A Scorecard to Measure Earth and Environmental Science Repository Resilience Both During Normal Times and in Times of Crisis.

Lesley Wyborn[1], Joseph Gum[2], Ruth Duerr[3], Reyna Jenkins[4], Rebecca Farrington[5], Tim Rawling[5], Kelsey Druken[6]

[1]NCI, Australian National University, Canberra, Australia; [2]ESIP, USA, [3]Ronin Institute for Independent Scholarship, USA, [4]World Data System, Victoria, Canada, [5]AuScope Ltd, Melbourne, Australia, [6]ACCESS-NRI, Canberra, Australia

We acknowledge and celebrate the First Australians on whose traditional lands we meet and pay our respect to our Elders past, present and future

Artist: Lynnice Letty Church – Tribes: Ngunnawal, Wiradjuri & Kamilaroi (ACT and NSW)
Gadi - "to search for" in Ngunnawal language - January 2020 for NCI Gadi Supercomputer

Abstract

Earth and environmental (E&E) datasets play a crucial role in Sustainable Development of our Planet, contributing to the prediction of natural hazards, effective development of natural resources, long term monitoring of vegetation changes, etc. Since the 1990s there have been concerted efforts to create global data networks (e.g., OneGeology, Federated Digital Seismology Network (FDSN), Global Biodiversity Information Facility (GBIF), Earth System Grid Federation (ESGF). As these networks become accepted as critical inputs into E&E data supply chains, their vulnerability is becoming of concern. Contemporary events have highlighted the need to formally assess the resilience of infrastructures underpinning these networks.

Recently members of the Sustainable Data Management Cluster of the Earth Science Information Partners (ESIP) developed a repository scorecard (https://zenodo.org/records/15208172) to enable repositories to assess their resilience based around four scenarios:

1. Incoming Natural Disaster forecast to hit the repository in 48 hours. The focus is on local destruction of the facility;
2. Loss of Organizational Funding requiring shut down within one month. The focus is total loss of the repository, but with time for mitigating actions;
3. Cyberattack/Organisational Infiltration with hostile agents gaining control of the cyberinfrastructure. The focus is on sudden denial of access to data; and
4. Loss of Technical Expertise critical to running the repository. The focus is on loss of expertise to keep the repository running.

This presentation will outline key elements of measuring repository resilience, and will present suggestions to develop international consensus on policies and agreements for sustaining critical global E&E data supply chains.

Why are we here?

Earth Science Information Partners

1. Events of early 2025 were a wakeup call: many who had their data and/or whose research depended on data in repositories started to worry:
   a. How safe are my research dependent datasets?
   b. How secure are the repository infrastructures they are stored in?
2. The US-based ESIP Sustainable Data Management Cluster sprang into action:
   a. Developed the Repository Crisis Scorecard within 2 months;
   b. Led to asking just what are our impactful Earth & environmental datasets
3. In Australia, we were inundated with requests to host threatened datasets: led to informal discussions on global collaboration of E&E RIs to create international offsite backups -
   a. But which were the primary datasets?
   b. What should we prioritise?
4. At a policy level, the Academies of Science interntionally were asking whether we could form policies to sustain critical data supply chains.

Not mentioning any names.....

Wikipedia
The Free Encyclopedia

Search Wikipedia | Search

2025 United States government online resource removals

3 langua

Article | Talk | Read | Edit | View history

From Wikipedia, the free encyclopedia

Contents hide

(Top)
Background
Removed and modified content
CDC website
Science and research websites
Justice and crime websites
Healthcare and social services websites
Other websites
Datasets
Legal responses
Reactions
References
Further reading
External links

The **2025 United States government online resource removals** are a series of web page and dataset deletions and modifica across multiple United States federal agencies beginning in January 2025. Following executive orders from President Donald T administration, governm
changes primarily affec
environmental policy, a
which saw over 3,000 p
While some content wa
and sparked legal chall

Background [edi

Agencies of the United
applications which rely
practice.[2] There is little
deletions occur.[2][3] De
modifications to public
previous government ch

Science and research websites [edit]

See also: NOAA under the second presidency of Donald Trump

In January, NASA undertook a comprehensive removal of DEI-related content from its public-facing websites. An internal directive instructed employees to "drop everything" and immediately eliminate references to terms such as "DEIA", "indigenous people", "environmental justice", "underrepresented groups/people", and content specifically targeting women, including content about "women in leadership".[21][22] This purge resulted in the deletion of various materials, including interviews with Black and female NASA employees, LGBTQ-related content, and two NASA-created comic books about women astronauts.[23][24]

NOAA webpage replaced by "THIS FILE IS DELETED BY EXECUTIVE ORDER."

