Meeting the Challenges of DAS Data Archival & Dissemination

Jerry A Carter
EarthScope Data Services

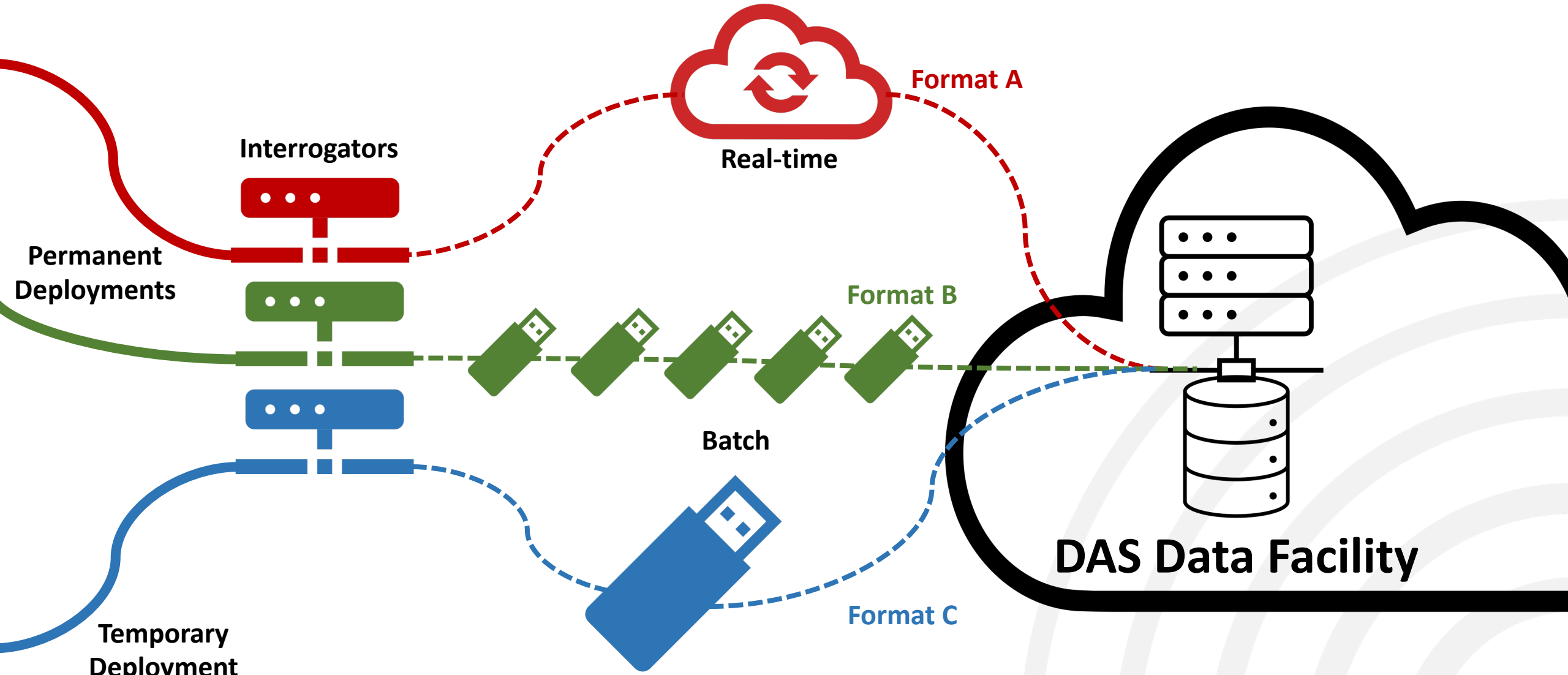
NSF USGS science for a changing world NASA GAGE SAGE Operated by EarthScope Consortium

