

## DATA MANAGEMENT PLAN

This SSI will not directly produce significant amount of traditional scientific data; however, it will produce a significant amount of software (source code), curricula for Data Carpentry, and documentation. Additionally, during the development process, ephemera such as code review comments, mailing list discussions and so forth will be generated.

The data collected has no privacy implications. Metadata for produced source code will be in the form of mercurial or git changesets and code review comments, both of which are industry standard. Licensing of source code, as detailed in the full proposal, will be under permissive, non-copyleft BSD licenses, which allow re-use, re-distribution and the production of derivatives. If sample data is required for development, we will either generate that sample data as a part of the project and license it under Creative Commons Zero, or will solicit it from community members and distribute it only under their terms.

Design documents developed in the course of this proposal will be versioned and available freely both in original source form and in rendered form on the `yt` website.

### Management of `yt` Source and Support Materials

All of the software described in this data management plan is developed in a completely open source manner. `yt` is freely available and open source. Any newly developed software as part of this SSI will be made available under an open source license *at the time of development*, both prior to and during code review for acceptance into the main `yt` distribution. We will store our source code on `BitBucket.org`; prior to review and acceptance upstream the source will be stored either under a specialized account or under personal developer accounts, with the intention of merging upstream as quickly as possible.

Documentation for `yt` developed in the course of this SSI will take three forms. The first will be narrative, example or design-oriented textual documentation. This documentation will be version-controlled and made available at the time of writing. Currently, all documentation for `yt` is currently available in both source and built form. The second form of documentation will be in the form of executable, narrative Jupyter notebooks. This, too, will be version-controlled and made available at the time of writing. The final form of documentation will be in the form of recordings of presentations about `yt` and the various sub-projects and milestones enumerated in the proposal. As often as possible these will be recorded and made available on YouTube. In this manner, we will endeavor to ensure that all educational materials are readily and freely available.

The curricula developed for Data Carpentry will be developed in the open, versioned using git, and provided through standard Data Carpentry channels. These will be licensed, made available, and developed in accordance with standard Data Carpentry practices and licensing schemes, which are permissive, non-copyleft and freely available. Additionally, developed lessons and lesson plans will be distributed through the Midwest Big Data Hub website.

We emphasize that all data products, discussions, documentation and source code are both preserved for the purposes of provenance and reproducibility and provided openly and publicly. Where we have built on external software projects, we will submit our contributions upstream in a license-compatible way to enable their inclusion.