More than 3,000 pages from the Census Bureau website were removed as of February 2, primarily including articles filed under research and methodology.[3] Pages relating to data stewardship as well as survey and data set documentation were also removed.[3]

The Food and Drug Administration (FDA) removed more than 100 pages as of February 2, including dozens of regulatory guidelines on topics such as increasing diversity in clinical trials and the potential for addiction and abuse in drug studies.[3]

Close to 50 research papers from the Office of Scientific and Technical Information – part of the Department of Energy – were removed as of February 2. The removed papers covered a range of subjects, such as chemistry, optics, and experimental medicine.[3]

Twenty pages from the National Institute of Standards and Technology (NIST) website were removed as of February 2, including a page documenting the organization's zero-tolerance harassment policy.[3]

The environmental justice mapping and screening tool, EJScreen, was removed from the Environment Protection Agency (EPA) on February 5, along several related pages.[25] Public Environmental Data Project (PEDP) published a reconstruction of one of its earlier versions.[26]

In March 2025, an unknown executive order signed by President Donald Trump resulted in the NOAA *Radar Next Program Overview* document being removed from NOAA servers.[27]

The NOAA maintains a list of resources and products it retires.[28] On May 31, the entire climate.gov team was fired, likely shutting down the site.[29] The National Climate Assessment reports, congressionally mandated under the Global Change Research Act of 1990, were taken offline,[30] and the 400 scientists working on the 2027 assessment were fired.[31]

https://en.wikipedia.org/wiki/2025_United_States_government_online_resource_removals

But is this something new?

Repositories have always worried about archiving content, providing backups..

- The Open Archival Information System (OAIS) reference model provides recommendations on setting up archives delivering long-term preservation of and access to information (in particular, digital information) and creating preservation packages
- But how many worried about total destruction of their physical repository structure?

Offsite backups are part of any properly run repository

- In the old Bureau of Mineral Resources in Canberra in the 1970's, we had managed backups in Sydney but..
- Recent events show it needs to be International.

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2. *Consultative Committee for Space Data Systems*, https://public.ccsds.org/Pubs/650x0m2.pdf (2012)

Crisis Scenarios discussed (Categories, Examples Non Exhaustive)

1. Loss of **Materiel**
    a. Natural disasters e.g., hurricanes (Helene, 2024), wildfires, flooding
    b. Conflict damage e.g., war
2. Loss of **Funding**
    a. Pulling of contracts
    b. Payment processors stop processing transactions?
3. Loss of **Functionality**
    a. No access to facilities, e.g. cyber attack on servers
    b. Physical occupation of facilities
4. Loss of **Personnel**
    a. Retirements
    b. Injury or deaths
    c. Poaching by other organizations

Increasing the Resilience of Global Earth and Environmental Science Data Supply Chains

Chair: Lesley Wyborn
Moderator: Reyna Broadhurst
Speakers/Panelists: Joseph Gum, Lesley Wyborn, Tim Rawling, Adrian Burton

Detailed Session Description is here

INTERNATIONAL DATA WEEK
13-16 October 2025
Brisbane, Australia

Topics discussed:

1. Measuring Resilience of Individual Repositories (Joseph Gum, ESIP Sustainable Data Management Cluster)
2. Democratically Identifying Impactful E&E Science Datasets (Joseph Gum)
3. Global networking of Earth Science Research Infrastructures for support in times of crisis (Tim Rawling, AuScope)
4. Developing policies for sustaining critical E&E data supply chains during times of crisis (Adrian Burton, ARDC)