Takeaway

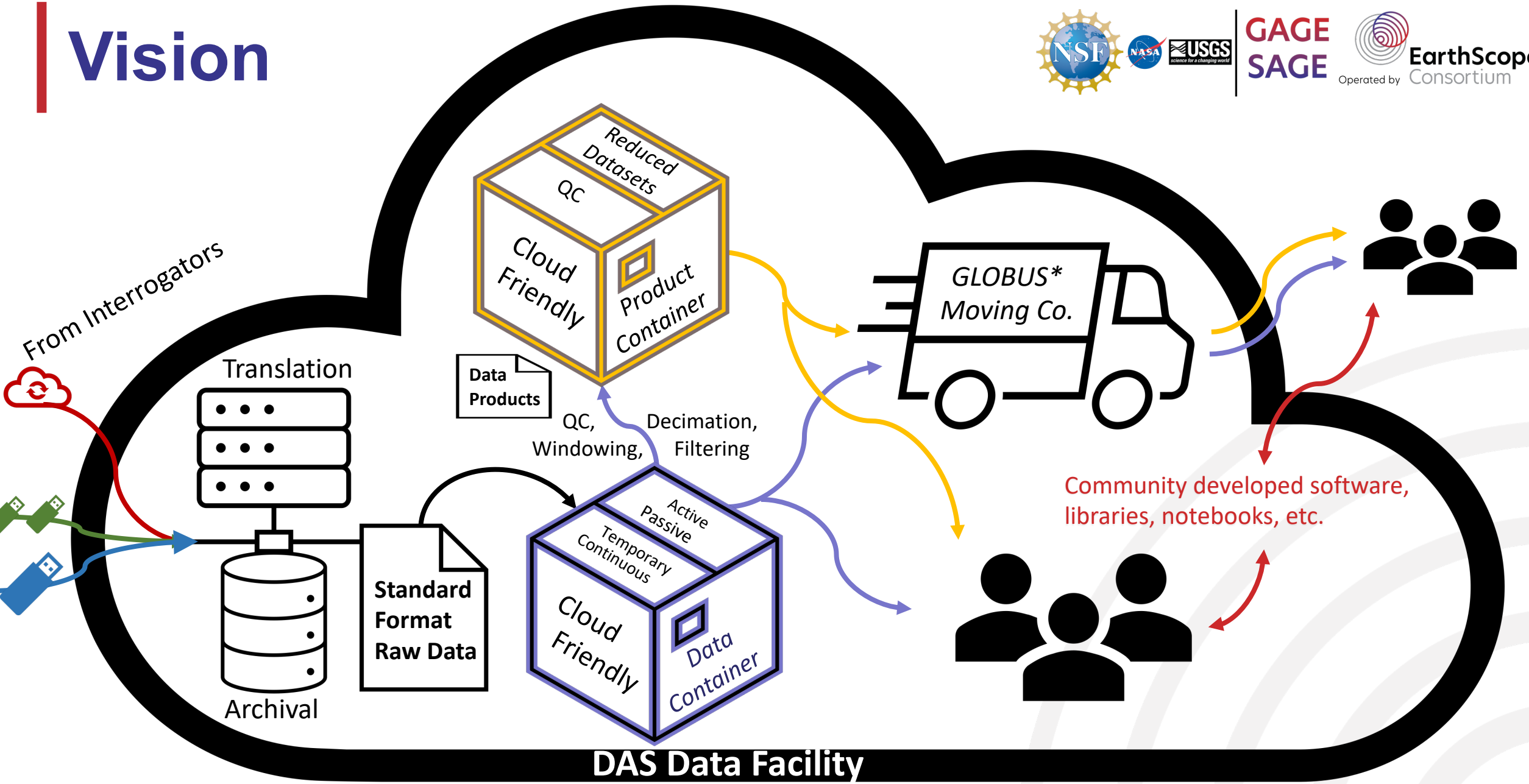
- A facility supporting open DAS data handling and persistent data storage provides significant value to the research community.

- Many of the initial technical challenges have been (or are being) met through community involvement.

Vision


Format A
Real-time
Interrogators
Permanent Deployments
Format B
Batch
Temporary Deployment
Format C
DAS Data Facility

Vision



* Chard, K., Foster, I., & Tuecke, S. (2017). Globus: Research data management as service and platform. In Proceedings of the practice and experience in advanced research computing 2017 on sustainability, success and impact (pp. 1–5).

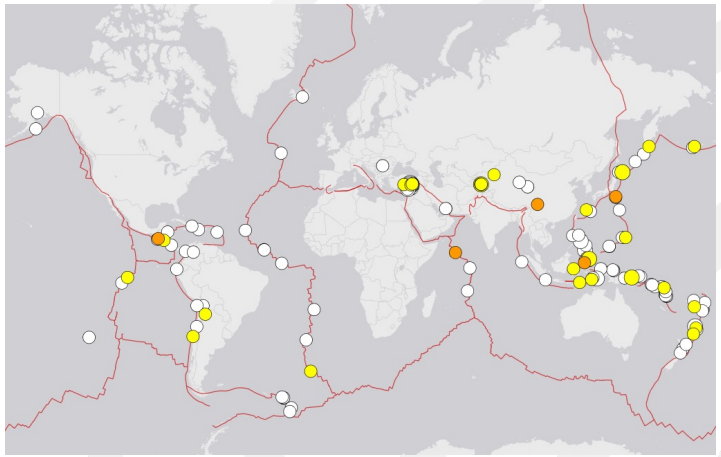
Vision

A DAS data facility for the research community with:

- Standardized data containers/formats, metadata content/formats
- Proximity to compute resources suitable for large DAS dataset research
- Standardized data products derived from original data
- Ability to deliver large datasets & products to researchers
- Perpetual storage (PIs focus on research, not data stewardship)
- Public domain DAS data processing tools

