

Data Management Plan

1. Expected Data

The proposed work will result in several software products, specifically

- A new kaitai target for the awkward array data structure
- An additional release of the XIA python-based analysis library
- A prototype release of an analysis library for the Super Cryogenic Dark Matter Search data
- A release of the Ruby program that allows scanning through binary data

The proposed work will not generate new data. Instead, this work will use data already collected. Small, example datafiles may be prepared for inclusion with the software as test cases.

2. Data Format

The software source code and documentation will be stored in ASCII text files.

Small data files meant to allow testing of the software will be stored in their original, custom binary formats. The descriptions of these formats will be stored in ASCII text files per the data description standard.

The point of the software developed under this work is to provide easy access to data stored in non-standard formats; documentation for the use of this software will be heavily tested.

3. Access to Data and Data Sharing Practices and Policies

The software created by this project will be publicly available for download from a cloud-based repository host such as github or gitlab.

Additionally, all software products will be registered and archived on Zenodo and be available for download from this website as well.

Software products will be publicly available throughout their development; releases will be used to guide users to stable versions. Released versions of the software will all be archived and available on Zenodo.

Papers related to the software products will generally be preceded by a software release; the availability of the software products is otherwise independent from publications.

Individuals and organizations who request the software will be able to download the code through the public channels.

The licenses on the software will be as permissible as possible to encourage maximum community adoption. All code created by the PI's group will be licensed under a permissive open-source license. In the case of software built with another collaboration the license will need to be negotiated.

Permissive open-source licenses (MIT, Apache, CC-BY 4.0) allow others to re-use the software but does not require that they grant the same license to users of their product. This makes it easier for companies to use the software since they're not required to share the source code.

Copyleft open source licenses (GPL, BSD) allow others to re-use the software and requires that they make the software available under the same or similar license terms. Copyleft licenses prioritize keeping source code freely available.

While the PI finds copyleft a sympathetic aim, this work will only be sustainable and impactful with the broadest possible community adoption. Companies in this sphere face significant constraints and the PI feels that permissive licenses align best with the adoption goal.

- New Katai target: MIT license
- XIA python-based analysis library: MIT license
- Analysis library for the Super Cryogenic Dark Matter Search data: still under discussion but likely Apache v3
- Ruby program that allows scanning through binary data: GPL 3.0

4. Policies for Re-Use, Re-Distribution

Scientists who use the code produced as a result of this work will be asked to cite the version they use using the appropriate Zenodo DOI.

All software downloads will include citation instructions.

Written content on the project website and documentation and any images will be licensed under CC-BY 4.0.

All code will be licensed with an open-source license that allows re-use of the code for commercial purposes.

Articles written about the software use and development will be published as open access wherever possible.

5. Archiving of Data

All software and documentation will be archived on Zenodo.

Zenodo is a collaboration between CERN and OpenAIRE and has an operation plan for the next twenty years.

Zenodo saves their data on two physically distinct disk servers. For low-use data, they reserve the right to store the data to tape.