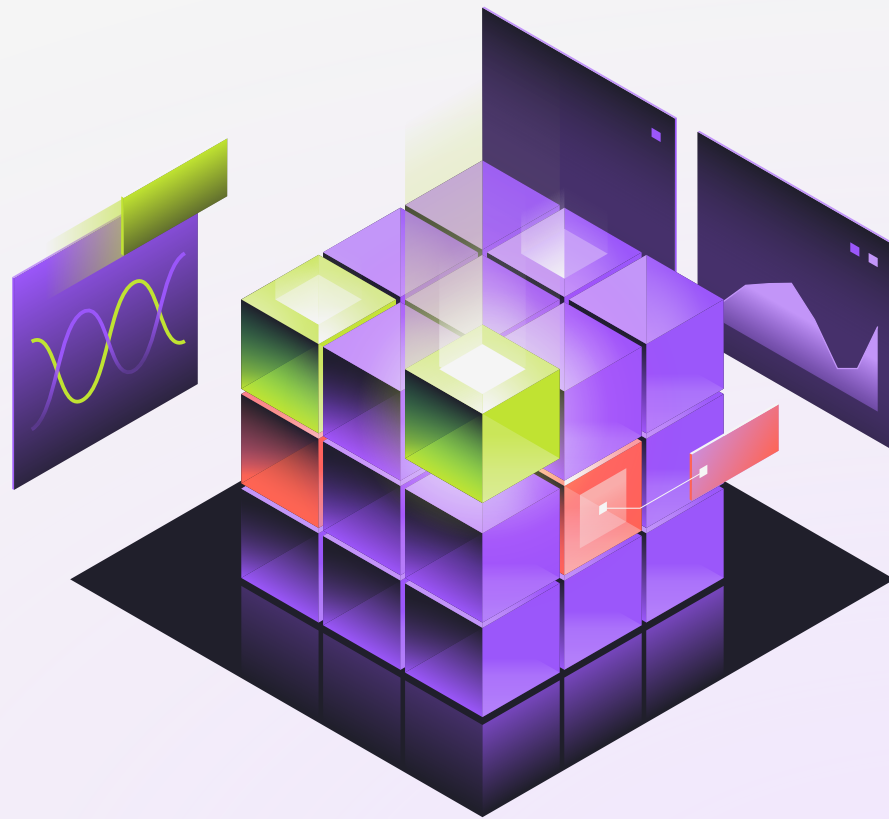




the future of earth system data infrastructure

KM-SCALE SUMMIT | 2026

Ryan Abernathey
CEO, Earthmover PBC



MITgcm LLC4320

CIRCA 2014

Dimitris Menemenlis (NASA
JPL) & Chris Hill (MIT)

Deformation radius: **10 km**

Spatial resolution: **1-2 km**

Grid points: **10^8**

Vertical levels: **100**

One scalar field: **40 GB**

State variables: **10**

Temporal resolution: **1 hour**

Run length: **1 year**

Raw dataset size: **40 PB**

Compressed dataset size: **2 PB**



MY JOURNEY



Earth-System Scientist

PhD in Climate Physics at MIT

Obsessed with ocean mesoscale eddies

Tenured Professor at Columbia University /
LDEO until 2023.



Open Source Hacker

Core contributor to **Xarray** and other popular
Python data tools

Member of the **Zarr** steering council

Co-creator of the **Pangeo** project



Startup Founder and CEO

Building Earthmover!



IMPORTANT TRENDS IN EARTH-SYSTEM MODELING IN 2026



KM-Scale Models

Earth system models are reaching higher and higher spatiotemporal resolution.



AI/ML Workloads

AI has a **voracious appetite** for training data. Researchers need **high throughput access** to **entire archives** for training and validation.



Agent-based Workflows

Query data with **natural language**, not code.

Deploy agents to explore data and **perform independent research**.



Cloud Computing

Cloud computing enables any organization to access **large-scale storage and compute** without huge up-front capital cost. **Object storage** over filesystem storage.