INTERNATIONAL DATA WEEK
13-16 October 2025
Brisbane, Australia

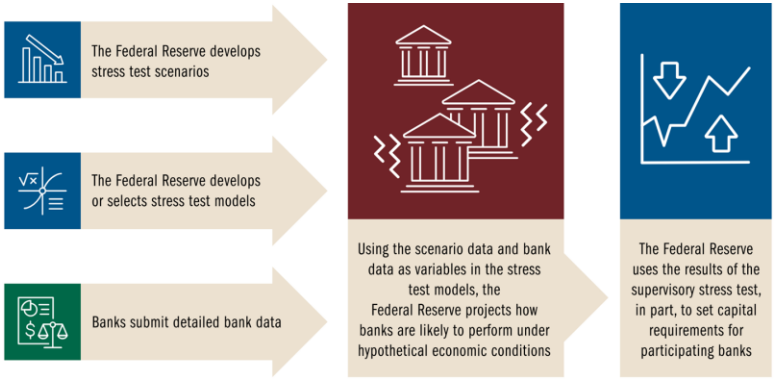
Repository Crisis Scorecards: An Assessment of Organizational Resiliency to the Unexpected

Joseph Gum

ORCID: 0000-0002-3045-0593

History of the Repository Crisis Scorecards (RCS)

- Created by the Earth Science Information Partners (ESIP) Sustainable Data Management Cluster, Joseph is the current chair (and unemployed)
- We started **February 2025** and [version 1.0 published](#) **April 3, 2025**
- Inspiration from the US Federal Reserve Banking Stress Tests - if a bank is stressed, can operations continue as usual, what remedies are needed?
- Conceived and published before knowing of [NDSA Levels of Digital Preservation](#) (which has since been remedied!)
- Format of the scorecards is borrowed from the NSF EarthCube Model Data RCN



2025 Stress Test Scenarios. Board of Governors of the Federal Reserve System. (n.d.). https://www.federalreserve.gov/publications/2025-stress-test-scenarios.htm

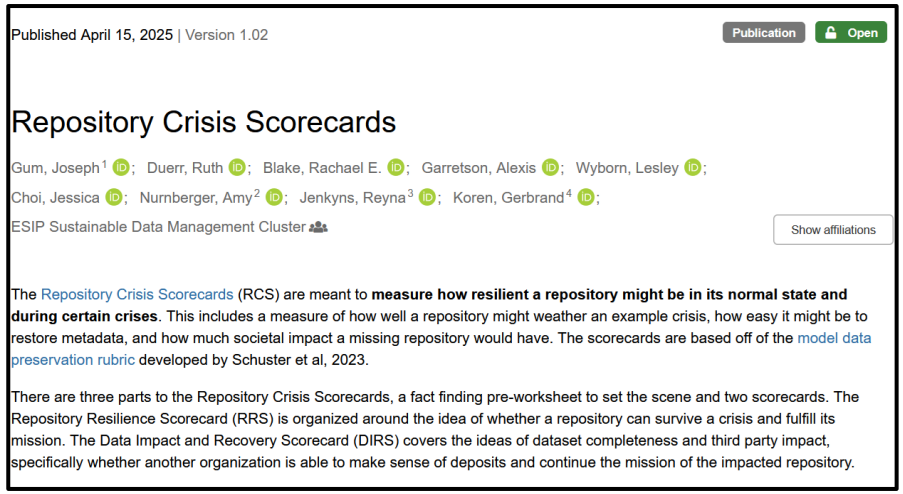
Practicals of the Repository Crisis Scorecards

Pre-worksheet: Questions to inventory holdings

- How much data is stored at the repository is the primary reference or only copy of the data? prepare for scorecards

Two scorecards:

- Repository Resilience Scorecard (RRS)
  - “original” scorecard with focus on how crises affect normal operations
- Data Impact and Recovery Scorecard (DIRS)
  - DIRS is made from crisis invariant questions that we didn’t want to throw out - if it feels disjointed/afterthought, it’s true!



