

Putting metadata to work in research with Marble and Beaverdam

D'Mello, Fiona^{1*}; More, Heather^{1,2*}; Steffen Brinckmann³, Sonja Grün², Michael Denker², Volker Hofmann¹, Stefan Sandfeld¹

¹Institute for Advanced Simulation – Materials Data Science and Informatics (IAS-9), Jülich Research Centre, Jülich, Germany.

²Institute 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.

³Institute for Energy and Climate Research – Structure and Function of Materials (IEK-2), Jülich Research Centre, Jülich, Germany.

*authors contributed equally to this work

Marble

MetadAta in pRoprietary BiNaRY files

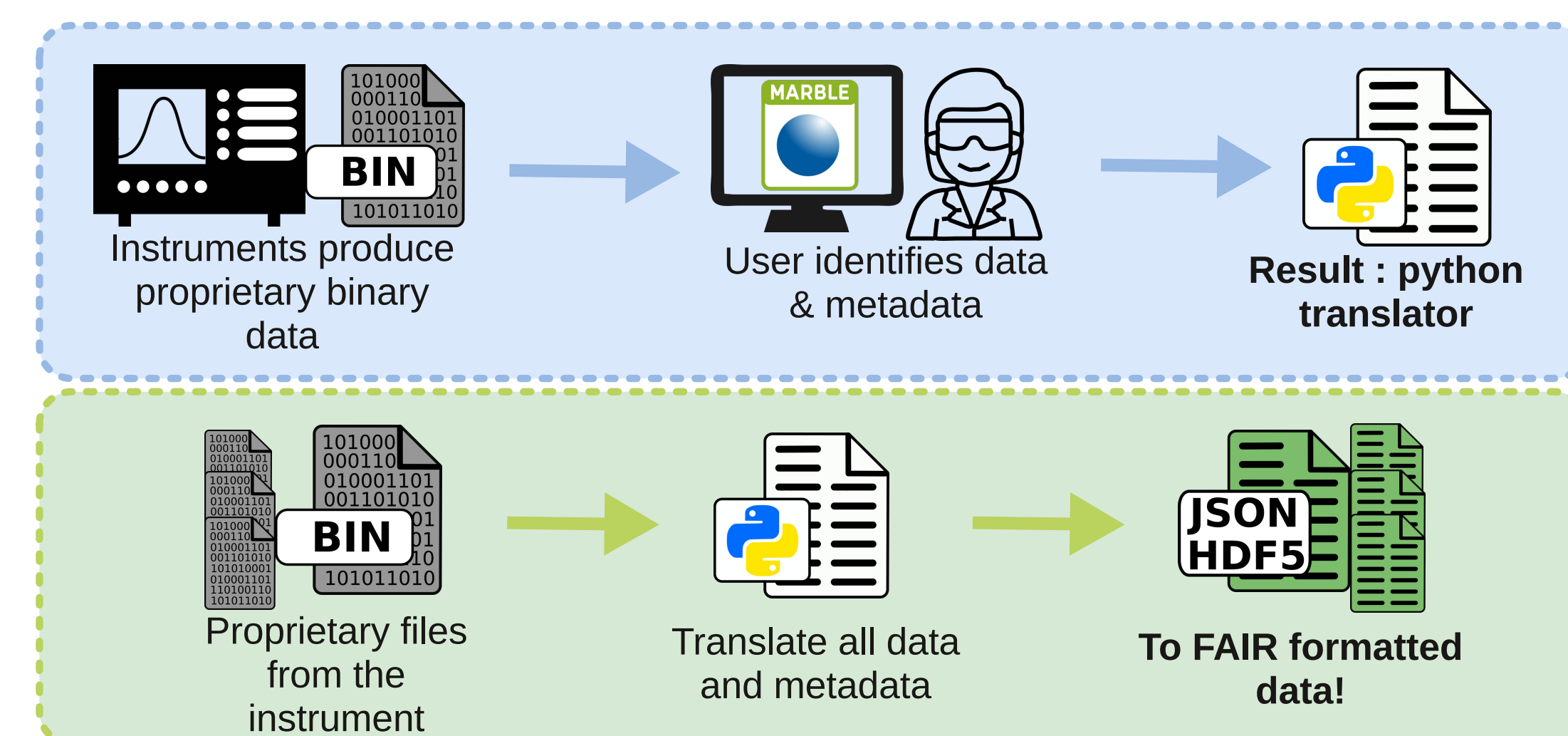
The problem

Files generated by scientific instruments contain primary data and metadata. To varying degrees **valuable metadata** remain **inaccessibly hidden**.

Hidden metadata **negatively impacts** interoperability, reusability, data provenance and reproducibility of experimental data.



Software deciphers binary data



Marble helps!

