

Climate Data Resilience Under Conditions of Institutional Failure

“Stay Ready so you don’t have to Get Ready”

March 2026

Stephen C. Diggs, Scripps Institution of Oceanography / UC San Diego,
[0000-0003-3814-6104](https://www.scripps.edu/people/diggs/)

Mark A. Parsons, National Snow and Ice Data Center, Univ. of Colorado, Boulder
[0000-0002-7723-0950](https://www.nsidc.org/people/parsons/)

1. Climate Data as Critical Infrastructure

Climate and environmental data underpin modern society. They inform decisions about water quality, air quality, food systems, navigation, public health, and disaster response. These data are not peripheral research outputs. They are critical infrastructure that are sustained by professionals.

Like all infrastructure, climate data systems face threats from natural hazards, technical failure, workforce instability, funding volatility, and political disruption. Resilience, therefore, must be understood operationally as the capacity to absorb disruption and recover function quickly.

Other infrastructure domains have confronted similar challenges. Fire management is particularly instructive, not because of its imagery, but because it has evolved toward systematic prevention, preparedness, professional response, and adaptation. Climate data stewardship must undergo a comparable transition.

2. A Practical Definition of Resilience

In this framework, resilience consists of four distinct but interdependent, human-driven elements:

1. **Prevention**
2. **Preparedness through risk assessment and prioritization**
3. **Crisis response**
4. **Adaptation over time**

Each element serves a different purpose, and confusing them undermines the whole. Focusing on only one, however effectively, is not just insufficient, it can be actively damaging. It can create a false sense of security while the underlying weak points that drive crises in the first place go unaddressed.

South African Archbishop Desmond Tutu made this point plainly: “There comes a point where we need to stop just pulling people out of the river. We need to go upstream and find out why they're falling in.” Emergency data rescue is necessary work, and the people doing it deserve full support. However, a community organized *primarily* around rescue will always be reacting to crises it could have reduced.

The river keeps drowning people.

Fire management offers a more complete model. Modern wildfire resilience doesn't choose between prevention and response. It integrates fuel management, community preparedness, professional standing response capacity, and post-fire ecosystem adaptation into a single operational doctrine. Each element depends on the others; remove any one, and the system becomes more fragile.

Climate data stewardship requires the same integrated thinking. The following sections address each element in turn, but the core argument is holistic: resilience is the product of all four elements working together, not just whichever one receives the most attention during the current crisis.

3. Prevention: Reducing Fragility Before Crisis

Fire prevention relies on norms, regulation, and engineering. Public education discourages risky behavior. Building codes reduce ignition risk. Controlled burns reduce fuel accumulation.

Climate data prevention requires a similarly layered approach.

Culturally, data must be recognized as a shared public asset whose value depends on accessibility, documentation, and stewardship. Norm-building around data sharing and responsibility remains incomplete but essential.

Practically, several actions are immediately available:

- Good data practices should be defined and [established at project inception](#), not deferred.
 - Data Producers can (and should) be data stewards / custodians too.
 - Stewards can be collectors.
 - Define these roles in advance.
- Researchers and data professionals must collaborate early to ensure data are intelligible and usable (e.g., [Borycz 2021](#), [Parsons et al. 2011](#)).
- Repositories should demonstrate resilience through certification (e.g., [Core Trust Seal](#)), governance transparency, and stress testing (e.g., [Repository Crisis Scorecards](#)).

- Repository communities must actively reduce unnecessary complexity, opaque provenance, and uncoordinated redundancy e.g. by adopting explicit provenance standards ([Greis et al. 2018](#)).
 - Coordinate collection and instrument configuration and synchronization.
 - Document openly and collaboratively.
 - Be clear on licenses and rights.

Prevention is rarely visible, but it is the most cost-effective form of resilience.

4. Preparedness: Prioritization is Not Enough

Growing investment in identifying “priority datasets” for rescue or reinforcement is important. Determining which datasets have the highest scientific, societal, and operational value is legitimate, important work. Some datasets genuinely are more critical than others, and value assessment shouldn't be abandoned.

