



NASA Roses F.7 (OSTFL) LOI

*Note: This was submitted on **May 3, 2024** . The submission was limited to 4000 plain text characters.*

Proposal Title: Supporting the development and adoption of Zarr

Type of award: Foundational

License of the project: MIT license

To whom it may concern:

We are writing to express our intent to apply to the ROSES-24 F. 7 Support for Open-Source Tools, Frameworks, and Libraries (OSTFL) funding opportunity to support the development and adoption of [Zarr](#) across the NASA science community. Zarr is a data format specification and a set of community-developed open source libraries for accessing multidimensional arrays and structured metadata. Today, Zarr is being utilized in multiple science mission areas across NASA, including Earth Science, Biological & Physical Sciences, Astrophysics, and Sun and Heliophysics.

The Zarr project is made up of two main components: a community developed open source specification and independent implementations of the specification in various languages (e.g. Python, C++, Rust).

1. The specification defines the Zarr format for chunked compressed multidimensional arrays. Critically, the specification is storage backend agnostic, allowing it to take advantage of a wide array of storage backends, including modern cloud object storage. In 2023, a new version of the Zarr core specification, V3, was approved. The new specification was motivated by community needs for greater interoperability and extensibility.
2. The various implementations of the specification provide APIs for reading and writing arrays and metadata. Zarr-Python is the canonical and most widely used of the implementations today and will be the focus of our proposal. It provides a high-level API allowing users to efficiently read and write arrays and metadata, integrating with the broader Scientific Python ecosystem through a NumPy-like interface.

Our proposed work plan will have three main objectives, all supporting the growing adoption of Zarr across the NASA science community.

1. Ongoing maintenance and development of Zarr-Python. This objective covers the variety of tasks needed to provide the community with the highest quality Zarr API possible,

from testing and bug-fixes to incremental development to improve the library based on user feedback.

2. Development of high-value extensions to Zarr-Python. The development of V3 specification has opened the door to a number of high-value extensions to the Zarr specification that were previously not possible. Work under this objective will focus on the development of specific extensions in Zarr-Python (e.g. variable chunk grid, storage transformers) and associated work on the specification.
3. Establishment and demonstration of Zarr metadata conventions. While the Zarr specification is relevant for any application that requires storage and efficient access to N-dimensional arrays, the growing adoption of Zarr has shown the need for additional community-established conventions for the organization of data and metadata within Zarr. Today there are a number of burgeoning metadata conventions for Zarr V3, including GeoZarr, OME-Zarr, NetCDF-Zarr, and multiscale data pyramids. This objective will focus on establishing and demonstrating the use cases of high-impact Zarr V3 metadata conventions for multiple science mission areas.

In summary, our intent to apply for the OSTFL funding opportunity signals our dedication to advancing Zarr's adoption within the NASA science community. With the recent approval of the V3 specification, Zarr is poised to make a significant impact, offering unparalleled versatility and efficiency in multidimensional data management. Through our proposed objectives, we aim to capitalize on this momentum, ensuring ongoing development, creating high-value extensions, and establishing vital metadata conventions. We are enthusiastic about the potential for Zarr to revolutionize scientific data handling and analysis.

Joseph Hamman
CTO
Earthmover, PBC
joe@earthmover.io