Published April 15, 2025 | Version 1.02

Publication Open

Repository Crisis Scorecards

Gum, Joseph[1]; Duerr, Ruth; Blake, Rachael E.; Garretson, Alexis; Wyborn, Lesley; Choi, Jessica; Nurnberger, Amy[2]; Jenkyns, Reyna[3]; Koren, Gerbrand[4];

ESIP Sustainable Data Management Cluster

Show affiliations

The Repository Crisis Scorecards (RCS) are meant to **measure how resilient a repository might be in its normal state and during certain crises**. This includes a measure of how well a repository might weather an example crisis, how easy it might be to restore metadata, and how much societal impact a missing repository would have. The scorecards are based off of the model data preservation rubric developed by Schuster et al, 2023.

There are three parts to the Repository Crisis Scorecards, a fact finding pre-worksheet to set the scene and two scorecards. The Repository Resilience Scorecard (RRS) is organized around the idea of whether a repository can survive a crisis and fulfill its mission. The Data Impact and Recovery Scorecard (DIRS) covers the ideas of dataset completeness and third party impact, specifically whether another organization is able to make sense of deposits and continue the mission of the impacted repository.

Gum, J., Duerr, R., Blake, R. E., Garretson, A., Wyborn, L., Choi, J., Nurnberger, A., Jenkyns, R., Koren, G., & ESIP Sustainable Data Management Cluster. (2025). Repository Crisis Scorecards (1.02). Zenodo. https://doi.org/10.5281/zenodo.15208172

Current Status/Feedback

Available on Zenodo since 3 April 2025:

- as at 8.00 AM 23rd October 2025: **1891 Views, 1868 downloads**
- All submitted scores/responses: **5**
- Currently submitted scores/responses after publication: 0

Feedback:

- Useful as an **self assessment tool** and conversation starter
- Worried about public **scores used to justify cutting grants/repositories**
- Need **assurance on anonymization** of responses
    - Working on Data Privacy Policy
    - Adopting best practices to anonymize responses

Future Work

Sloan Award (Oct 5 2025) to continue including:

- Finish **Repository Crisis Scorecard (RCS) description paper**
- Publish an **updated version** of the RCS
- **Collect data from four cohorts** (interest form)
  - repository administrators
  - scientific and technical domain experts
  - data users
  - citizen archivists
- Publish a **white paper** summarizing findings from the RCS and provide recommendations for strengthening repository resilience

Engage with Us

- **Learn more, submit scores, and interest form**
  - esipfed.org/rcs
- **Join a Sustainable Data Management Cluster meeting**
  - first Monday of the month at 4 pm US ET (8:00 AM AEDT)
- **Questions?** rcs@esipfed.org

Online RCS form

RCS Website

Thank You!

ESIP EARTH SCIENCE INFORMATION PARTNERS

ALFRED P. SLOAN FOUNDATION

Impactful Datasets

23 October 2025

Joseph Gum, Independent

Shelley Stall, VP, Open Science Leadership

AGU OPEN SCIENCE

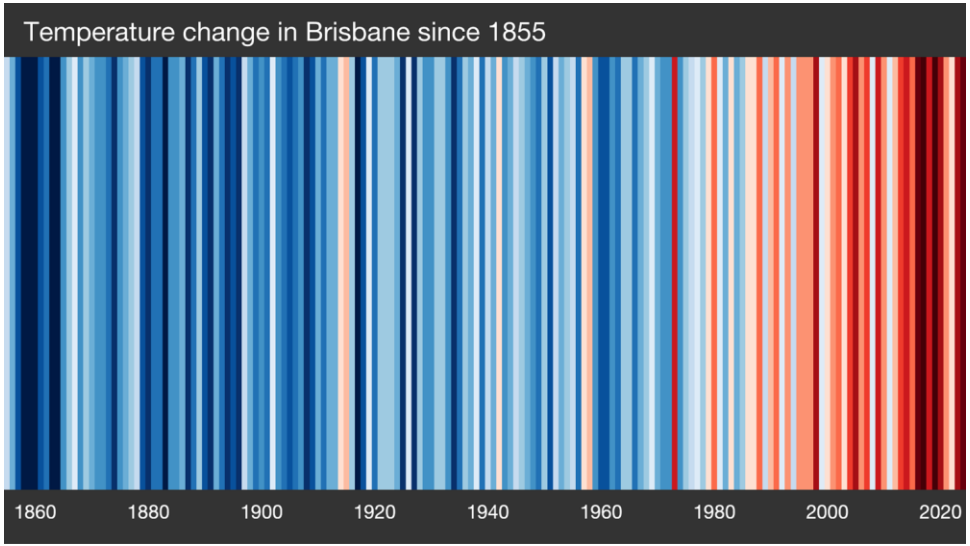
Impactful Datasets

Idea formed February 2025 – what data might be at risk of being lost?