This framework doesn't dispute any of that.

But the limitations of prioritization as a primary strategy need to be stated plainly, because the literature and community practice haven't always done so.

What this framework does not argue:

- That priority lists are useless
- That some datasets are more critical than others
- That value assessment should be abandoned.

What this framework does argue:

- Priority lists proliferate rapidly when combined across communities.
- Priority lists often obscure labor, custody, rights, and political risk.
- Priority lists can create a false sense of control under highly volatile conditions.

Case Study: Data Rescue in Action

As described by de Sherbinin ([2026](#)), the 2025 rescue of the NASA Socioeconomic Data and Applications Center (SEDAC) illustrates our principles in practice.

When NASA funding was abruptly terminated and SEDAC datasets were removed from public catalogs, more than six terabytes of globally significant data became inaccessible with no advance notice. The threat was institutional rather than technical.

The response succeeded because it emphasized readiness over ranking. Trusted data professionals mobilized quickly, repository options were evaluated against operational criteria, and a generalist repository with migration support and institutional backing was selected. Redundancy was treated as a feature rather than a flaw.

SEDAC now exists on multiple platforms with independent identifiers. This outcome reduced fragility and restored access quickly, demonstrating that distributed stewardship outperforms assumed permanence.

The first problem is simple arithmetic. Any single organization can produce a coherent, well-reasoned priority list. But when lists from multiple, diverse communities - climate science, oceanography, public health, agriculture, ecology, economics - are laid side by side, their aggregate demands quickly exceed available institutional capacity. Aggregating rankings doesn't produce a unified strategy; it produces an unranked pile with a ranked aesthetic. It makes you feel like you've done something useful but only adds confusion.

The second problem runs deeper. Priority frameworks typically score scientific or informational value reasonably well. They are not as good at evaluating the operational realities underneath a dataset's continued existence: Who holds custody, and under what terms? What tacit expertise maintains the processing chain? What legal or jurisdictional constraints govern mirroring? A dataset can score highly on every value dimension while remaining effectively inaccessible for reasons no priority scoring approach captures. Lists that don't reflect these realities don't just underperform, they *actively mislead* by suggesting that scoring a dataset is equivalent to securing it.

The third problem is temporal. Priority lists are calibrated to conditions at the time of their construction. Under volatile conditions, the situation can change faster than any list gets updated. What was low-risk last week may be in acute danger today.

Dataset prioritization is a preparedness activity. It informs planning, investment strategies, staffing decisions, and mirroring strategies during non-crisis periods when deliberate action is possible. It doesn't function well as an emergency gate—a real-time mechanism for deciding what to save when capacity is already overwhelmed and the window of intervention is closing. That requires pre-positioned infrastructure, pre-negotiated custody agreements, and teams that already know what they're doing.

A spreadsheet of ranked datasets doesn't substitute for any of these critical preparedness activities..

The communities doing the most effective prioritization work understand this distinction. The communities most at risk are those treating the prioritization list as the destination rather than the foundation for harder operational and governance work that must follow.

An ensemble approach to climate data preservation integrates dataset utility, institutional vulnerability, and observational continuity. Like ensemble forecasting in weather prediction, it improves preparedness without claiming certainty. Its proper role is strategic, not operational.

5. Response: Standing Capacity, Not Ad Hoc Rescue

When a fire occurs aboard a ship at sea, there is no external assistance. The U.S. Navy addresses this risk through standing, trained damage-control teams. These teams respond immediately to any alarm, focus first on containment, and escalate when necessary. Every sailor has baseline training, but specialists lead the response.

Climate data emergencies share key characteristics with shipboard casualties. They unfold rapidly, often without warning, and are frequently driven by institutional rather than technical failure. Under these conditions, ad hoc volunteer efforts and prolonged deliberation increase risk.

