Practical Data Stewardship for Salmon Biologists

DRAFT. A Blueprint for Domain-Specific Best Practices in Fisheries Data Stewardship

Brett Johnson Scott Akenhead Katie Barnas
Jennifer Bayer Tomas Bird Samuel Cimino Graeme Diack
Lara Erikson Nancy Leonard Catherine Michielsens
Fiona Martens Emily Lescak Gottfried Pestal Matt Jones
Mark Saunders Yi Xu

Keywords: Salmon data stewardship, data interoperability, FAIR principles, persistent identifiers (PIDs), controlled vocabularies, metadata standards, application programming Interface (API), data citation, ontology development

Abstract

Fisheries research, management, and conservation increasingly generate vast and diverse data crucial for timely decision-making. Yet these data remain largely fragmented across jurisdictions, disciplines, and outdated infrastructure, limiting their use in responsive management. Biologists are increasingly taking on data stewardship responsibilities to address these challenges, often without clear guidance, training, or support. Shared, community-agreed practices for implementing domain-specific data standards are needed to move beyond generic data management guidance toward fit-for-purpose tools and workflows. To address this gap—and to show how other communities can do so—we develop seven practices for salmon data stewardship and demonstrate their application through a real-world case study. We provide practical guidance for those transitioning into these essential stewardship roles, including domain-specific tools, templates, and examples from salmon research and management. We argue that effective salmon management depends on formally establishing data stewardship as a dedicated, institutionally supported professional role. These practices integrate both sociocultural and technical approaches to ensure data meet modern open science principles and respect Indigenous Data Sovereignty. Through a case study of a historical sockeye salmon productivity analysis spanning Pacific Coast jurisdictions, we highlight how clearly defined data stewardship practices enhance

data reproducibility, integration, and management efficacy. With a foundation of shared practices, data stewards will enable faster, more transparent decision-making, support development of machine-actionable datasets with high-quality metadata and consistent semantics that enable automated analysis, and expand the use of cross-jurisdictional datasets—ultimately strengthening the management and conservation of salmon populations and the ecosystems they inhabit, and, by extension, other data-rich fisheries data domains.

The Data Stewardship Challenge

Integrated, timely, and high-quality data are essential for effective fisheries research, management, and conservation. Such data underpin robust stock assessments, inform adaptive management strategies, enable rapid responses to emerging threats, and support transparent decision-making. Yet across the fisheries domain, biologists face persistent challenges in achieving these goals. Data on fish populations, health, and environmental conditions are often fragmented, inconsistently measured, and incomplete across time, space, and life-history stages (NOAA Data Governance Committee 2024). These issues limit the utility of these data for research and management.

Cross-Jurisdictional Data Fragmentation

The challenges are especially pronounced in salmon science, where data must be integrated across multiple ecological regions and jurisdictional boundaries (Groot and Margolis 1991). Salmon biologists routinely collect information managed by diverse agencies and institutions, often in isolation and without a focus on interoperability. While localized successes in data coordination exist—particularly within regional fisheries management offices and treaty commissions—salmon data integrated across agencies for each phase of the salmon life cycle is uncommon and costly for most programs (Marmorek et al. 2011; Inman et al. 2021; Diack et al. 2024). Most salmon datasets remain confined within institutional silos, often undocumented, stored in outdated systems, or formatted according to internal standards that are incompatible with broader integration efforts. Even within organizations, data can be siloed by data type or program with freshwater data going in one data system while estuary, open-ocean, and commercial fishery data each housed in separate data systems with limited ability to easily re-connect the data through shared identifiers.

This fragmentation is compounded by the number of disciplines and organizations involved. Geneticists, oceanographers, freshwater ecologists, stock assessment biologists, and fisheries managers all contribute data using their own field-specific conventions and workflows. Meanwhile, data is distributed across federal, state, provincial, tribal, and academic institutions—each with its own mandates, technologies, and metadata requirements. Many salmon data-holding organizations rely on aging infrastructure or opaque, undocumented standards that lag behind modern open-science practices. This tangle of disciplinary and institutional fragmentation slows integration, hinders reproducibility, and delays analyses that could otherwise inform time-sensitive management decisions, conservation actions, or restoration plans. When critical datasets are hard to find, access, or interpret, biologists and analysts lose valuable time trying to reconstruct or harmonize them, reducing transparency, increasing the risk of errors, and delaying urgent conservation or management responses.

