



Contribution ID: 121

Type: POSTER&PITCH

Interactively exploring metadata with Beaverdam

Monday 4 November 2024 15:00 (1 hour)

Scientists frequently need to get an overview of their experiments by summarizing information spread over multiple files and storage locations. This metadata may include items such as experimental conditions, subject details, and characteristics of the experimental data. It is common for researchers to spend time developing their own solutions tailored to their specific use case. However, overviews of metadata have similar requirements across research fields. We leveraged these similarities to develop generic software for efficiently exploring collections of metadata, which scientists can quickly customize for their own work.

Our software Beaverdam (Build, Explore, And Visualize ExPeRimental DAtabases of Metadata) combines metadata from multiple experiments into a central database, then builds an interactive dashboard to explore the contents of the database. Graphs show a high-level overview of multiple experiments, a table shows details of each experiment, and interactive filters help researchers identify experiments meeting specific criteria. Users customize the dashboard using a single configuration file. We developed Beaverdam in Python and have released it as a Python package which users can run from the command line or incorporate into their own code.

Although we designed Beaverdam for all sizes of datasets, its automated approach is particularly useful for datasets with many experiments and/or extensive metadata. We tested Beaverdam with metadata from a neuroscience dataset in which each of the hundreds of experimental sessions contains thousands of items of metadata. Using Beaverdam, researchers were able to efficiently identify experimental sessions meeting their criteria for further analysis – a task that would have been impossible by hand.

We expect that Beaverdam will help scientists efficiently explore their metadata, identify gaps, and inform further analyses.

Beaverdam on GitHub (open source): <https://github.com/INM-6/beaverdam>

Funding: This work is supported by the Helmholtz Metadata Collaboration (HMC), EU Grant 945539 (HBP SGA3), and the NRW-network iBehave (NW21-049).

Please specify "other"

In addition, please add 3 to 5 keywords.

database, metadata, software, visualization, Python

Please specify "other"

For whom will your contribution be of most interest?

Researchers

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

Data professionals and stewards

Primary authors: MORE, Heather (Institute for Advanced Simulation (IAS-6 and IAS-9), Forschungszentrum Jülich); DENKER, Michael (INM-10, Forschungszentrum Jülich); Prof. GRUEN, Sonja (FZJ); SANDFELD, Stefan; HOFMANN, Volker

Presenter: MORE, Heather (Institute for Advanced Simulation (IAS-6 and IAS-9), Forschungszentrum Jülich)

Session Classification: Poster Session B

Track Classification: Connecting research data: 7. Infrastructure and common practices for consolidation of (meta)data