Progress Examples

- **Standard Data Formats / Containers Suitable for Cloud**
  - ➢ TileDB, Zarr (Yiyu Ni, Marine Denolle, EarthScope)
- **Metadata Content**
  - ➢ Metadata Model White Paper (Mellors, R. J., Hodgkinson, K. M., Hui Lai, V., & the DAS RCN Data Management Working Group)
- **Proximity to Computational Resources**
  - ➢ Common Cloud Platform (EarthScope DS)
- **Standardized Data Products**
  - ➢ Global DAS month (Andreas Wuestefeld)



Progress Examples

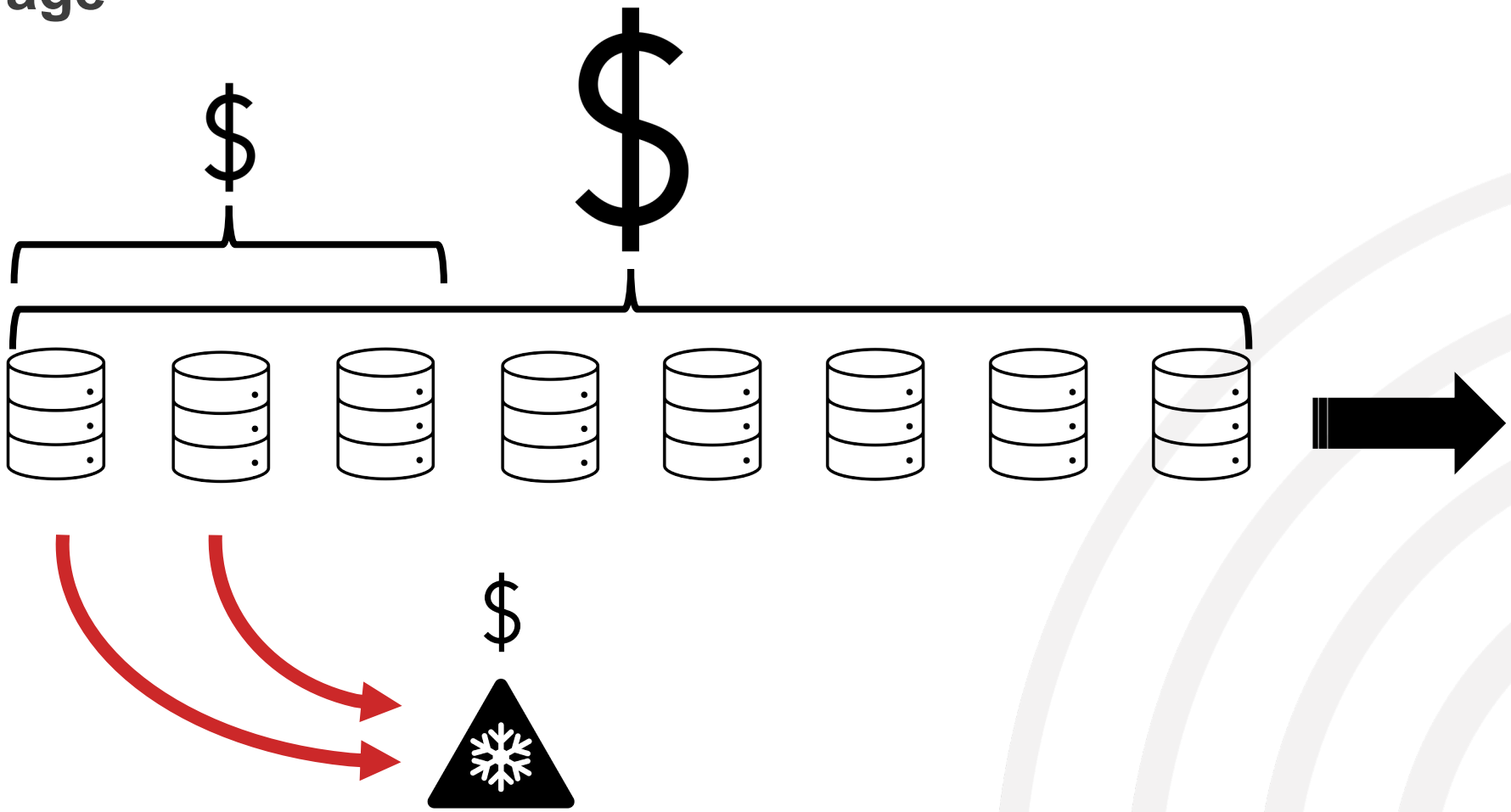
- **Large Data Transport**
  - PubDAS: A PUBlic Distributed Acoustic Sensing Datasets Repository for Geosciences, SRL (2023) 94 (2A): 983–998. https://doi.org/10.1785/0220220279 (Spica et al.)