The Need for Coordinated Stewardship

The growing complexity of fisheries management, combined with escalating environmental uncertainties due to climate change, demands rapid, integrated, and robust data analyses (Bull et al. 2022, ward_surveyjoin_2025). The mismatch between fragmented data systems and fixed administrative boundaries creates an urgent need for interoperable, dynamic, multi-scale data stewardship that can adapt to shifting ecological and management priorities. Despite the scale and importance of these datasets, biologists who collect and manage salmon data are often expected to act as de facto data stewards without training, guidance, institutional support, or access to community-agreed best practices. Tasks such as documenting methods, aligning terminology, formatting for data sharing, and publishing data are typically performed off the side of a biologist's desk. A lack of institutional support (Diack et al. 2024), training (Volk, Lucero, and Barnas 2014), and dedicated roles for data management further relegate critical data stewardship tasks to an *ad hoc* status. The absence of clear roles, standards, and community-endorsed practices leaves even motivated scientists unsure how to structure their data for future use. As a result, data stewardship is inconsistent and reactive, and data integration remains a major bottleneck to adaptive management and ecosystem-scale learning.

Both researchers and managers struggle to align their data with community-agreed principles such as FAIR (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al. 2016) and Indigenous Data Sovereignty frameworks like the CARE principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) (Carroll, Rodriguez-Lonebear, and Martinez 2019; Jennings et al. 2023). Adhering to CARE data management principles is all the more important when it comes to salmon related data given the sociocultural importance of salmon to First Nations, Tribes, and Indigenous communities throughout the North Pacific and North Atlantic regions (Ween and Colombi 2013; Earth Economics 2021). Large volumes of data collected through long-term monitoring programs hold tremendous value, especially for secondary users—but are often inaccessible due to a lack of time, resources, and incentives for data producers to publish them (LINDENMAYER et al. 2012). Without clear support and guidance, well-intentioned practitioners are left with ad hoc approaches that limit reuse and interoperability. This gap can only be bridged by equipping both data producers and stewards with tools, support, and institutional backing to publish interoperable, machine-readable metadata and datasets in alignment with shared principles.

Framework for Action

In this paper, we provide actionable practices, examples, and workflows to help salmon biologists improve the usability, reproducibility, and long-term impact of their data. We develop seven best practices for salmon data stewardship and demonstrate their application through a real-world case study of cross-jurisdictional sockeye productivity analysis. Our case study shows how cross-jurisdictional alignment of terms and reproducible pipelines can enable faster status assessment updates and more responsive management decisions. Our goal is to support salmon biologists and the broader research and management community to effectively steward salmon data. To keep this broadly useful, we emphasize patterns—lifecycle planning, metadata governance, vocabulary alignment, reproducible publishing, and role clarity—that any taxa-centric community can adopt, substituting their own standards and tools. We also map the seven practices to widely used data-lifecycle models to make adoption straightforward outside salmon contexts.

A coordinated approach to stewarding salmon data should follow established open science standards and adhere explicitly to FAIR principles, tailored specifically for salmon research and management. Our practices build upon existing standards and vocabularies including Darwin Core, OBIS, schema.org, and OBO Foundry ontologies, ensuring compatibility with broader biodiversity informatics infrastructure rather than reinventing foundational frameworks. Achieving meaningful interoperability demands both breadth and depth. **Broad interoperability** integrates diverse scientific domains, systems, and formats, requiring structured, machine-readable data and metadata published openly for maximum discoverability. **Deep interoperability** demands precise definitions of salmon-specific terms and methods, ensuring data remains meaningful and usable across contexts. Salmon data stewards can improve conservation outcomes for salmon by coordinating across boundaries to develop a shared foundation of data stewardship practices. To address these foundational challenges, we must establish clear data stewardship roles and practices that span the entire data lifecycle—from collection and documentation through integration and long-term preservation.

Defining Data Stewardship in Salmon Science

Data stewardship encompasses the coordinated practices, roles, and responsibilities necessary to effectively manage, share, and reuse data throughout its lifecycle (NOAA 2007; Plotkin 2014; Peng et al. 2018). Within fisheries science, stewardship involves ensuring data quality, compliance with agreed-upon standards, and the establishment of clear governance to guide data collection, documentation, integration, and preservation. However, salmon data stewardship goes beyond mere technical data management; it involves actively facilitating collaboration, communication, and consensus-building among data producers and users across multiple institutions and jurisdictions.

Data stewardship represents a critical subdiscipline within the broader field of data science. While data science is often narrowly associated with machine learning and statistical modeling, we adopt a more comprehensive view that encompasses how we treat, handle, and represent data, along with the social and technical information systems that enable data use for science. Data stewardship focuses on the practical implementation of these principles—ensuring that data infrastructure, standards, and practices actually serve scientific and management needs rather than remaining theoretical constructs.