- Access all information from proprietary formats automatically
- Create and/or reuse instrument-specific python-based translator
- Translate, standardize your experiments data

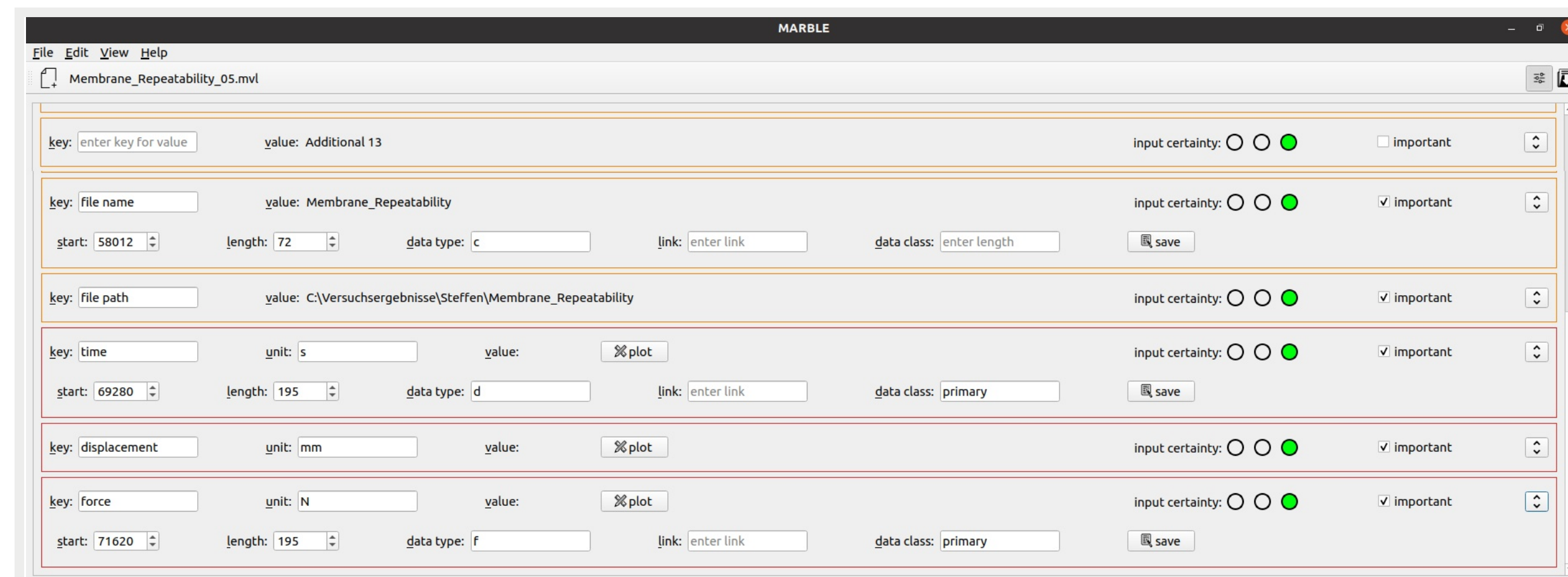
My instruments all produce different metadata files -- how do I decipher their information?

How do I process files from an instrument?

Is my data sensible to others who might use it?

Features

- Structured, annotated presentation of all decoded information
- Standardized output as HDF5 file enriched with JSON metadata
- Multiple versions Available both in GUI and CLI



Acknowledgements

This work was carried out at the Hub Information of the Helmholtz Metadata Collaboration (HMC) Platform and the Institute for Energy and Climate Research (IEK-2) of Forschungszentrum Jülich (FZJ)

Beaverdam

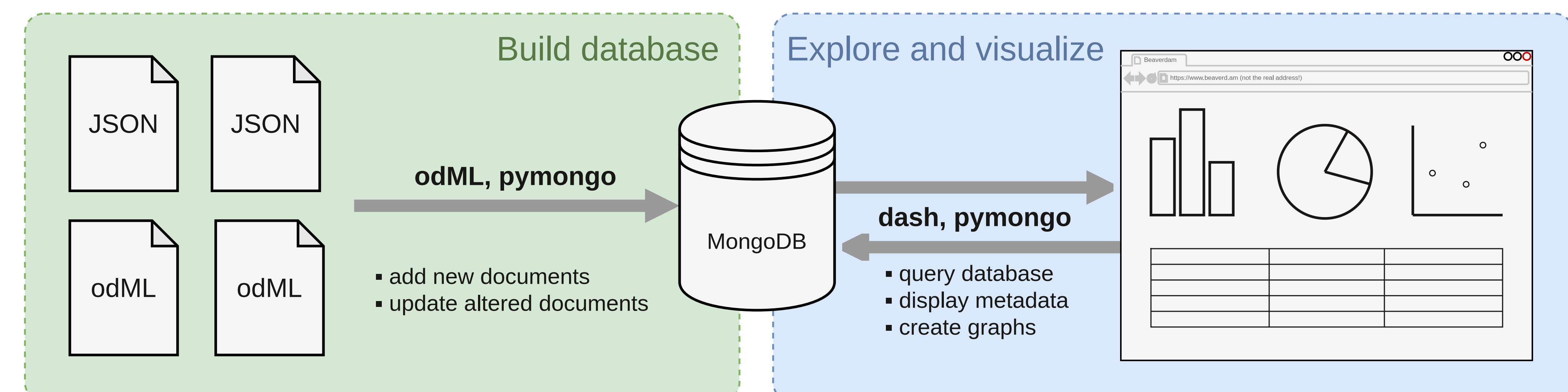
Build, Explore, And Visualize ExpeRimental DAtabases of Metadata

The problem

Metadata often **isn't human-readable** and is spread over **many files**, e.g. for each experiment and instrument.