- **Perpetual Storage**
  - GorDAS (Andrew Barbour, USGS)

- **Public Domain DAS Data Processing Tools**
  - MinIO (Yiyu Ni, Marine Denolle)
  - DASDAE, DASCore (2022 AGU abstract) - DASDAE: Open-source Python Packages for Easier DAS Data Use (Eileen Martin et al.)

Challenges

Data Storage

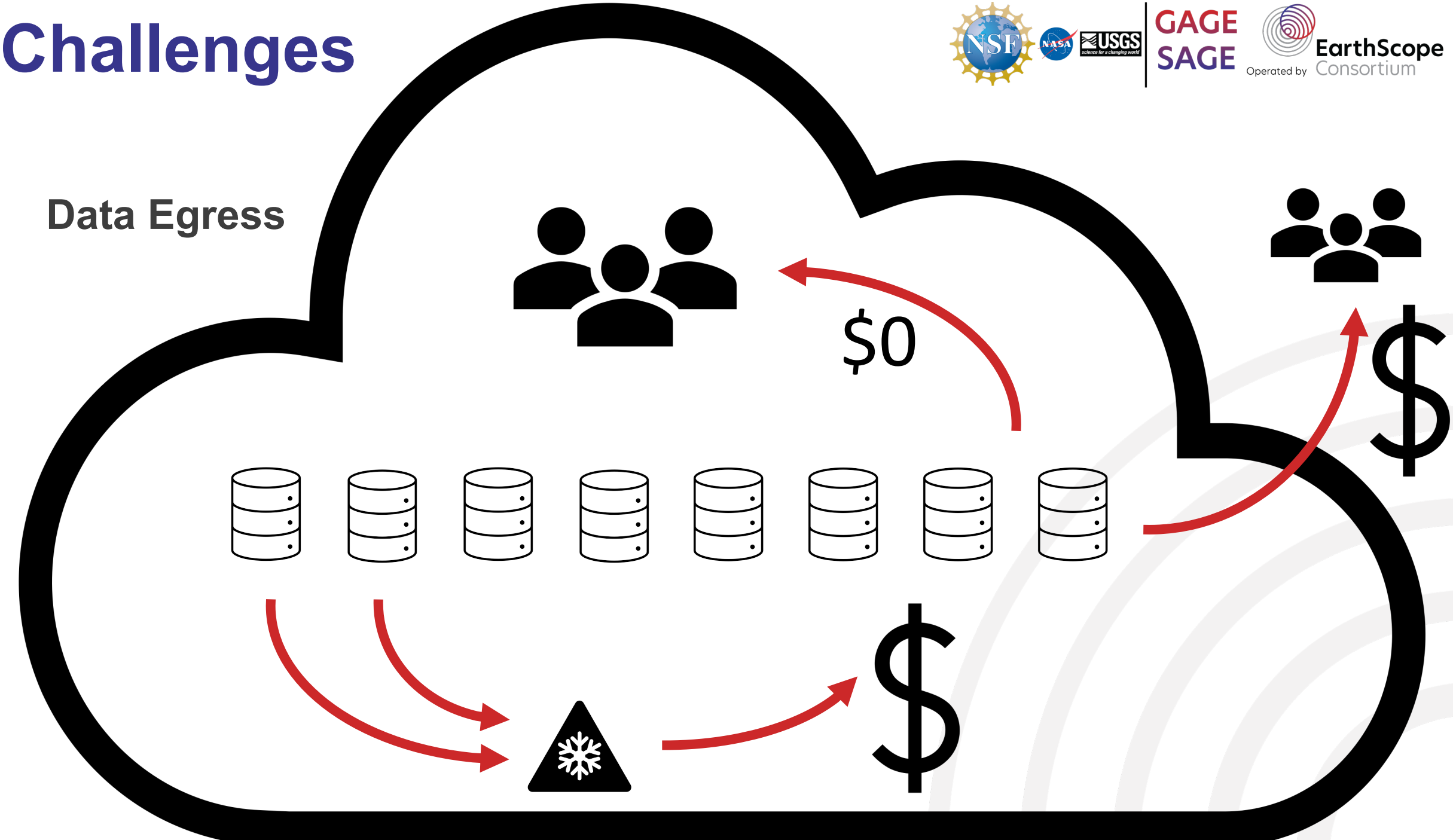


Challenges

GAGE SAGE

Operated by EarthScope Consortium

Data Egress



Way Forward

DAS Data Facility Funding

- Express community support for DAS Data Facility
- Emphasize importance to missions of stakeholders (NSF, DOE, USGS, DOD, …)
  - ✓ DAS is enhancing current seismological research, hazard monitoring, environmental monitoring, national security monitoring, …