Effective salmon data stewards serve as boundary spanners and community managers, convening diverse stakeholders across agencies, First Nations, Tribes, and academic institutions to build sustained communities of practice. This boundary-spanning role is particularly critical in transboundary contexts where data integration requires navigating complex jurisdictional and cultural boundaries (Ward et al. 2025). By facilitating communication, translating between different organizational cultures and technical systems, and maintaining long-term relationships, data stewards create the social infrastructure necessary for effective cross-boundary data collaboration.

Box 1: Critical Functions of Salmon Data Stewards

Effective salmon data stewards perform several critical functions:

- **Technical oversight**: Ensuring metadata completeness, adherence to standardized terminologies and vocabularies, and robust quality assurance protocols.
- **Social and organizational facilitation**: Leading stakeholder engagement, capacity-building activities, and negotiation of data access and sharing agreements, including addressing First Nations, Tribes, and Indigenous Peoples' rights and interests in data governance.
- **Institutional advocacy**: Championing the institutional recognition of data stewardship roles, promoting sustained investment and dedicated resources for data management infrastructure and practices.
- **Implementation and adoption facilitation**: Actively promoting data use and ensuring that standards and practices remain practical and relevant by maintaining close contact with real-world applications. This includes monitoring data utilization, gathering feedback from users, and iteratively refining standards based on actual implementation challenges to prevent theoretical approaches that fail in practice.

Data stewards can implement FAIR and CARE principles through concrete technical and governance mechanisms they control, such as documenting consent constraints and access levels in metadata, using controlled vocabularies to ensure consistent terminology, and establishing repository roles that enforce data sovereignty requirements. For example, stewards can document consent constraints in metadata fields and enforce access restrictions via repository user roles, ensuring that Indigenous data sovereignty is respected while maintaining data discoverability and appropriate reuse. This governance approach is particularly critical for sensitive data such as Traditional Knowledge and sensitive habitat locations, where stewardship practices must balance open science principles with appropriate access controls and cultural protocols.

Data stewards play a critical role bridging the gap between biologists and Information Technology (IT) staff by translating data needs into application or data system features. A user-centred design approach to salmon data stewardship is critical and focuses on creating tools that align with biologists' needs. When data management is separated from biologists, accountability weakens, and quality issues go unnoticed. While IT expertise is essential for infrastructure and security, effective data system design requires IT to act as an enabler, rather than gatekeeper, provisioning self-serve data infrastructure. The Data Steward, serving as a translator between IT and biologists, enables biologists to engage independently with data systems, fostering ownership and accountability and ultimately improving data quality for research and management.

Dedicated stewardship roles empower salmon biologists to bridge disciplinary divides and jurisdictional barriers, transforming fragmented datasets into cohesive, interoperable resources. By proactively defining, implementing, and maintaining data standards and workflows, salmon data stewards create conditions for timely, accurate, and reproducible analyses. Such stewardship positions salmon biologists to better inform adaptive management decisions, ultimately strengthening salmon conservation and resilience.

Updating Pacific-wide Sockeye Productivity: A Case Study for What Agencies Could Do Now

This case study revisits a Pacific Coast-wide sockeye productivity dataset assembled from diverse agency sources by academic researchers (Peterman and Dorner 2012). We reflect not on the significant work the research team accomplished, but rather on the preventable institutional and technical barriers that impeded their work—and continue to burden data updates and reuse efforts today. Their study examined productivity trends across 64 sockeye salmon stocks spanning Washington, British Columbia (B.C.), and Alaska. However, attempting to replicate or build upon this analysis today is an arduous, time-consuming, and error-prone endeavour due to fragmented data sources, inconsistent formats, and lack of standardized practices among the key institutions involved: the Washington Department of Fish and Wildlife (WDFW), Fisheries and Oceans Canada (DFO), and the Alaska Department of Fish and Game (ADF&G).

Each section below highlights a key challenge faced by the team and proposes practical steps based on our best practices (Table 2) that data-holding agencies could do to enable easier integration, validation, and updating of salmon datasets across jurisdictions and decades. This case study illustrates how implementing the foundational concepts and practical recommendations outlined in this paper can transform data stewardship practices within these organizations. By doing so, they can significantly enhance data accessibility, quality, and interoperability—ultimately enabling more efficient and accurate analyses that support salmon conservation and management.

Challenge 1: Interpreting the Data — What do these numbers actually mean?

