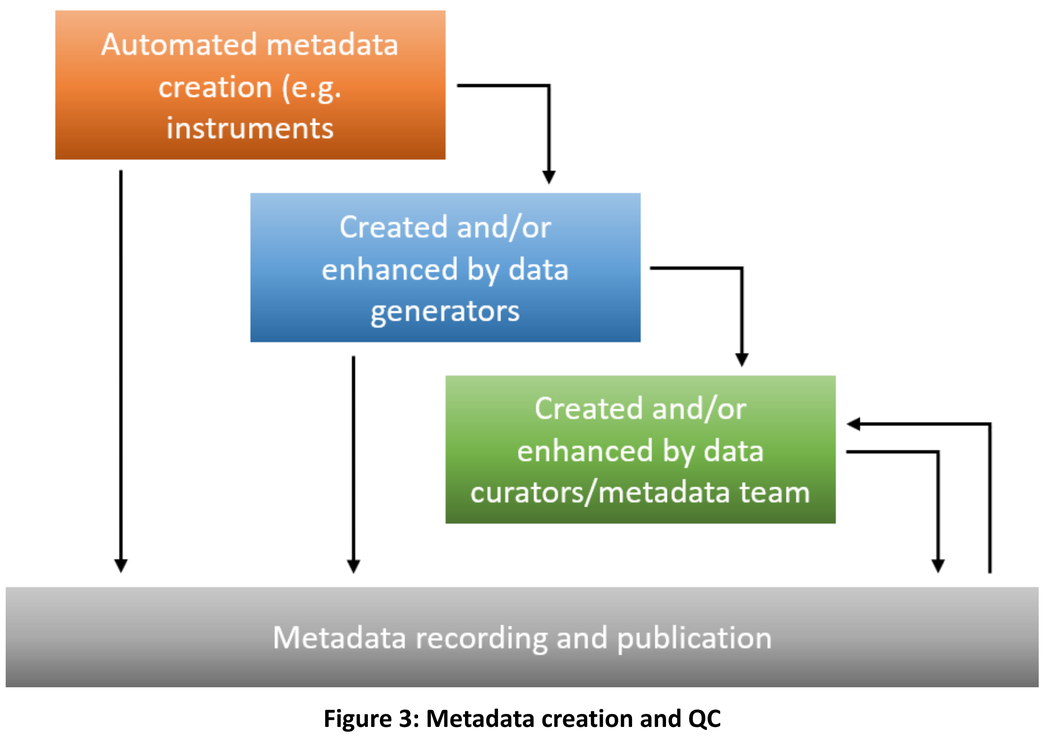

Figure 3: Metadata creation and QC

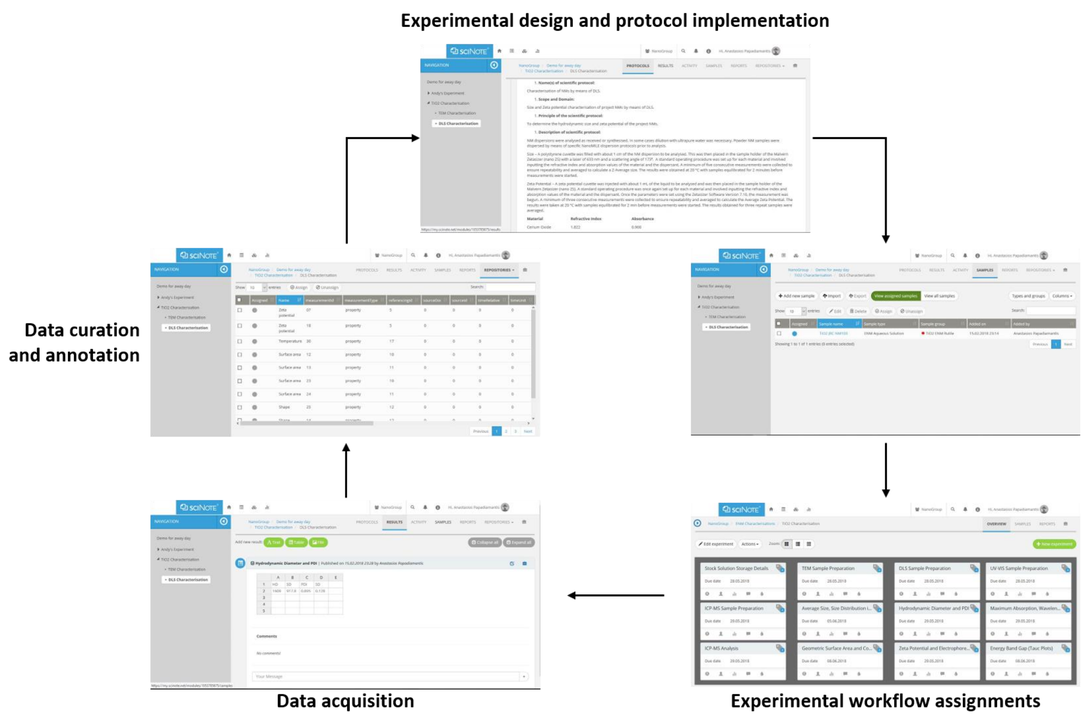

Figure 4: Online lab-books experimental workflow, data curation and annotation

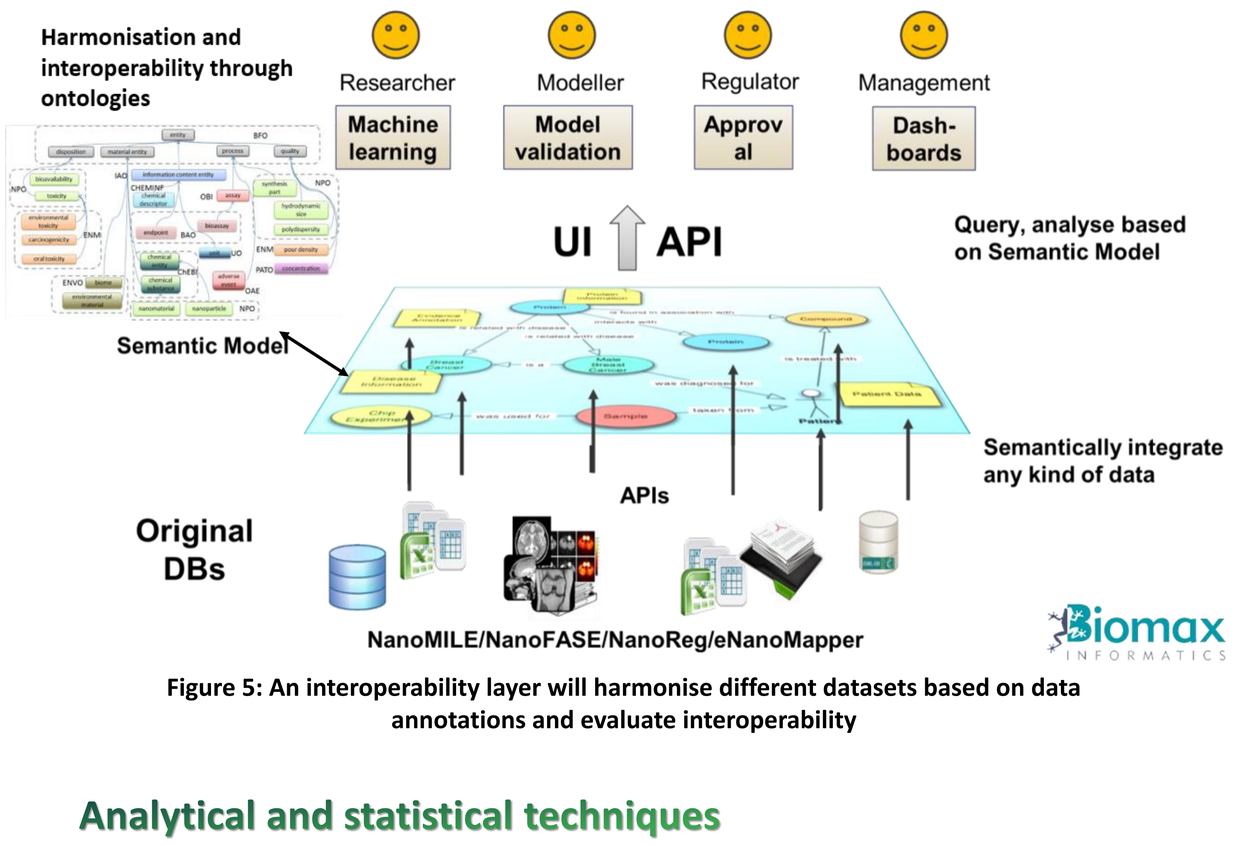

Figure 5: An interoperability layer will harmonise different datasets based on data annotations and evaluate interoperability

Types of Metadata/Data Annotation

- Metadata provides critical information and allows experimental repeatability and QC of data (Figure 4)
- Metadata cover the whole data lifecycle and can be divided in: bibliographical, experimental (e.g. instruments, characterisation methods) and analytical (e.g. statistical protocol, models)
- Ontologies provide a hierarchical structure overview along with the relationships between experimental terms and endpoints
- Ontologies act as a vocabularies for term harmonisation, making sure that different datasets "speak the same language"
- Datasets are annotated using ontological definitions, allowing findability and cross-dataset interoperability (Figure 5).

Analytical and statistical techniques

- UoB offers guidance and support in the linear and non-linear statistical analysis of complex data
- UoB provides training on the types (linear, non-linear, normality) and dataset size requirements for the identification of the most appropriate analytical techniques
- UoB expertise covers a wide range of statistical approaches and training using software (e.g. SPSS) and coding (e.g. R, MATLAB) practices
- UoB has experience in the analysis and Rietveld refinement of XRD data and EXAFS analysis of Synchrotron data.

Contact:
A.Papadiamantis@bham.ac.uk
info@nanocommons.eu

This project has received funding from the EU H2020 Programme:
Grant Agreement nº 731032

Complete the NanoCommons survey and help us expand the tools and services we offer