Effective crisis response requires a **new, standing** capacity (a “fire brigade” with in-house and external specialists) that can:

- Assess risk rapidly ([Mayernik et al. 2020](#))
- Engage existing custodians
- Preserve full context, **not just files**
- Escalate across institutional boundaries when needed.
 - Many data crises hit walls that a single institution can't handle on their own. Legal obstacles, Indigenous data sovereignty, and similar challenges would need to rely on pre-established relationships with colleagues who hold authority at those boundaries: national centers, international bodies, legal counsel, agency leadership. Without pre-negotiated escalation pathways, a response team that reaches one of those walls simply stops.

Copying alone is not preservation. Without documentation, provenance, and stewardship continuity, copied data may be unusable.

Crisis response remains the least developed element of this framework and, arguably, the most consequential. Progress on prioritization and prevention is real. The standing capacity, governance structures, and pre-negotiated agreements that effective crisis response actually requires do not yet exist. The repository community needs to treat building that capacity as urgent, practical work, starting now.

6. Adaptation: Designing for Change

Fire ecology has shown that total suppression increases long-term risk. Healthy systems adapt.

Climate data infrastructure must do the same. Resilience does not imply preserving systems unchanged. It requires:

- Continuous maintenance and migration
- Periodic refactoring of services
- Consolidation where fragmentation increases risk
- Planned retirement of obsolete components.

Adaptation requires sustained maintenance, funding, and professional oversight. Neglect increases brittleness and amplifies future crises.

7. People First

Addressing these elements of prevention, preparedness, response, and adaptation requires that we consider people first. It is communities who suffer when data are lost, and data rescue is not a one-off. It is an on-going activity because the relevant people and their knowledge must be continually supported.

Across all elements of resilience, one ordering principle holds:

- 1. People**
- 2. Products**
- 3. Data**

Workforce expertise and trust networks are the most fragile components. Products make the data more useful and usable. The value of data persists only when the first two are supported.

Ignoring this ordering produces brittle systems that fail quietly and suddenly.

Data are only valuable if people do something with them. At the same time, data are a public good. Sustaining experts and their knowledge is an imperative and it takes some risk.

None of this happens without professional courage. Standing up for data resilience means standing on something specific: the shared values this community already holds around open access, FAIR principles, and the conviction that data friction is a solvable problem with real human consequences. Most of us already agree on these things. What the current crisis demands is the willingness to act on them visibly, on the record, even when institutional incentives push back. That is not a technical challenge. It is a professional and ethical challenge.

We know what needs to be done.

8. Conclusion

The assumption that climate data will be stewarded indefinitely by any single institution is no longer tenable. Resilience now depends on distributed responsibility, professional readiness, and a clear recognition of the need for both preparedness and response.

Priority frameworks remain essential planning tools. They are not substitutes for readiness.

Preserving climate data under conditions of institutional instability requires holistic investment in professionals, standing response capacity, and systems designed for redundancy and change. This shift is not optional. It is the cost of operating responsibly in an increasingly volatile environment.

9. References

- Borycz, Joshua. 2018. "Implementing Data Management Workflows in Research Groups Through Integrated Library Consultancy." *Data Science Journal* 20: 9. <https://doi.org/10.5334/dsj-2021-009>.
- De Sherbinin, Alex. 2026. "Things Fall Apart: Lessons from a Defunded Data Repository." *Data Science Journal* 25: 9. <https://doi.org/10.5334/dsj-2026-009>.
- Gries, Corinna, Amber Budden, Christine Laney, et al. 2018. "Facilitating and Improving Environmental Research Data Repository Interoperability." *Data Science Journal* 17: 22. <https://doi.org/10.5334/dsj-2018-022>.
- Mayernik, Matthew S., Kelsey Breseman, Robert R. Downs, Ruth Duerr, Alexis Garretson, and Chung-Yi (Sophie) Hou. 2020. "Risk Assessment for Scientific Data." *Data Science Journal* 19: 10. <https://doi.org/10.5334/dsj-2020-010>.
- Parsons, Mark A., Øystein Godøy, Ellsworth LeDrew, et al. 2011. "A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science." *Journal of Information Science* 37, no. 6: 555–69. <https://doi.org/10.1177/0165551511412705>.
-