Peterman's team frequently worked with datasets that lacked basic contextual information. Fields such as "year," "return," or "age class" were often undefined or inconsistently used. For example, some datasets recorded returns by calendar year while others used brood year, and few included metadata to clarify the distinction. In many cases, the team had to reconstruct metadata by back-checking against reports or simulating assumptions (e.g., about age structure) to interpret the data correctly.

Remedies:

- **Best Practice 3: Make Data, People, Projects, and Outputs Discoverable, Linked and Citable with Persistent Identifiers (PIDs).** Assigning PIDs such as digital object identifiers (DOIs) to protocols, methods, and people (via ORCIDs) and linking them together using data stores and catalogues links data to its provenance and ensures that methods, context, and interpretation decisions are traceable.
- **Best Practice 4. Use Shared Data Models, Vocabularies and Metadata to Enable Integration.** To prevent this kind of ambiguity, agencies can now adopt internationally recognized metadata schemas such as ISO 19115 or Ecological Metadata Language, data models (Darwin Core Data Package) to model age and age type data concepts, and use controlled vocabularies to restrict the permissible values in the age field to calendar year, brood year, or otherwise.

Challenge 2: Accessing and Using the Data — Where is it stored, and how do I get it?

The Peterman dataset was compiled from multiple files scattered across email inboxes, regional offices, and grey literature. Data were stored in inconsistent formats, lacked clear versioning, and were difficult to discover outside of specific research networks. Even today, no API or structured access mechanism exists to update or query the data programmatically. As a result, researchers hoping to build on the dataset may have to start from scratch.

Remedies:

- **Best Practice 2: Reuse Proven Infrastructure to Save Time and Increase Interoperability**
  Rather than developing bespoke data catalogues or repositories, agencies should adopt existing catalogues used beyond their own institution such as the Ocean Biodiversity Information System, Zenodo, or the Knowledge Network for Biocomplexity). These are proven platforms with a broad user base that support persistent storage, discoverability, and interoperability.

- **Best Practice 5: Store and Analyze Data in Ways That Others Can Easily Access, Use, and Trust**
  Agencies can use open-access data repositories or their own institutional data repositories or catalogues that make data discoverable using PIDs and provide programmatic access to data possible using Application Programming Interfaces.

Challenge 3: Sustaining the Dataset — Who is responsible, and why should I contribute?

Once Peterman and his team completed their analysis, no formal plan existed for sustaining or updating the dataset. Responsibility for ongoing maintenance fell informally to former students and collaborators. Despite its national and international relevance, the dataset was never adopted by an agency as a living product. Moreover, the original data contributors often lacked incentives, support, or recognition for their efforts—conditions that persist in many data environments today.

Remedies:

- **Best Practice 1: Make Data Governance Explicit to Support Trust and Reuse**
  Agencies should define roles, responsibilities, and decision-making processes through formal governance mechanisms such as data product charters. Use a Data Management Plan with a responisibility matrix such as "responsible, approver, consulted, informed" (RACI) to clarify govermamce, assign maintenance responsibility, and ensure continuity across staff turnover and institutional change.
- **Best Practice 6: Incentivize and Track Data Sharing and Reuse** Visibility, credit, and metrics are critical for motivating data sharing. Agencies can embed citation guidance in metadata and track dataset reuse through COUNTER-compliant dashboards or DataCite APIs.
- **Best Practice 7: Build Community Through Co-Development and Mutual Benefit** Effective data stewardship requires collaboration between biologists, First Nations, Tribes, Indigenous communities, managers, and data professionals. Participatory design ensures that systems and standards meet user needs and are adopted over time. *Practical application:* Facilitate cross-jurisdictional working groups to co-develop data standards and align on shared outcomes for priority datasets.

While the analytical contribution of the Peterman productivity dataset remains significant, the barriers encountered in compiling, interpreting, and maintaining the data are instructive. These challenges are not unique to Peterman's team—they reflect systemic gaps in data governance, documentation, infrastructure, and incentives. By adopting the seven best practices detailed in Table 2, agencies and researchers can transform legacy datasets into living resources, enabling reproducibility, easing collaboration, and accelerating insight across the salmon research and management community.

The challenges and solutions demonstrated in this salmon case study generalize across fisheries and environmental monitoring domains. Cross-jurisdictional data harmonization, quality assurance and control patterns, standardized metadata requirements, and long-term archiving strategies are universal needs that extend far beyond salmon science. Similar barriers and solutions apply to trawl survey data integration, invertebrate monitoring programs, and water quality datasets that span multiple agencies and jurisdictions.

