



Contribution ID: 96

Type: Session

Emerging technologies in the global context: challenges and opportunities for the long-term environmental data management lifecycle.

Tuesday 14 October 2025 11:30 (1h 30m)

Conversations around the data lifecycle from creation to re-use frequently revolve around the challenges of limited resources, lack of understanding by parties at various points of the process of what is involved in making it successful, and lack of interest from funding and re-use communities (e.g. Borgman & Groth, 2025, Specht et al., 2025, Stahlman, 2022). This is exacerbated when the data collected and preserved for reuse is to be sustained over the long term. Maintaining the integrity and quality of data collection, data deposition, curation and discovery while technology, funding, and expectations from those data evolve presents persistent challenges. Moreover, questions increasingly surround the provision of sufficient (super)computer access to properly curate data holdings while allowing effective augmentation in the data stocks, as well as open access to those holdings.

This 90-minute session will attempt to identify critical factors affecting the data lifecycle over the long term. These include the inherent expansion of data over time across domains and knowledge contexts, the effect of emerging technologies, the increasing energy and financial cost of handling, curating, and using data, and who takes responsibility for the curation of data for the common good. We shall use environmental data for research as our domain focus.

Invited panellists from various stages of the data lifecycle (see details below) will use their experiences and how they have overcome the challenges of maintaining good practice against adversity to answer the three questions listed below. Each panellist will present pre-prepared short statements (not all panellists will necessarily have comments against every question) followed by a short Q & A with the audience.

How do we balance the promise of emerging technologies with the practical risks of data loss and preservation challenges across the long-term data lifecycle?

What does true democratization of environmental data look like—and who might be left out?

How can we ensure that our environmental data management practices remain sustainable—both environmentally and ethically—in a rapidly shifting global context?

The session will close with audience discussion guided by various data lifecycle stages in relation to the panel discussion (20 minutes). This will help formulate key themes from the session, and participants (panellists and attendees) will be invited to contribute to a paper aimed to describe a clear understanding of current status and provide guidance for future work, such as (a) the Identification of possible roles of new technologies, (b) the quantification of the cost of high quality, persistent and abundant data, and (c) the identification of risks to data when supported by big corporations, and mechanisms to reduce risk.

The session will be supported by a questionnaire of the audience and an on-line space for future collaboration.

References

Borgman CL, Groth P (2025) Harvard Data Science Review 7, doi:10.1162/99608f92.35d32cfc

Specht A, et al. (2025) Data Science Journal 24, 1. doi:10.5334/dsj-2025-001

Stahlman GR (2022) Journal of the Association for Information Science and Technology 73, 1692–1705. doi:10.1002/asi.24687

Panellists

Our assembled experts represent various components of the data lifecycle. Insights into the challenges of data retention will be provided by Wim Hugo, a member of the EOSC long-term data retention task force, while experience of the data demands imposed on an observatory and repository will be provided by Siddeswara Guru. Greg Maurer of the US-LTER can provide practical ways to achieve consistency in data acquisition across a distributed network of research sites, while Gretchen Stahlman brings expertise in data curation ed-

ucation and legacy data integration. Stephen Bird (or colleague from the Queensland Cyber Infrastructure Foundation, part of an Australia-wide network that provides cloud facilities for research organisations) will bring understanding of the options and limitations of cyber infrastructure support for the present and the future. Shelley Stall is a pre-eminent global thinker and promoter of Open Data practices for researchers and an expert in scholarly publishing, while Tavita Su'a provides fundamental understanding of the requirements for building and supporting an emerging data network and repository. As chair, Alison Specht brings theory and practice in field (including transdisciplinary), experimental, and synthesis centre research, data management and education, with an eye to timeliness.

Stephen Bird: Queensland Cyber Infrastructure Foundation (QCIF), 0009-0001-9846-0990; Siddeswara Guru: TERN, University of Queensland, 0000-0002-3903-254X; Wim Hugo: Chief Technology Officer for DANS, an institute of the KNAW (Royal Dutch Academy of Science) 0000-0002-0255-5101; Greg Maurer: US Long Term Ecological Research Network, State University of New Mexico, 0000-0002-3007-8058; Gretchen Stahlman: Florida State University School of Information, 0000-0001-8814-863X; Shelley Stall: Data Vice-President, American Geophysical Union, 0000-0003-2926-8353; Alison Specht: TERN, University of Queensland, 0000-0002-2623-0854; Tavita Su'a: South Pacific Regional Environmental Portal and Data Hub.

Primary authors: SPECHT, Alison (TERN, University of Queensland); Dr STAHLMAN, Gretchen (Florida State University School of Information)

Presenter: SPECHT, Alison (TERN, University of Queensland)

Track Classification: SciDataCon Persistent Themes: Data Stewardship