CRRDI Draft Strategic Plan: Community Comment

Deadline: 30 September 2026
Feedback Form

Strategic Plan: https://zenodo.org/records/21497401

Repository Resilience Maturity Model: https://zenodo.org/records/21608822

Webinar recording:
youtube.com/watch?si=2qMTE0rUfKuci6Eo&t=5&v=LDl9L3HQQKo&feature=youtu.be

Webinar slides: https://zenodo.org/records/21842235

The Coalition for Resilient Research Data Infrastructure (CRRDI) has released a draft strategic plan for community comment. To ensure this plan is truly a community-driven strategy, we are releasing this draft at an early stage for community comment to understand whether the strategies serve the ecosystem's needs, are feasible given current realities, and resources and existing work that should be referenced. CRRDI is a coalition still in formation, and this plan is intended to serve as an anchor around which the community can organize rather than a finished blueprint handed down by the Strategic Planning Committee.

We want to know whether the strategies serve the ecosystem's needs, whether they're feasible given current realities, and what existing work we should be building on. We're interested in feedback on the coalition model itself and its governance (as outlined in Appendix A), as well as gauging interest in joining CRRDI's efforts to implement the strategic plan.

*required

Name: Meredith Goins

Organization: World Data System

What's your relationship to this work? *(select all that apply)*

- ☐ Repository or infrastructure provider
- ☐ Data librarian or curator
- ☐ Researcher or data user
- ☐ Funder

- ☐ Policy or government
- ☐ Industry
- ☐ Advocacy or civil society
- ☐ Other: international research data infrastructure network

May we contact you about your comments?*

- ○ Yes
- ○ No

[If yes is selected, next question follows. If no is selected, this question does not appear]

Please provide your email, so we're able to get in touch with you.*

Wds-ipo@utk.edu

Do the strategies serve the ecosystem's needs as you understand them, with respect to supporting long-term repository resilience? What's missing?

- Please share how CRDDI is different than the other efforts to shore up infrastructure over the last two decades. How is this going to move the needle?
- Consider a discussion on the business acumen and fundraising and development experience of data repository leaders. Development officers at universities play a role in sustaining colleges and programs, why not repositories? Is there a synonymous entity at data repositories?
- As indicated on the website "While CRRDI is focused on federally funded research data in the U.S. context, many of the strategies outlined in the plan should be considered within a global framework." Given the broad relevance (beyond federally funded research and the US) and the magnitude of effort for pillar success, it may be better to frame this focussed concern as a priority objective within a fuller-scope initiative which can build upon this foundation for a myriad of other repository resiliency issues. While federal funding is a current crisis, repositories face other funding barriers and other resiliency risks that factor into the maturity model and the ultimate sustainability of these repositories.

Are these strategies feasible given current realities? Where do you see obstacles?

- The strategic plan is for a 3-year period, from 2027-2030. It is unclear what would be the anticipated outcome at that point.

- Given the emphasis on federal funding, what tactics will increase the likelihood that outputs can withstand evolving political circumstances within and beyond the strategic plan timeline?
- Additional feedback regarding obstacles and feasibility are contained in pillar-specific responses.

What existing work, resources, or initiatives should this plan be referencing or building on?
CRRDI is committed to amplifying and building on existing efforts, incorporating them into our implementation roadmap, rather than initiating competing or duplicative efforts.

- We give numerous specific resources to consult in the pillar-specific questions. We are aware of numerous resiliency-focussed activities, including repository network replication services and disciplinary-specific coordination. Many of these are already known to the planning committee members (e.g., an AGU-lead resiliency coalition for earth sciences).

Pillars — Specific Comments

Pillar 1 — Assess and Monitor the Repository Landscape

Understanding where the ecosystem is most vulnerable and how those risks evolve is essential to advancing resilience. This pillar develops a shared, community-led maturity model to help those supporting federal data infrastructure adopt resilient risk management strategies, building on existing frameworks rather than duplicating them.

(1.1) Model:

- Do we really need another repository maturity model and all the pieces that may have their own models within? Why is this better than others? A few more recent examples, although this has been in discussion for multiple decades:
    - Data repository lifecycle model discussed in: Lin, D., McAuliffe, M., Pruitt, K.D. *et al.* Biomedical Data Repository Concepts and Management Principles. *Sci Data* **11**, 622 (2024). https://doi.org/10.1038/s41597-024-03449-z
    - Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K. and Stall, S. (2020) 'The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments', *Data Science Journal*, 19(1), p. 41. https://doi.org/10.5334/dsj-2020-041
    - Corcho, O., Ekaputra, F.J., Heibi, I. *et al.* A maturity model for catalogues of semantic artefacts. *Sci Data* **11**, 479 (2024). https://doi.org/10.1038/s41597-024-03185-4
    - Francis Chemorion, Marc-Antonio Bisotti. Data curation maturity model. *TechRxiv*. 14 November 2024. https://doi.org/10.36227/techrxiv.172963242.27715654/v2
- There is a significant leap between emerging and established. Perhaps there needs to be a middle level of "Operationalizing."