How our seven practices align to data lifecycle models

Our seven best practices map directly to established data lifecycle models, demonstrating their broad applicability beyond salmon science. The NOAA Data Lifecycle provides a widely recognized framework with six sequential stages (Plan, Obtain, Process, Preserve, Access, Disposition) and four cross-cutting elements (Document, Track and Monitor, Quality, Security) (NOAA Data Governance Committee 2024). This alignment ensures our practices are grounded in established federal data management standards and can be readily adopted by other agencies and research communities.

The mapping shown in Table 1 demonstrates how each practice addresses specific lifecycle stages while the cross-cutting elements ensure comprehensive data stewardship throughout the entire lifecycle. For example, Practice 1 (Data Governance) spans the entire lifecycle from planning through disposition, while Practice 4 (Shared Data Models) primarily supports the Process and Preserve stages. This systematic alignment with established frameworks enhances the credibility and portability of our approach across different domains and institutions.

Table 1: Mapping of seven best practices to NOAA Data Lifecycle stages and cross-cutting elements

Best Practice	Plan	Obtain	Process	Preserve	Access	Disposition	Cross-cutting
1. Data Governance							Document, Quality
2. Reuse Infrastructure							Track and Monitor
3. Persistent Identifiers							Document, Track
4. Shared Data Models							Quality
5. Accessible Storage							Security, Quality
6. Incentivize Sharing							Track and Monitor

Best Practice	Plan	Obtain	Process	Preserve	Access	Disposition	Cross-cutting
7. Community Building							Document, Quality

Metadata governance as a cross-cutting foundation

Unlike the sequential stages of the data lifecycle, metadata governance operates as a continuous, cross-cutting practice that spans all phases simultaneously. While data moves through Plan → Obtain → Process → Preserve → Access → Disposition, metadata governance must be active throughout, ensuring that documentation, quality control, and discoverability are maintained at every stage. This cross-cutting nature means that metadata governance failures at any point can compromise the entire data stewardship effort, making it a critical foundation rather than a discrete step in the process.

The lifecycle mapping in Table 1 reveals that data governance elements appear in every stage: planning metadata requirements (Plan), documenting collection methods (Obtain), structuring and validating metadata (Process), ensuring long-term preservation (Preserve), enabling discovery and access (Access), and managing final disposition (Disposition). This pervasive presence underscores why metadata governance must be treated as an institutional capability rather than a project-specific task, requiring dedicated resources, trained personnel, and systematic processes that operate continuously across all data activities.

Table 2: Best practices and practical applications of salmon data stewardship

Best Practice	Practical Applications
1. Make Data Governance Explicit to Support Trust and Reuse. Establishing clear governance structures ensures quality, accountability, and compliance with FAIR and CARE principles. It enables trust and long-term stewardship across multi-organizational projects.	- Document roles and responsibilities using a Data Product Governance Charter and structured frameworks (e.g., DACI or RACI). - Integrate CARE principles to respect First Nations, Tribes, and Indigenous data rights. - Form a governance or oversight committee to review data standards, timelines, and agreements.

Best Practice	Practical Applications
2. Reuse Proven Infrastructure to Save Time and Increase Interoperability. Leveraging existing platforms and technologies reduces costs and improves long-term interoperability and sustainability.	- Use domain-specific repositories like OBIS or GBIF. - Publish and archive data with KNB or Zenodo.
3. Make Data, People, Projects, and Outputs Discoverable, Linked and Citable with PIDs. Persistent identifiers (PIDs) connect data with researchers, institutions, and outputs—supporting data citation, reuse, and automated attribution.	- Encourage use of ORCID iDs for researchers. - Use ROR IDs for institutions. - Assign DOIs via DataCite for data packages. - Embed DOIs in dashboards and metadata.
4. Use Shared Data Models, Ontologies and Metadata to Enable Integration. Common vocabularies, metadata standards, and ontologies support integration across systems and preserve semantic meaning.	- Adopt ISO 19115, EML, or DataCite metadata standards. - Re-use terms defined in Salmon Domain Ontology. - Model datasets using the Darwin Core Data Package Model.
5. Store and Analyze Data in Ways That Others Can Easily Access, Use and Trust. Structured and accessible data formats ease reusability, and support integration with analytical tools and applications while data analyzed or wrangled using programmatic scripts (R, Python etc.) enable reproducibility and increase trust.	- Provide APIs using FastAPI, Flask, or Django REST. - Archive in trusted repositories (e.g., GBIF, FRDR, USGS). - Write scripts in a programming language to wrangle, transform, and analyze data. - Use GitHub to host code for collaboration and transparency and the GitHub / Zenodo integration for DOI assignment and preservation.