Flagged words in the US include “climate change, clean energy”

Leading voices will speak up for leading datasets

How to give the rest of the community a voice in nominating datasets that mean something to them?

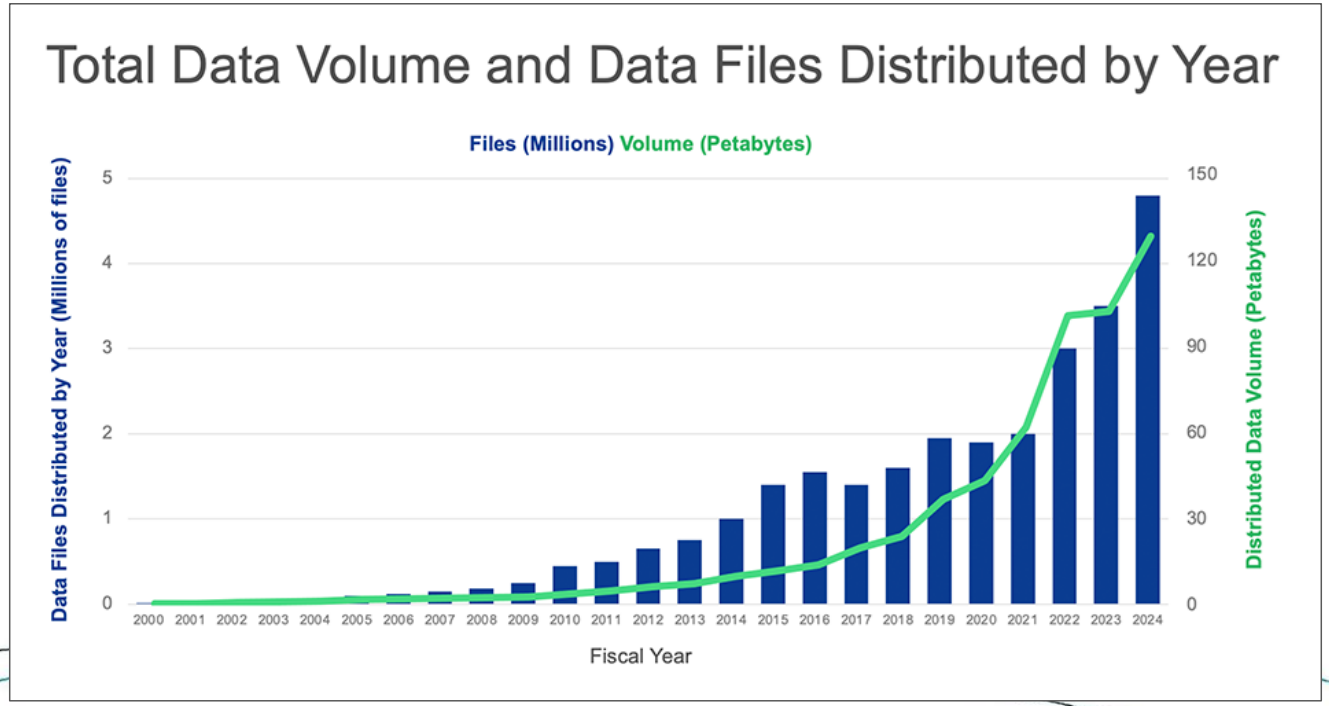


https://showyourstripes.info/l/australasia/australia/brisbane

Data Rescue is not Data Hoarding

NASA Earth Science Division Statistics:

- 4.5 billion files delivered
- 8+ million distinct data and service users
- 128+ petabytes of Earth science data



https://www.earthdata.nasa.gov/

Most Impactful Datasets

Democratic process for the community to nominate impactful datasets…

…what are the criteria for “impactful datasets”?