Private Sector Participation



Microsoft



Brightband



WindBorne



Jua



Open Science

Journals and funders increasingly value **open sharing of code and data** as part of the publication process.



earthmover

🕒 Our Mission

To empower **people** to use **scientific data** to solve humanity's **greatest challenges**.

- Founded in 2022 as a **public benefit corporation** by **Ryan Abernathey** and **Joe Hamman**.
- 14 people spread across US, Europe, and South America.
- "Open core" business model - investing in **community open source** and providing a **managed cloud data platform** for institutions.



ACCELERATING EARTH-SYSTEM SCIENCE

- 1 evolving data storage
- 2 evolving data-intensive computing
- 3 connecting the ecosystem



EVOLVING DATA STORAGE

from files to arco data cubes

Humans and **agents** don't want to deal with thousands of individual files. They want ARCO data cubes.

Analysis Ready

All timesteps, variables, ensemble members, etc. are organized into a single logical object (data cube).

Cloud Optimized

Applications can efficiently stream data directly from cloud object storage with low latency and high throughput.

Zarr is the most widely used ARCO data format for Earth-system-model data.

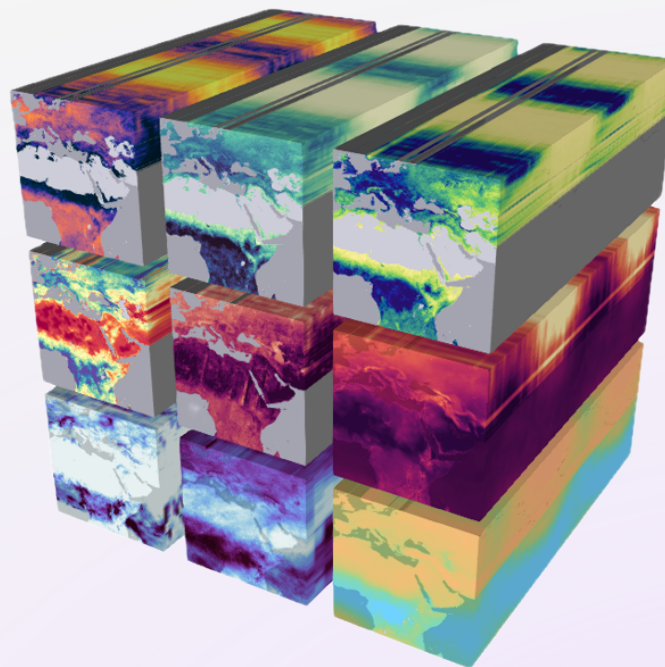


Image credit: Mahecha et al. 2020. <https://doi.org/10.5194/esd-11-201-2020>



EVOLVING DATA STORAGE

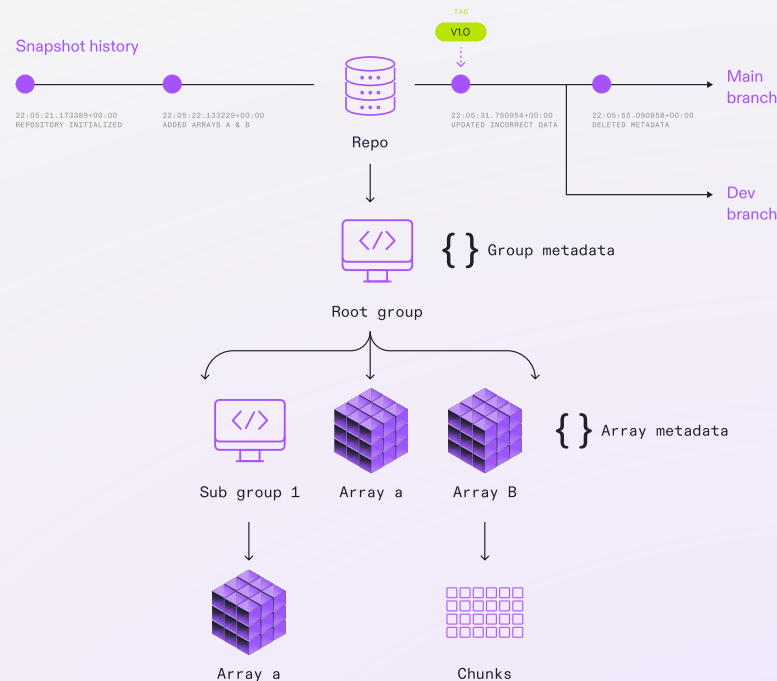
icechunk: zarr for production

A specialized **Zarr Store** which supports:

- **ACID transactions** in a distributed context
- **Version history** (snapshots, branches, tags) with optional expiration and garbage collection
- Complex metadata operations (move, shift, reindex)
- **Virtual chunks** - pointers to blobs of numerical data in external files (NetCDF, GRIB)
- **Open source** (Apache 2.0), **serverless**

Implemented in Rust with a **high-performance I/O layer**.

<https://icechunk.io>





EVOLVING DATA STORAGE

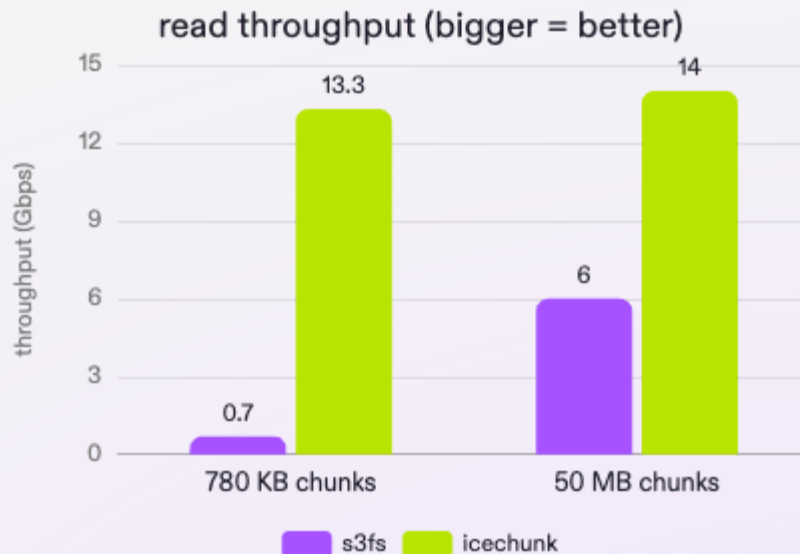
icechunk: zarr for production

A specialized **Zarr Store** which supports:

- **ACID transactions** in a distributed context
- **Version history** (snapshots, branches, tags) with optional expiration and garbage collection
- Complex metadata operations (move, shift, reindex)
- **Virtual chunks** - pointers to blobs of numerical data in external files (NetCDF, GRIB)
- **Open source** (Apache 2.0), **serverless**

Implemented in Rust with a **high-performance I/O layer**.

<https://icechunk.io>



Single-node throughput benchmarks (15 Gbps NIC)

<https://www.earthmover.io/blog/i-o-maxing-tensors-in-the-cloud/>



EVOLVING COMPUTE

zax: pure-rust tensor compute engine

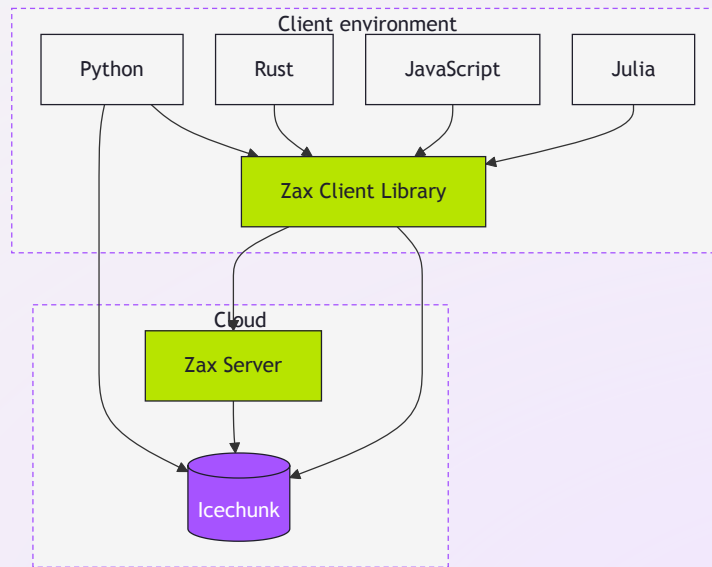
Inspired by:

- **Xarray** - Same NetCDF-based data model and user-friendly API
- **Polars** - Database-style planner / optimizer / executor architecture with full-stack awareness

Key Features

- **No Dask!** - End-to-end execution in Rust.
- Bindings to **all major languages**.
- **Client / server mode** — Run compute next to data
- **Open Source** (Apache 2.0)

data flow



(Public release Fall 2026)



EVOLVING COMPUTE

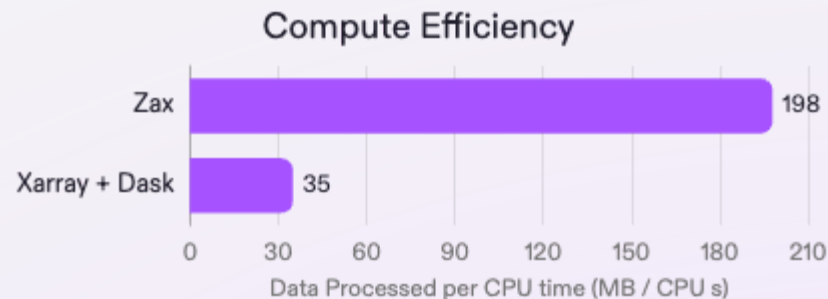
zax: pure-rust tensor compute engine

Inspired by:

- **Xarray** - Same NetCDF-based data model and user-friendly API
- **Polars** - Database-style planner / optimizer / executor architecture with full-stack awareness

Key Features

- **No Dask!** - End-to-end execution in Rust.
- Bindings to **all major languages**.
- **Client / server mode** — Run compute next to data
- **Open Source** (Apache 2.0)



Important

Zax is > 5x more efficient at processing data than the dominant Python / Xarray / Dask stack.

Faster science and **lower energy usage** per calculation.



PLATFORM

the earthmover stack

platform

Flux

Scalable Tensor Compute Services

Tiles

EDR

DAP

Zax Server

Zax SQL

OpenEO

Arraylake

Data Management and Governance

Data Catalog

Credential Vending

Access Controls

Data Marketplace

open source

Icechunk

Tensor Storage Engine

ACID Transactions

Data Versioning

Best-in-class I/O Performance

Virtualization of Legacy Formats

cloud

BYOB

Bring your own bucket (or use ours)



generic S3-compatible



TRACTION

earthmover public sector adoption



Providing core data infrastructure for the **Tipping Points** project, integrating 26 independent research groups.



Helping **Integrated Modeling Virtual Institute** unlock value from model data in the cloud. (MERRA-21C, ECCO, etc.)



Building **CIRRUS** with Booz Allen Hamilton. Pilots ongoing within OAR (PMEL, GFDL).



Building the Destination Earth Data Lake **Data Cube** project, in partnership with EODC, DevSeed, etc.



Scoping pilot for operational / research forecast data delivery system.



Scoping pilot for BARRA model delivery.



WHAT THIS ENABLES



Nikolay Koldunov • 1st

Climate scientist. Senior scientist at Alfred Wegener Institute
4mo • Edited •

Our team built a demo agent application for working with ERA5 data based on [Earthmover](#) Arraylake. Fast data access lets us pull the right slices of ERA5 quickly, iterate, and visualize, without turning every question into a coffee break.

What Eurus does:

- Natural language → precise retrieval + analysis for ERA5 workflows
- Multiple interfaces: interactive CLI, web chat UI, and an MCP server for tool-using assistants/IDEs
- Built-in Python REPL (pandas/xarray/matplotlib/cartopy)

See an example of the workflow in the video below.