Best Practice	Practical Applications
6. Incentivize and Track Data Sharing and Reuse. Recognizing data contributors and tracking reuse promotes a culture of sharing and supports professional recognition.	- License data with CC-BY 4.0. - Include citation text and visible credit fields. - Use COUNTER metrics and DataCite APIs to monitor reuse. - Encourage dataset citation in references.
7. Build Community Through Co-Development and Mutual Benefit. Engaging users early ensures tools and standards meet real-world needs and enhances long-term stewardship.	- Participate in RDA Salmon Interest Group. - Facilitate workshops for metadata and vocabulary alignment. - Support community-engaged research with tangible benefits.

Conclusion

Salmon biologists and data stewards across the globe have generated extensive datasets on salmon abundance, environmental conditions, and biological characteristics. When integrated, these data become valuable assets, a fact powerfully demonstrated by studies such as (Peterman and Dorner 2012). However, as noted by reports to the Cohen Commission (Marmorek et al. 2011), these data are often incomplete, inconsistently collected, and fragmented across institutions and jurisdictions. Integrating across such diverse sources can be done, but requires effort that is often not accounted for in smaller-scale studies. This fragmentation is a missed opportunity to deepen our understanding of the drivers of change across salmon life stages and regions, and limits the effectiveness of management decisions, particularly in the face of climate change and biodiversity loss.

But this limitation also reveals an opportunity. By adopting shared best practices in data governance, metadata standardization, persistent identification, infrastructure reuse, and community co-development we can radically improve the transparency, reusability, and interoperability of salmon data. A coordinated, future-oriented data stewardship strategy can transform the role of salmon data in science and management. The case study presented in this paper—drawn from one of the Pacific Region's most influential salmon survival syntheses (Peterman and Dorner 2012)—illustrates how technical and social data management gaps directly obstructed efforts to answer pressing questions. If some of the best practices we propose had been adopted by the data producers—such as documenting their datasets more thoroughly, storing data in accessible formats, or using persistent identifiers—substantial time and resources could have been saved. The case offers a clear and cautionary tale, as well as a hopeful roadmap.

The emergence of the data stewardship role (Plotkin 2014) represents one of the most critical institutional shifts needed to realize this vision. Historically, the work of managing, documenting, and maintaining data has been diffuse and undervalued—often falling to biologists without support, training, or recognition. As the volume and complexity of scientific data grow, so too does the need for clearly defined data stewardship responsibilities embedded within research teams and organizations. Training biologists in the principles and practices of data stewardship—while also supporting dedicated professionals who specialize in this work—is essential to sustaining trustworthy, reusable, and interoperable salmon data systems.

Realizing this vision requires concrete institutional commitments organizations should formally appoint dedicated data stewards with clear roles, responsibilities, and reporting structures. Agencies can adopt centralized metadata repositories and establish compliance metrics to track progress toward FAIR and CARE principles. Key implementation steps include: (1) designating stewardship roles within existing organizational structures, (2) investing in metadata management infrastructure, (3) establishing data governance committees with cross-organization representation, and (4) developing performance indicators that measure data discoverability, interoperability, and reuse. These institutional changes ensure that data stewardship becomes embedded in organizational culture rather than remaining an ad hoc responsibility.

The visionary future state is one where salmon researchers and stewards—across agencies, Indigenous Nations, academic labs, and community groups—can easily access and contribute to well-documented, versioned, and machine-readable datasets. In this future, field biologists, Indigenous guardians, modelers, and policymakers interact with a living knowledge system—one that is flexible, easy to implement, and rooted in principles of FAIRness Indigenous Data Sovereignty. Metadata standards, controlled vocabularies, and shared governance frameworks are not afterthoughts but integral to the culture of data collection and use. Scientists receive credit for publishing high-quality data, and users trust the provenance and structure of the datasets they rely on to make critical management decisions.

Realizing this vision will require investment in both people and systems. Key to this transformation is the emergence of the data steward as a professional role: a hybrid expert who understands operational field biology, information science, governance protocols, and community needs. As highlighted by Roche et al. (2020), institutionalizing data stewardship roles ensures long-term capacity for data governance, quality control, and interoperability—functions that are often neglected or left to informal actors. We must not only train new data stewards but also support and upskill biologists to take on stewardship responsibilities in collaborative, interdisciplinary settings. This is essential to address the "technical debt" of unmanaged data and to modernize research practices in line with open science norms. By embedding these practices into the everyday work of data generation, documentation, publication, and reuse, we can move salmon science decisively into the era of data-intensive discovery.