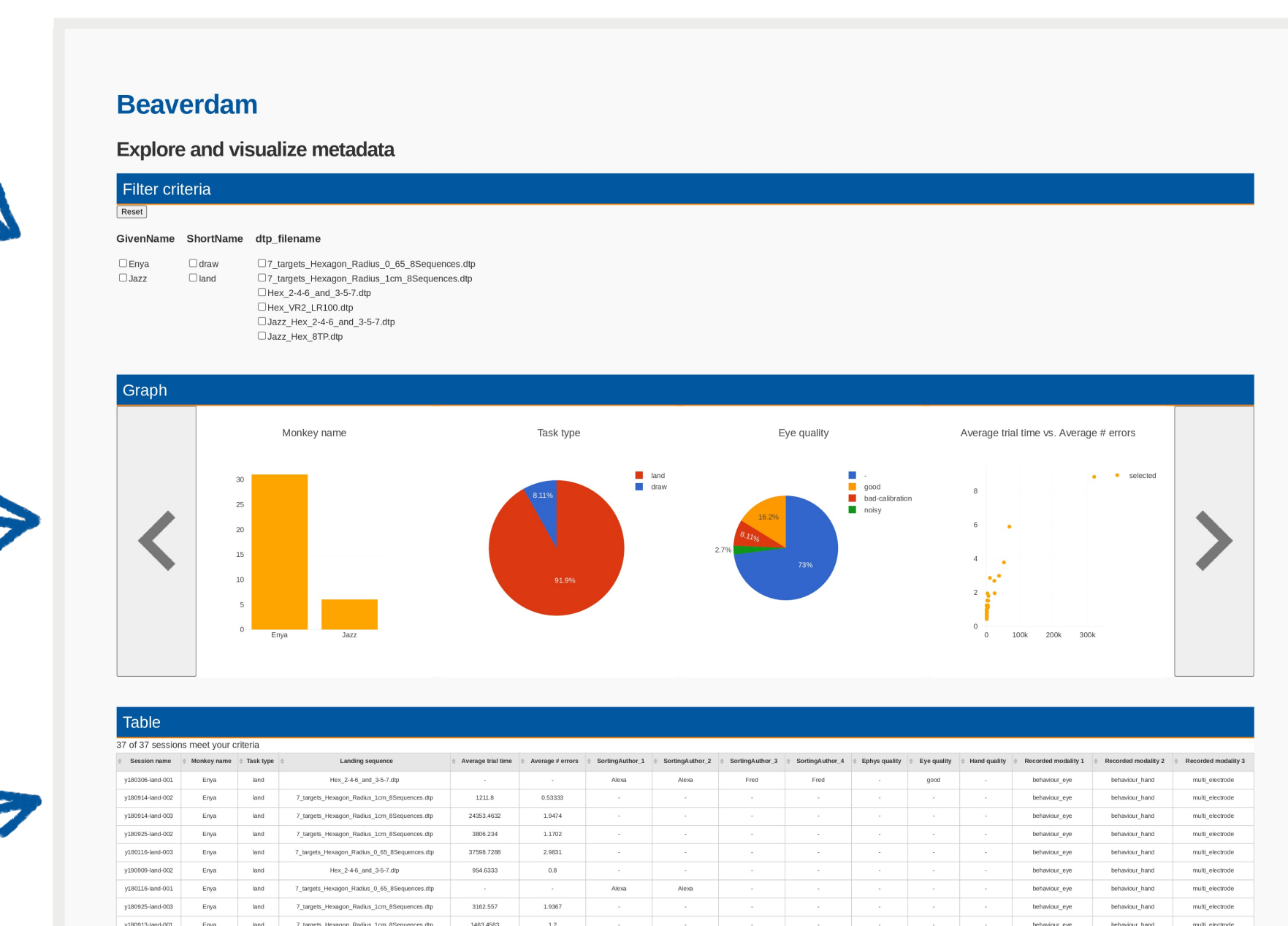
It can therefore be **time-consuming or impossible** to get an overview of all experiments, and to identify which experiments meet criteria for further analysis.

Built with Python



Features

- Search and filter for experiments meeting criteria
- Summarize results with interactive graphs
- View details of experiments meeting search criteria



My instruments all produce different metadata files -- how can I combine their information?

How many experiments were longer than 15 minutes?

In which experiments did subject A use condition B?

Acknowledgements

We thank Anton Pirogov and Annika Strupp for their help in choosing the database for Beaverdam.

An initial version of Beaverdam (called Owl) was produced by Lena Blind, Annika Röthenbacher, Jana Schelter, Julia Wellmann, and Jianing Sun.

This work was supported by the Hub Information of the Helmholtz Metadata Collaboration (HMC) Platform, and the Ministry of Culture and Science of the State of North Rhine-Westphalia, Germany (iBehave).