- It may be very difficult for a repository to assign themselves within a level for some of the dimensions (because some of the dimensions are attempting to flatten multiple dimensions within); for example, data stewardship and integrity encompasses a lot of services that merit independent maturity evaluations (e.g., data deposit curation versus long-term active preservation). A similar challenge is faced by the FAIR dimension since many repositories do well at the FA but would score less well for the IR.
- The "Community & Use" dimension doesn't seem to adequately capture the various types of use, nor acknowledge the immense difficulty for open data repositories to comprehensively and accurately know how data are being used. The idea that formal agreements are expected with its users is not widely realistic, nor sensible for most circumstances. It is also unclear to what degree those users need to be directly involved in the resiliency planning and response. Wouldn't it make more useful to have these maturity levels reflect the extent to which the repository understands its community and the impact of the data usage?
- It is somewhat awkward to have a dimension exclusively for 'Crisis Readiness and Resilience' given that is the intent of the overall model, unless this would be some kind of aggregate result based on the other dimensions?
- Although this matrix may provide some additional utility, it is somewhat redundant with the CoreTrustSeal application and may be burdensome to complete if supporting documentation is expected. Perhaps an initial estimate maturity level assignment could be executed by AI with existing content, such as CoreTrustSeal applications and re3data records (whilst promoting updates to these records for adequate content).

(1.2) Landscape analysis:

- This proposed strategic plan states includes a landscape analysis that proposes to itemize all federal data repositories that have a primary purpose of hosting federally funded research datasets. How will CRRDI manage inputs from the institutional repository and generalist repositories? It is known that not all data ends up in these federal repositories. How will CRRDI include the data listed in generalist and IR repos?

- In the strategic plan, the title and summary of this subsection emphasizes FAIR, but this is only one dimension of the maturity matrix. Is that intentional?
- How will this initiative build upon a multitude of existing approaches to FAIR assessments, and minimize duplication of effort (e.g., FAIR Maturity Matrix, F-UJI, FAIR Implementation Profiles, etc).
- What criteria are being used to determine whether data repositories have a primary purpose for federally-funded data? Will this be self-declared? Will metadata about funder information be considered (despite that existing funder information metadata is still not very mature)? What about repositories that have done the good work to be more diversified, yet still include federally funded data within their data holdings?

- Who owns/governs the maturity model data that will be collected for monitoring, and what rights do the repositories have with data about them?
- A data management plan will be needed for the maturity model dataset(s), so that it can itself be a citeable resource that is preserved, with versions or releases to account for corrections and updates over time.

(1.3) Resiliency dashboard:

- As acknowledged, the level of transparency for repositories to disclose the reality of their situation needs to be treated carefully to avoid inadvertent harm.
- Consider creating the governance and maintenance plan before building the tool. Not having these items does not create buy-in from the beginning and could lead the dashboard to be unowned, requiring more free labor to maintain the tool.

Pillar 2 — Develop Diverse, Sustained Funding and Governance

Even the best assessment is only useful if repositories have the funding, governance, and community to act on it. This pillar develops diverse, sustainable business models and establishes a multi-sectoral consortium where funders, repositories, and other stakeholders can align investment, share best practices, and work through the sustainability challenges no organization can solve alone.

- Many U.S. federal funders are willing to offer start-up costs, but then state that repositories need to build their own funding model that doesn't rely on future federal funds, even if the research is performed under federally funded research grants. What proof is there that the funders want to play a "core role in preserving and maintaining repository infrastructures?"
- How is the project incorporating the changing governmental requirements? It cannot be assumed that the federal government believes in the “public good” for data. Is it in scope for CRRDI to advocate for the "public good" for data?
- This conversation has been active for decades. How is your path forward different from the business models already discussed in the literature? See, but not limited to: (Baker & Millerand, 2024; Baker, 2012; Berman & Cerf, 2013; Bothner et al., 2025; Burns et al., 2013; Dillo & Hodson, 2015; Johnson, 2023; Mesman et al., 2024; Mishra et al., 2016; National Academies of Sciences et al., 2020; OECD Science, 2017; Romette et al., 2018; Schofield, 2010; Shetty et al., 2017; Strecker et al., 2023; Virapongse et al., 2024)
- Does CRRDI have the business and financial expertise to ideate new ways of building sustainable repositories? Data and computer experts aren’t enough. Science is a team sport and needs broader perspectives on the committee, including business, foundations and other funders.