To do items:

- Discuss differences between data management plans, data governance charters, data sharing agreements
- Incorporate ref to Streamnet Data Exchange Standards somehow
- Add in figures
- fill out appendix 1 more thoroughly
- Refine Reorg the content in appendix 2 (traning roadmap) and decide if it Makes sense to put some of that content into a 3rd column in table 1

Competing interests

Acknowledgements

References

Bull, C D, S D Gregory, E Rivot, T F Sheehan, D Ensing, G Woodward, and W Crozier. 2022. "The Likely Suspects Framework: The Need for a Life Cycle Approach for Managing Atlantic Salmon (*Salmo Salar*) Stocks Across Multiple Scales." Edited by Wesley Flannery. *ICES Journal of Marine Science* 79 (5): 1445–56. https://doi.org/10.1093/icesjms/fsac099.

Carroll, Stephanie Russo, Desi Rodriguez-Lonebear, and Andrew Martinez. 2019. "Indigenous Data Governance: Strategies from United States Native Nations." *Data Science Journal* 18 (1): 31. https://doi.org/10.5334/dsj-2019-031.

Diack, Graeme, Tom Bird, Scott Akenhead, Jennifer Bayer, Deirdre Brophy, Colin Bull, Elvira de Eyto, et al. 2024. "Salmon Data Mobilization." *North Pacific Anadromous Fish Commission Bulletin*, December. https://doi.org/10.23849/npafcb7/x3rlpo23a.

Earth Economics. 2021. "The Sociocultural Significance of Pacific Salmon to Tribes and First Nations." Tacoma, Washington. https://www.psc.org/wp-content/uploads/wpfd/preview_files/The-Sociocultural-Significance-of-Salmon-to-Tribes-and-First-Nations(5da9942da9fb4fe0d77eb32bd6165e43).pdf.

Groot, Cornelis, and L. Margolis. 1991. *Pacific Salmon Life Histories*. UBC Press. https://books.google.ca/books?id=I_S0xCME0CYC.

Inman, Sarah, Janessa Esquible, Michael Jones, William Bechtol, and Brendan Connors. 2021. "Opportunities and Impediments for Use of Local Data in the Management of Salmon Fisheries." *Ecology and Society* 26 (2). https://doi.org/10.5751/ES-12117-260226.

Jennings, Lydia, Talia Anderson, Andrew Martinez, Rogena Sterling, Dominique David Chavez, Ibrahim Garba, Maui Hudson, Nanibaa' A. Garrison, and Stephanie Russo Carroll. 2023. "Applying the 'CARE Principles for Indigenous Data Governance' to Ecology and Biodiversity Research." *Nature Ecology & Evolution* 7 (10): 1547–51. https://doi.org/10.1038/s41559-023-02161-2.

Johnson, Brett, and Tim van der Stap. 2024. “Data Mobilization Through the International Year of the Salmon Ocean Observing System.” *North Pacific Anadromous Fish Commission Bulletin*, December. https://doi.org/10.23849/npafcb7/6a4ddpde4.

LINDENMAYER, DAVID B., GENE E. LIKENS, ALAN ANDERSEN, DAVID BOWMAN, C. MICHAEL BULL, EMMA BURNS, CHRIS R. DICKMAN, et al. 2012. “Value of Long-Term Ecological Studies.” *Austral Ecology* 37 (7): 745–57. https://doi.org/10.1111/j.1442-9993.2011.02351.x.

Marmorek, David, Darcy Pickard, Alexander Hall, Katherine Bryan, Liz Martell, Clint Alexander, Katherine Wieckowski, Lorne Greig, and Carl Schwarz. 2011. “Cohen Commision Technical Report 6-Fraser River Sockeye Salmon: Data Synthesis and Cumulative Impacts.” Vancouver, B.C. http://www.cohencommission.ca/.

NOAA. 2007. *Environmental Data Management at NOAA*. National Academies Press. https://doi.org/10.17226/12017.

NOAA Data Governance Committee. 2024. “Management of NOAA Data and Information, Data Management Handbook,” January. https://www.noaa.gov/sites/default/files/2025-03/NAO_212-15B_-_Data_Management_Handbook.pdf.

Peng, Ge, Jeffrey L. Privette, Curt Tilmes, Sky Bristol, Tom Maycock, John J. Bates, Scott Hausman, Otis Brown, and Edward J. Kearns. 2018. “A Conceptual Enterprise Framework for Managing Scientific Data Stewardship.” *Data Science Journal* 17. https://doi.org/10.5334/dsj-2018-015.