Members of Earth Science Information Partners (ESIP) were consulted over a month to develop the criteria

Lots of debate when trying to determine a criteria to cover all Earth and Space Sciences datasets: in-situ, remote sensing, models…

Selection Criteria

A dataset IS:

- Publicly Accessible - Open Science
- FAIR - Findable, Accessible, Interoperable, Reusable
- Legally usable - Licensing

A dataset IS NOT:

- Dark data, e.g. private lab notebooks, hard drives, logs, emails
- An instrument, e.g. microwave sounders, thermistors
- A technique, e.g. spectral analysis

Submitter states why the dataset is **impactful** to them!

Promote Nominations and Follow Up

Cutoff for nominations is now 20 October 2025

Commentary to be released in AGU Advances November 2025

Special Collection – open in Earth and Space Sciences late 2025

Presentation at AGU 2025 in New Orleans

AGU Impactful Datasets Nomination Form

Global networking of Earth Science Research Infrastructures for supporting times of crisis

Presented by:
Dr Tim Rawling
ORCID: 0000-0002-8841-4384

The role of Earth & Environment National Research Infrastructures

- National Research Infrastructures (NRI's) **create, curate and protect** a variety of Earth, environmental and climate datasets for the benefit of the research community and society.
- Our mandate is different to Government Agencies.
  - We can operate as **trusted data providers**
- Always designed to provide **data resilience in times of crisis**
  - Many focussed on geohazard monitoring
    - Earthquake, tsunami, cyclone, flood, bushfire
    - Often networks can be damaged or destroyed
    - Data systems designed to be **mirrored in case of catastrophe**
    - Architected to allow **high bandwidth in times of crisis**