2.2 Establish a multi-sectoral consortium...

- Although the idea of a multi-sectoral consortium isn't new, CRDDI should be realistic in how long it will take to make something like this happen. It is possible to start a consortium, in the plan's timeline but how will you motivate senior leaders to come together and build a collaborative funding?
- What is CRDDI's long term strategy to stay funded to meet this longer-term timeline?
- How will you overcome the challenges of sharing of data at the international level, including data sovereignty and governance?

Pillar 3 — Develop a Shared Outreach, Advocacy, and Action Strategy

Long-term resilience depends on awareness, support, and the collective capacity to act. This pillar equips the community with shared advocacy tools, builds capacity to report and respond to threats, and connects related efforts rather than competing with them.

- The activity mapping is great. Pulling together advocacy toolkits is fine. Questions on implementation remain. Impact metrics on individual data repositories needs to be defined. CRRDI is asking individuals to do a lot in their free time. Volunteers don't have as much capacity as organizations wish.
- How will individuals who are federally funded and not able to advocate going to use your materials? Consider different terminology to be inclusive of those individuals.
- The alignment activity is important, but it is unclear what role the coalition would have as compared to the other ones such as the AGU coordinated resiliency coalition? How do you plan to account for potential conflict of interest with regard to funding sources that the strategic team has with other similar emerging initiatives to avoid potential conflicts of interest.

Coalition Governance, Accountability, and Sustainability

We'd welcome any input on the coalition model — concerns about the structure, what accountability would need to look like in practice, what your organization would need in order to participate, or what the charter must address that isn't reflected here.

- Staffing will occur with "dedicated support for core coordination functions" (p. 19). This area lacks detail, which doesn't work well in business development. It shows you are willing to keep the funding and coalition in a flexible state, depending on the organization(s) that fund it. That is troublesome as it looks like you could give your funders too much leeway in deciding the future of the coalition.
- The core staffing model is not defined at all – roles and responsibilities, decision-making and accountability abilities, administrative functions, remote vs in-office environment, etc. That merits a great deal more consideration to determine whether this entity would be viable. To have the initial staffing of the committees based entirely on volunteers is concerning. It does not lend itself well to solid accountability or high-quality expertise (especially beyond individuals that have an altruistic

commitment to research data to those that bring other important skill sets), and perpetuates burnout. Nor can the organizations that underly the coalition necessarily afford to have their staff devote much time to this endeavour for a significant stretch.

- It is unclear why there needs to be a membership model at all? Couldn't this be a service-driven entity without a membership, but rather identifying early adopters and collaborations with existing entities (ESIP, WDS, CTS, AGU) that have synergistic objectives? It seems overwhelming for repositories, who are struggling, to have one more administrative membership burden to manage. Metrics are important to demonstrate impact, but that can be done without an explicit membership model.
- On many levels, I agree with having a clear sustaining and sunsetting plan for the coalition itself; beyond being an important facet of good governance for itself, these criteria can serve as a demonstration that it is doing for itself what it hopes to do for repositories. It is also consistent with the Principles of Open Scholarly Infrastructure.

Please provide any additional comments or suggestions regarding the strategic plan that may not fit neatly into the boxes above.

- What happens when a repo is found at risk?
- What about closed data?
- How will you engage the funders?
- Given that this coalition seems to be US-focussed, is there any intent to coordinate internationally especially in recognition of cross-border data sharing and researcher collaborations.

Our goal is for CRRDI to be built by and with the community it serves — and there are many ways to contribute whether you want to help shape the final plan, contribute expertise or resources to CRRDI's formation and implementation efforts, or simply stay informed as the coalition takes shape, we'd like to know. Answering here isn't a commitment; it helps us understand where interest lies and how to involve you as CRRDI forms.

Would you or your organization be interested in:

☐ Joining CRRDI as a member

☐ Funding or otherwise resourcing CRRDI to implement this strategic plan

☐ Partnering on specific strategies or workstreams

☐ Contributing expertise without a formal role

☐ Staying informed

☐ Not sure yet

☐ No, but thanks for asking

If you're interested in specific strategies or workstreams, which ones?