Repo: <https://lnkd.in/deQMCNmH>

Live Hugging Face deployment: <https://lnkd.in/dAbhnKSr>

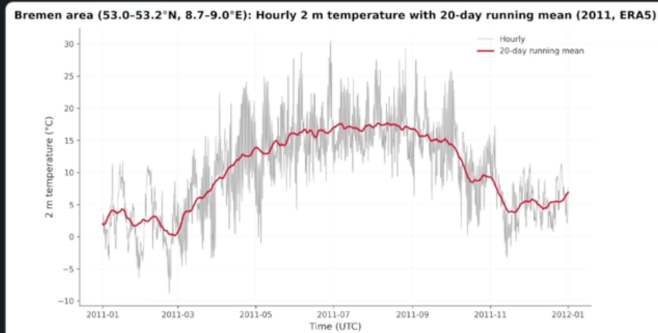
⚠ To run the demo (locally or on HF), you'll need:

- an OpenAI API key
- an Arraylake API key

If you're building LLM apps over large scientific arrays, this is a great example of how storage + retrieval speed unlocks new interaction patterns. It also offers a glimpse of what data processing and retrieval might look like in the next few months (not years :)).



Overlay added: a **20-day running mean** (computed as a centered **480-hour** rolling mean) plotted on top of the hourly Bremen 2 m temperature series for 2011.



Enlarge

Download

Show Code

plot temperature in Helsinki for the same time in h

Send

Enter to send, Shift+Enter for new line



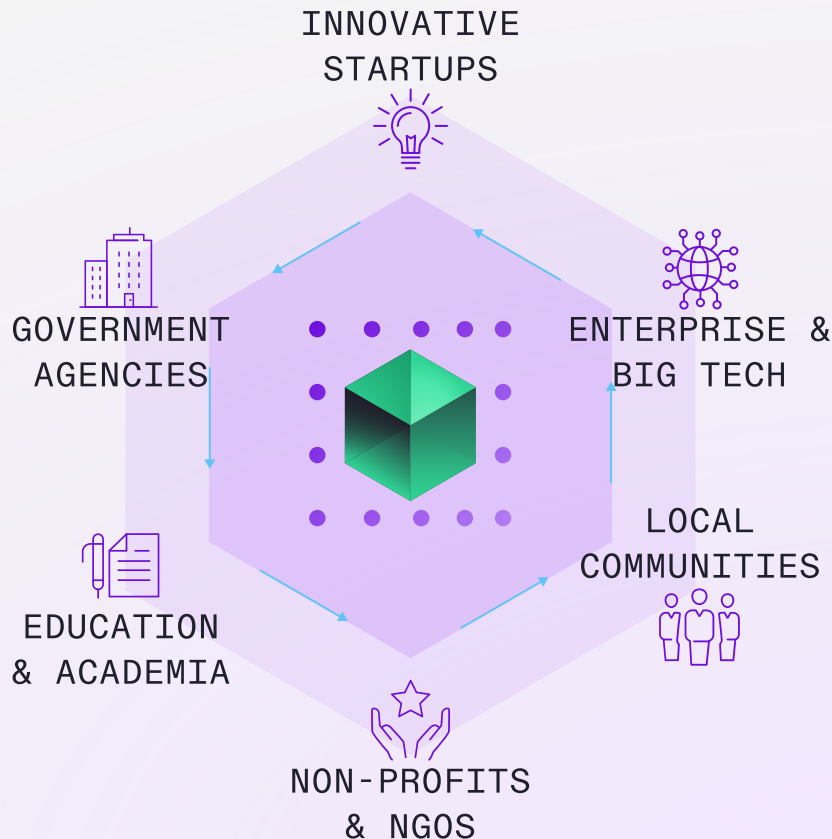
OUR VISION

a connected ecosystem

We envision a world in which scientific data can be **explored, visualized, analyzed, shared, and built-upon effortlessly.**

Our technology provides a foundation for people and organizations to create new scientific knowledge and data-driven products which will help humans thrive in sustainable harmony with each other and with our planet.

Our company becomes a key trusted partner to industry, government, and academia while helping to sustain a multi-stakeholder open-source ecosystem.





earthmover

earthmover.io - 📌 Join mailing list to stay informed

