

Putting metadata to work in research with Marble and Beaverdam

Researchers in many fields rely on complex data from specialized instruments and large numbers of experiments. Metadata is key to efficiently document and describe data's essential attributes, and help to generate overviews of large datasets. Manually collecting and curating the extensive amounts of metadata required –some of which might be even inaccessible –is a major challenge. To support scientists in this endeavour, we develop software tools that automatically extract and consolidate metadata, which can then be put to use during data selection and further processing.

Marble (Meta**d**ata in p**r**oprietary B**i**nary f**i**les) allows researchers to identify and extract metadata from proprietary binary files. Software that operates scientific instruments often stores data in proprietary formats which are inaccessible outside of the manufacturer's ecosystem. Even if the software supports data exports, exported files may contain only a subset of the metadata originally captured, thus valuable information might be lost. Marble supports researchers in identifying sequences of data and deciphering metadata from proprietary data formats. A user interface presents results to researchers, who can adjust the deciphering method and annotate the identified information. The method and file structure is stored in converter files, which can be reused and exchanged between users. Our long-term goal is to create a software framework which creates fully automatic converters to translate proprietary data into accessible and reusable formats.

Beaverdam (Build, Explore, And Visualize ExpeRimental DAtabases of Metadata) allows researchers to interactively explore large amounts of metadata. Users can combine similar metadata from multiple experiments into a database, then access a graphical user interface in a web browser to interactively identify subsets of experiments meeting specific criteria. A data table and interactive plots describe characteristics of the selected experiments on the fly. Beaverdam is suitable for use at multiple scales –individual researchers can build and access a database locally, or a research group can maintain a joint database which members access remotely. While we are testing Beaverdam on electrophysiological brain data, it is domain agnostic and will be useful for all research disciplines.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC staff

In addition please add keywords.

metadata, software, proprietary, database, dashboard

Primary authors: D'MELLO, Fiona (Forschungszentrum Jülich GmbH); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany); BRINCKMANN, Steffen (Institute for Energy and Climate Research –Structure and Function of Materials (IEK-2), Jülich Research Centre, Jülich, Germany); GRÜN, Sonja (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); DENKER, Michael (INM-10,

Forschungszentrum Jülich); HOFMANN, Volker; SANDFELD, Stefan (Institute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Presenters: D'MELLO, Fiona (Forschungszentrum Jülich GmbH); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany)

Session Classification: Postersession II

Track Classification: Postersession