- Pillar 1 — Assess and Monitor the Repository Landscape
- Pillar 3 — Develop a Shared Outreach, Advocacy, and Action Strategy

REFERENCES

Baker, K. S., & Millerand, F. (2024). The Incremental Growth of Data Infrastructure in Ecology (1980–2020). *Ecology and Evolution*, *14*(12). https://doi.org/10.1002/ece3.70444

Baker, M. (2012). Databases fight funding cuts. *Nature*, *489*(7414), 19–19. https://doi.org/10.1038/489019a

Berman, F., & Cerf, V. (2013). Who Will Pay for Public Access to Research Data? *Science.*, *341*(6146), 616–617. https://doi.org/10.1126/science.1241625

Bothner, B., Lusetti, S. L., Seville, R. S., Baker, J. E., Barnes, B., Hoffmann, P. R., & Hovde, C. J. (2025). RAIN: a multistate research alliance to facilitate collaboration, increase student opportunities, and share core facility resources. *Advances in physiology education*, *49*(2), 374. https://doi.org/10.1152/advan.00217.2024

Burns, C. S., Lana, A., & Budd, J. M. (2013). Institutional repositories: Exploration of costs and value. *D-Lib magazine*, *19*(1-2). https://doi.org/10.1045/january2013-burns

Dillo, I., & Hodson, S. (2015, June 4, 2025). Data Publication: Cost Recovery for Data Centres Interest Group. IASSIST 2015, Minneapolis, MN.

Johnson, T. R., Bourne, Philip E. (2023). The Biological Data Sustainability Paradox [epub]. Retrieved 3/20/25, from https://arxiv.org/abs/2311.05668

Mesman, J. P., Barbosa, C. C., Lewis, A. S. L., Olsson, F., Calhoun-Grosch, S., Grossart, H.-P., Ladwig, R., La Fuente, R. S., Münzner, K., Nkwalale, L. G. T., Pilla, R. M., Suresh, K., & Wain, D. J. (2024). Challenges of open data in aquatic sciences: issues faced by data users and data providers. *Frontiers in environmental science*, *12*. https://doi.org/10.3389/fenvs.2024.1497105

Mishra, A., Schofield, P. N., & Bubela, T. M. (2016). Sustaining large-scale infrastructure to promote pre-competitive biomedical research: lessons from mouse genomics. *New biotechnology*, *33*(2), 280–294. https://doi.org/10.1016/j.nbt.2015.10.002

National Academies of Sciences, E., Medicine . Committee on Biological Collections: Their Past, P., Future, C., Options for Sustaining Them. issuing, b., National Academies of Sciences, E., Medicine . Board on Life, S., National Academies of Sciences, E., Medicine . Division on, E., & Life, S. (2020). *Biological collections : ensuring critical research and education for the 21st century*. The National Academies Press.

OECD Science, T. a. I. P. P. (2017). Business models for sustainable research data repositories. In (pp. 80). Paris: OECD Publishing.

Romette, J. L., Prat, C. M., Gould, E. A., de Lamballerie, X., Charrel, R., Coutard, B., Fooks, A. R., Bardsley, M., Carroll, M., Drosten, C., Drexler, J. F., Günther, S., Klempa, B., Pinschewer, D., Klimkait, T., Avsic-Zupanc, T., Capobianchi, M. R., Dicaro, A., Ippolito, G.,…Lieutaud, P. (2018). The European Virus Archive goes global: A growing resource for research. *Antiviral Res*, *158*, 127–134. https://doi.org/10.1016/j.antiviral.2018.07.017

Schofield, P. N., Eppig, Janan, Huala, Eva, Hrabe de Angelis, Martin, et. al. . (2010). Sustaining the Data and Bioresource Commons. *Science*, *330*(6004), 592–593. https://doi.org/https://www.science.org/doi/10.1126/science.1191506

Shetty, S., Red, V., Kamhoua, C., Kwiat, K., Njilla, L., Blowers, M., Williams, J., & Hall, R. D. (2017, 2017). Data provenance assurance in the cloud using blockchain.

Strecker, D., Pampel, H., Schabinger, R., & Weisweiler, N. L. (2023). Disappearing repositories: Taking an infrastructure perspective on the long-term availability of research data. *Quantitative science studies*, *4*(4), 839–856. https://doi.org/10.1162/qss_a_00277

Virapongse, A., Gallagher, J., & Tikoff, B. (2024). Insights on Sustainability of Earth Science Data Infrastructure Projects. *Data science journal*. https://doi.org/10.5334/dsj-2024-014

-