Contribution ID: **136**   Type: **Presentation**

Improving Research Data Reuse Through Structured Data Management, Review Processes, and Persistent Identifiers

Wednesday 15 October 2025 14:00 (11 minutes)

Achieving genuinely open and reusable research data requires structured data management and robust documentation to ensure interoperability and practical reuse. Many openly published datasets remain underutilised due to inadequate documentation, incomplete metadata, or unresolved sensitivities.

To address these challenges, we implemented a structured data management framework within Australia's National Environmental Science Program Marine and Coastal Hub—a national funding initiative supporting hundreds of diverse environmental science research projects across multiple research organisations. These projects range from environmental restoration and mapping environmental assets to monitoring vulnerable species populations. Managing data across such varied disciplines and data types requires a rigorous yet flexible approach.

Our structured workflow begins with targeted "data discussions" at key project milestones: project initiation, annually during the project, and shortly before completion. During these discussions, data wranglers guide research teams to proactively address issues such as licensing, data sensitivities, Indigenous data governance, and detailed planning for dataset documentation and publication. For sensitive datasets, a structured Access Control Plan is developed to ensure the data remains FAIR (Findable, Accessible, Interoperable, and Reusable), even if restrictions on public access apply. This early and consistent engagement helps researchers resolve data management issues promptly, significantly improving documentation quality and reuse potential.

Another key aspect is our detailed dataset review, conducted as datasets are submitted, typically toward the project's end. Researchers know their datasets will be assessed for completeness, clarity, and usability. Researchers complete a structured "dataset reporting form," breaking metadata creation into clear, guided questions with practical examples. Any gaps or ambiguities identified are communicated back to researchers for clarification, ensuring the final documentation fully supports dataset reuse.

This structured review process significantly enhances metadata quality. A comparison of our repository's metadata records with similar repositories lacking structured intervention shows our records typically contain significantly more detail, helping to ensure datasets are easier to reuse with less ambiguity.

We systematically integrate Persistent Identifiers (PIDs)—including Digital Object Identifiers (DOIs) for datasets, ORCIDs for researchers, ROR identifiers for institutions, and RAiDs for research projects—into ISO19115-3 metadata records. DOIs ensure persistent access and straightforward citation, increasing dataset visibility and reuse. ORCIDs and RORs resolve ambiguity regarding authorship and institutional attribution, improving credit assignment. RAiDs track dataset impact back to research funding sources, demonstrating value and impact for funders, aiding future investment.

This combined approach—structured researcher engagement, supportive documentation reviews, and systematic PID integration—provides a transferable model significantly enhancing dataset quality, visibility, and reuse potential. Our experience demonstrates proactive data wrangling and strategic PID integration ensure research data achieves genuine openness, interoperability, and practical reuse, offering valuable, replicable insights for global data managers and repositories.

Primary author: Dr LAWREY, Eric (Australian Institute of Marine Science)

Co-authors: BABICCI, Suzannah (Australian Institute of Marine Science); Dr FLUKES, Emma (Institute for Marine and Antarctic Studies (IMAS), University of Tasmania); JOHNSTON, Jared (Australian Institute of Marine Science); MARTIN, Julia (Australian Research Data Commons)

Presenter: Dr LAWREY, Eric (Australian Institute of Marine Science)

Session Classification: Presentations Session 7: Open research through Interconnected, Interoperable, and Interdisciplinary Data

Track Classification: SciDataCon2025 Specific Themes: Open research through Interconnected, Interoperable, and Interdisciplinary Data