International Partners

USA

EarthScope Consortium

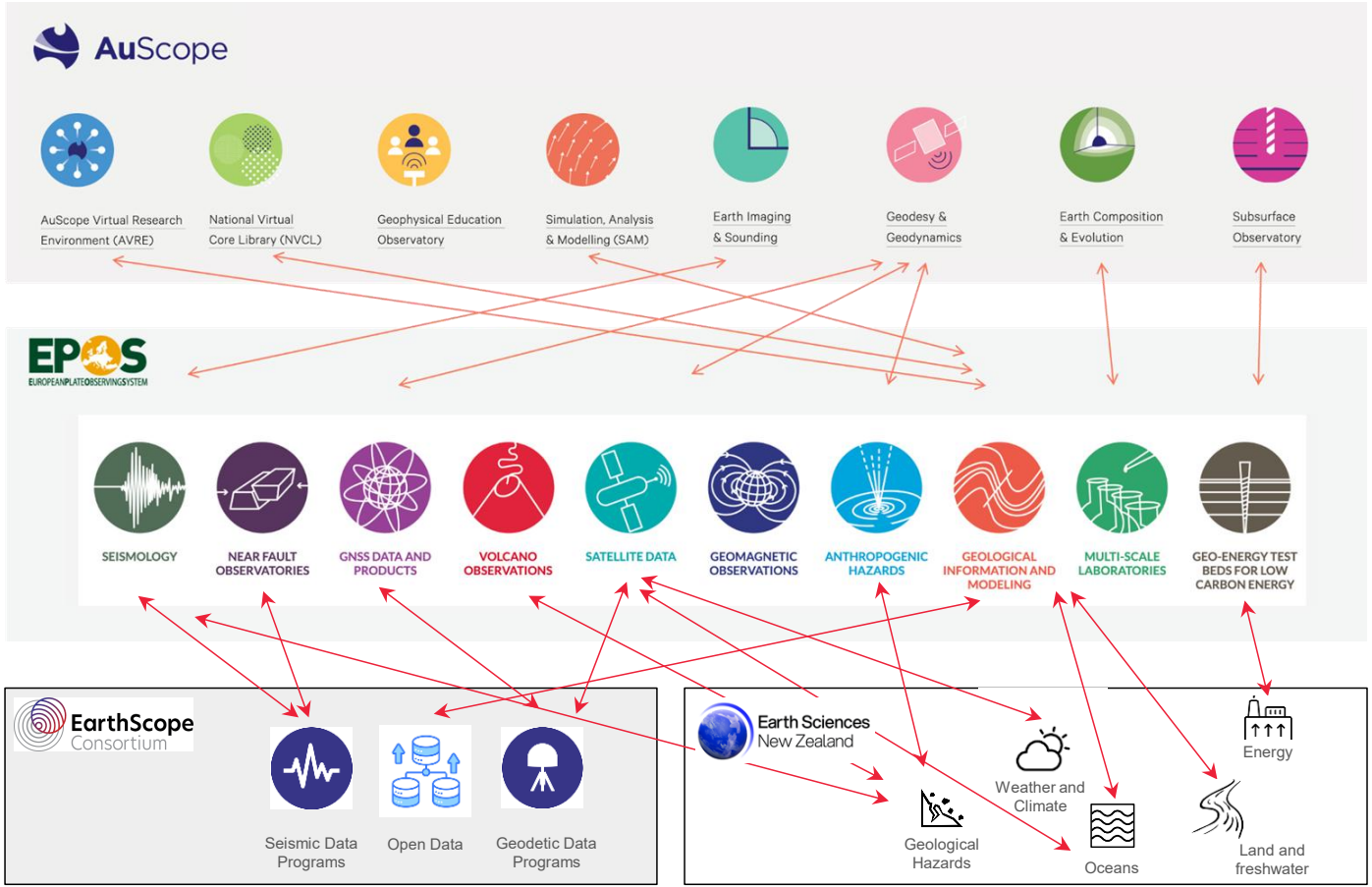
Australia & New Zealand

AuScope

Earth Sciences New Zealand

Europe

EPOS
EUROPEANPLATEOBSERVINGSYSTEM


AuScope
AuScope Virtual Research Environment (AVRE)
National Virtual Core Library (NVCL)
Geophysical Education Observatory
Simulation, Analysis & Modelling (SAM)
Earth Imaging & Sounding
Geodesy & Geodynamics
Earth Composition & Evolution
Subsurface Observatory
EPOS
EUROPEANPLATEOBSERVINGSYSTEM
SEISMOLOGY
NEAR FAULT OBSERVATORIES
GNSS DATA AND PRODUCTS
VOLCANO OBSERVATIONS
SATELLITE DATA
GEOMAGNETIC OBSERVATIONS
ANTHROPOGENIC HAZARDS
GEOLOGICAL INFORMATION AND MODELING
MULTI-SCALE LABORATORIES
GEO-ENERGY TEST BEDS FOR LOW CARBON ENERGY
EarthScope Consortium
Seismic Data Programs
Open Data
Geodetic Data Programs
Earth Sciences New Zealand
Geological Hazards
Weather and Climate
Oceans
Energy
Land and freshwater

Shared Vision for a GRI

- ✓Uphold the principle of **open science**: open, accessible, interoperable data
- ✓**Boost Visibility and Use** of Data in Global Research Ecosystems
- ✓**Elevate local data** to global levels, while **involving local communities**, ensuring they benefit from relevant scientific outcomes
- ✓**Secure** locally, nationally and globally significant datasets

Global Research Infrastructures create opportunities for international data resilience

- **Petabytes of data are available across borders, disciplines, and timescales** but have limited value in isolation.
- By creating **open, integrated, multidisciplinar**y data collections we enable wholesale data sharing and interoperable resilient data ecosystems.
- **Research Infrastructures (RIs)** across the continents provide:
  - High-quality, standardized, open data
  - Equitable access and collaboration opportunities
  - Foundations for multidisciplinary science
  - Resilient data networks

Academies of Science:
Policies for Research Supply Chain Resilience

- What are the cross border data flow dependencies for contemporary research?
- How could those data supply chains be disrupted?
- What measures can we take?

Crisis Scenarios: do they affect your data and/or the repository that you are using??

1. Loss of **Materiel**
    a. Natural disasters e.g., hurricanes (Helene, 2024), wildfires, flooding
    b. Conflict damage e.g., war
2. Loss of **Funding**
    a. Pulling of contracts
    b. Payment processors stop processing transactions?
3. Loss of **Functionality**
    a. No access to facilities, e.g. cyber attack on servers
    b. Physical occupation of facilities
4. Loss of **Personnel**
    a. Retirements
    b. Injury or deaths
    c. Poaching by other organizations