

Data Management Plan for the Cosmic Explorer Project

Cosmic Explorer Gravitational-Wave Observatory Conceptual Design. The Cosmic Explorer Project is currently in the Conceptual Design phase. The primary products of this research will be design documentation, such as subsystem requirement specifications, interface specifications, underlying models and software, information and data on potential Cosmic Explorer sites, as well as the Project Execution Plan, Design and Development Plan, Cost Book, Cost Model Data Set, Systems Engineering and Interface Management Plans and other management documentation. The primary repository for this documentation is the Cosmic Explorer Document Control Center¹, as well as the existing GitHub and GitLab Cosmic Explorer repositories². PASSCAL seismic data will be archived at the Incorporated Research Institutions for Seismology (IRIS) Data Management Center³ (DMC). The products will also include related scientific articles, which will be published as open access via the arXiv pre-print server, stored in the Cosmic Explorer Document Control Center, and submitted to a journal customary to the field.

Cosmic Explorer Project and Consortium Computing Infrastructure. Syracuse University operates the Cosmic Explorer Project and Consortium computing infrastructure that will be used to manage the data for this project. This infrastructure is deployed using Internet2's InCommon Trusted Access Platform architecture. The hardware is provided by the Syracuse University Academic Virtual Hosting Environment (AVHE) which uses the high-availability VMWare environment to host servers that provide user registry, mail, and document control services for the Cosmic Explorer Consortium. All of these services use InCommon Research and Scholarship Federated Identity Management, allowing users from participating campuses to log in using their campus credentials. An ORCID gateway is provided for users whose campus has not yet adopted InCommon Research and Scholarship.

The Cosmic Explorer user registry is implemented using Internet2's COmanage platform, which provides federated identity management services and enables authentication and authorization in a single system. COmanage automatically creates user accounts and access controls for the Cosmic Explorer services: GitHub, mailing lists, and access to the Document Control Center. Mailing lists are provided using the widely supported GNU Mailman platform. The Cosmic Explorer Document Control Center uses a clone of the LIGO Document Control Center software, which is a fork of the Fermilab DocDB software. The software thus has demonstrated its scalability, ensuring the support for the envisioned growth of the Cosmic Explorer project.

Access to data and data sharing practices and policies. The Cosmic Explorer Document Control Center (CEDCC) supports several levels of access permission (director, management, project, consortium, public). All final products and deliverables will be made available to the public. Project members will have access at the project level, while the more restricted access level will only be used if necessary, such as for confidential documents and proprietary vendor information.

Policies and provisions for re-use, re-distribution, and the production of derivatives. Whenever possible, Conceptual Design Documentation will be made public through the CEDCC, and, if appropriate, submitted to the arXiv pre-print server. Papers will be published in a peer-reviewed conference papers, scientific journals, or books that publish in English and/or languages and preferred modes of sharing as determined by local and Indigenous communities associated with potential Cosmic Explorer observatory locations. All scientific articles will be made available as green open access via the arXiv pre-print server,

¹<https://dcc.cosmicexplorer.org>

²<https://github.com/cosmic-explorer>, <https://gitlab.com/cosmic-explorer>

³<https://www.passcal.nmt.edu/content/data-archiving>

the CEDCC, and submitted to a journal customary to the field. All software and data will immediately be made available to the community freely under open source software distribution licenses. Publications, data, and software may be distributed freely in academic environments and may be cited with appropriate attribution according to standard academic practice. PASSCAL seismic data will be archived at the IRIS Data Management Center using the Standard for the Exchange of Earthquake Data (SEED) file format for data and metadata, adhering to the IRIS PASSCAL Data Delivery Policy⁴. Upon approval by IRIS, the complete seismic dataset will be openly accessible through the Data Management Center Web Services⁵. Indigenous knowledge holders will be listed as co-authors or cited as they deem appropriate, and all dissemination of information will respect Indigenous data sovereignty (see below).

Center for Evaluation and Research for STEM Equity (CERSE) Data. The University of Washington Center for Evaluation and Research for STEM Equity (CERSE) will collect and analyze demographic data on the Cosmic Explorer teams, to be summarized and incorporated into reports.

Standards: Methods used to collect data are electronic surveys and interviews. The analysis of the qualitative data will use a pre-determined coding schema as well as open coding to identify unanticipated themes in the data. Analysis of the quantitative data will involve descriptive statistics.

CERSE Data Access: Surveys and interview notes will be stored on a password-protected, restricted access server. Data will be reported in such a way that individuals will not be identified such as by stakeholder group or institutional affiliation.

Indigenous Data Sovereignty and Data Sharing. Every process in the data management plan is subject to CARE Principles for Indigenous Data Governance developed by the International Indigenous Data Sovereignty Interest Group.⁶ In addition to the above, a priority of the Cosmic Explorer Indigenous Partnership Program (IPP) will be to work with Indigenous communities with whom partnerships develop to build protocols for the use and sharing of data in data systems, facility conceptual design, computing infrastructure, and future uses in a way that is beneficial to these communities. This work will collaborate with the IPP to ensure these data management processes are in compliance with protocols developed in partnership with Indigenous communities.

Data Management Plan for the Cosmic Explorer Observatory. This document will evolve into the data management plan of the Cosmic Explorer observatory. The current plan is intended to be sufficient for the Conceptual Design phase, and a number of steps of evolution are anticipated as we pass through Preliminary, Final, Construction, and Observatory phases. It will ultimately include the management of scientific output data (strain data, skymaps, timestamps, etc), establishing policies for data collection, access, sharing, re-use and production of derivative works, and preservation. To maximize the scientific return of CE, our long-term vision is for Cosmic Explorer data to become open to the extent possible, ensuring the highest possible scientific impact of the data. The full data management plan for the Cosmic Explorer Observatory will be developed during the design stage of the project.

⁴<https://www.passcal.nmt.edu/content/general-information/policy/data-delivery-policy>

⁵<https://service.iris.edu/>

⁶Collective benefit, Authority to control, Responsibility, and Ethics (CARE) Principles, DOI: 10.5334/dsj-2020-043