

## Preserving open data as a world heritage to secure our future

*Achieve accessibility, transparency and reproducibility in research.*

Addressing global challenges and taking action towards a sustainable future—in regards to ecology, biodiversity, water and food security and climate—requires both solution-driven research and a deep understanding of Earth and space systems. Trustworthy, reproducible science requires that data, including their provenance, be accessible for evaluation and further study. International collaboration and open data sharing are essential for addressing these challenges and promoting new scientific advances. Although strides have been made in improving data sharing platforms, as well as data and metadata standards, there is still work to be done to facilitate careful collection, stewardship and reuse of data across the Earth and space sciences.

To achieve an ecosystem where data are consistently collected, well described, curated, shared, preserved and reused there needs to be significant continuous evolution in (i) equitable access to data and infrastructure, (ii) a change in the research culture and the scientific rewarding system to value data contributions, and (iii) standards around describing and documenting data across disciplines and domains.

This statement takes the broadest interpretation of the term ‘data’ according to the [Beijing Declaration on Research Data](#): data can be collected, generated or compiled by humans or machines and include (but not limited to) metadata, samples, methods, software and algorithms. Infrastructure includes hardware, intangible assets such as software and human capital.

### **Ensure equitable access to trusted and reusable data**

Researchers have a responsibility to collect, document and share data—raw and processed—in an ethical manner that is as open and transparent as possible.

Persistent funding is crucial for maintaining open and trusted data infrastructures, such as repositories, to facilitate ease of deposit and long-term reuse. This infrastructure supports the management and curation of data throughout the research process, ensuring that data are findable, accessible, interoperable and reusable ([FAIR](#)) for both people and machines. To meet these expectations, investment in support staff, such as data stewards and IT developers, along with research and development of data architectures is imperative.

In alignment with the [UNESCO Recommendation on Open Science](#), this statement takes the position that data should be ‘as open as possible, as closed as necessary,’ which recognizes the challenges associated with data related to national security, intellectual property and cultural sensitivities. Future data ecosystems should implement federated infrastructure to ensure equitable access without compromising national security, privacy and other concerns.

It is highly important to acknowledge the context and provenance of all data, including derived data, with respect to nature and people, according to the [CARE Principles for Indigenous Data Governance](#). Additionally, we need to recognize and address the social and technical barriers that hinder marginalized groups from participating in the current process through co-development.

Equitable access to trusted data is critical for facilitating transparent, reproducible and trustworthy artificial intelligence (AI) systems. We should enhance data documentation and curation to mitigate the risk and bias in AI systems in alignment with the AGU community [guidelines on ethical and responsible use of AI](#).

### **Change research culture to recognize data contributions**

The behavior of researchers is strongly influenced by the culture of institutions, publishers and funders. It is critical that these entities recognize and reward contributions by people serving different roles across the life cycle of data, in line with the [San Francisco Declaration on Research Assessment \(DORA\)](#).

Data management planning is a crucial first step in this process. Robust data stewardship necessitates cooperation between individual researchers, professional support staff, IT developers, disciplinary communities and repository personnel. Thus, dedicated data managers and curators are integral to building synergies among these stakeholders and should be appropriately funded and recognized for their contributions. In addition, researcher training should incorporate a curriculum on best practices for data stewardship, which should consider disciplinary norms. Finally, institutions and researchers must openly identify and address academic incentive structures that disincentivize data sharing.

### **Implement metadata standards across disciplines and domains**

In our interdisciplinary and diverse research community, data allows researchers from different domains to communicate effectively about specific measurements and analysis. However, which salient features of data are captured, and how they are described, often differs by domain, creating challenges to collaboration and data reuse. For this reason, domain-specific repositories are recommended over generic repositories for better disciplinary data curation.

Additionally, describing and accounting for the specific provenance and contexts—such as permissions, terminologies, uncertainty and sampling biases—is essential for the responsible reuse of data in alignment with FAIR and CARE principles, including reuse via automated data-driven analysis and AI. Thus, it is paramount that our community continuously improve data documentation and adopt standards for transparent, machine-readable and understandable metadata, and data properties.

### **Take collective responsibility**

Robust data stewardship, preservation and sharing requires individual actions by all players in the global research ecosystem: researchers, institutions, funders, publishers and research infrastructure stewards. Data infrastructure supporting the complete data life cycle should be globally distributed across public, private, commercial and not-for-profit entities. Expanding open data infrastructure, such as repositories, across national boundaries will advance scientific progress and address global challenges.

All entities have a responsibility to ensure their infrastructure facilitates equitable and sustained ease of use for both data depositors and users, is interoperable and prioritizes collective reuse of data. To take action:

- **Researchers** should engage with research infrastructure and repositories to ensure their data comply with FAIR and CARE principles and that data are responsibly developed, used and cited.
- **Research infrastructure providers**, especially repositories, can be key partners for researchers in making their data FAIR and CARE-compliant. They should ensure that metadata capture is robust and standardized, including adherence to disciplinary standards and the use of persistent identifiers.
- **Publishers and Journals** should require data, as defined broadly in the beginning of this statement, be archived in federated, recognized and domain-specific repositories. The data should be properly cited and made available upon publication while mitigating concerns of privacy, national security and Indigenous Data Sovereignty.
- **Institutions** should recognize, reward and incentivize the work associated with responsible data stewardship and support persistent funding of research infrastructure.
- **Funders** should require the inclusion of responsible, FAIR and CARE-compliant data sharing in all funding agreements, with consequences for non-compliance and direct persistent funding to research infrastructure commensurate with the central roles it plays in achieving the vision outlined in this statement.

Statement adopted by the American Geophysical Union 29 May 1997; Reaffirmed May 2001, May 2005, May 2006; Revised and Reaffirmed May 2009, February 2012, September 2015; November 2019; September 2024.