Peterman, Randall M., and Brigitte Dorner. 2012. “A Widespread Decrease in Productivity of Sockeye Salmon (*Oncorhynchus Nerka*) Populations in Western North America.” Edited by Jordan S. Rosenfeld. *Canadian Journal of Fisheries and Aquatic Sciences* 69 (8): 1255–60. https://doi.org/10.1139/f2012-063.

Plotkin, David. 2014. *Data Stewardship*. Elsevier. https://doi.org/10.1016/c2012-0-07057-3.

Roche, Dominique G., Monica Granados, Claire C. Austin, Scott Wilson, Gregory M. Mitchell, Paul A. Smith, Steven J. Cooke, and Joseph R. Bennett. 2020. “Open Government Data and Environmental Science: A Federal Canadian Perspective.” Edited by Tanzy Love. *FACETS* 5 (1): 942–62. https://doi.org/10.1139/facets-2020-0008.

Volk, Carol J., Yasmin Lucero, and Katie Barnas. 2014. “Why Is Data Sharing in Collaborative Natural Resource Efforts so Hard and What Can We Do to Improve It?” *Environmental Management* 53 (5): 883–93. https://doi.org/10.1007/s00267-014-0258-2.

Ward, Eric J, Philina A English, Christopher N Rooper, Bridget E Ferriss, Curt E Whitmire, Chantel R Wetzel, Lewis Ak Barnett, et al. 2025. “‘Surveyjoin‘: A Standardized Database of Fisheries Bottom Trawl Surveys in the Northeast Pacific Ocean.” https://doi.org/10.1101/2025.03.14.643022.

Ween, Gro, and Benedict Colombi. 2013. “Two Rivers: The Politics of Wild Salmon, Indigenous Rights and Natural Resource Management.” *Sustainability* 5 (2): 478–95. https://doi.org/10.3390/su5020478.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.

Appendices

Appendix 1. Real-world Example Applications of the Best Practices

Here we provide detailed descriptions of the seven best practices for salmon data stewardship, along with practical applications and real-world examples. This is not an exhaustive list, but rather a starting point for salmon biologists and data stewards to implement effective data stewardship practices in their work based on examples from the salmon research and management community.

1. Make Data Governance Explicit to Support Trust and Reuse

Clear governance defines roles, responsibilities, and procedures ensuring data quality, long-term maintenance, accountability, and compliance with community principles such as FAIR and CARE. Effective governance fosters trust, facilitates data sharing, reduces ambiguity regarding decision making, and is critical for coordinating both technical and sociocultural aspects of data stewardship.

In collaborative international or multi-organizational settings, establishing governance at the outset of a project is crucial for aligning diverse groups, including biologists, data managers, Indigenous communities, policymakers, and other participants. Early governance planning should establish clear, collaborative frameworks that respect each group's expertise and needs from the beginning.

Practical Applications:

1.1 Document roles and responsibilities clearly at project start using a Project or Data Product Governance Charter and structured frameworks (e.g., DACI or RACI charts) that relate to organizational data policies.

- Example of a Data Management Plan from the California Department of Water Resources
- Data Management Plan Templates from DMPTool, and NOAA Data Management Handbook

1.2 Integrate CARE principles to ensure ethical governance and respect Indigenous data rights.

- Northwest Indian Fisheries Commission use password protected website to host all the WDFW and tribal data in a one-stop shopping website for co-managers to pull data they need for decision-making process. https://fisheriesservices.nwifc.org/

1.3 Create a governance or oversight committee for regular data practice reviews and decision making regarding data structures, timelines, data sharing agreements and interoperability protocols

- Pacific Salmon Commission has formed a Technical Committee on Data Sharing including both US and Canadian data contributors. https://www.psc.org/membership-lists/

2. Reuse Proven Infrastructure to Save Time and Increase Interoperability

Building custom solutions should be avoided where possible. Maximizing existing platforms and technologies reduces costs, accelerates implementation, and increases data interoperability. Building modular, interoperable systems grounded in proven technologies ensures sustainable long-term stewardship.

Practical Applications:

2.1 Use the Ocean Biodiversity Information System or the Global Biodiversity Information Facility to standardize and host your data

2.2 Use free data catalogue services such as the Knowledge Network for Biocomplexity (KNB) or Zenodo

- The Pacific Salmon Foundation's spawner surveys dataset on Zenodo (Carturan and Peacock 2025) received more views within weeks than a analogous dataset in the Salmon Data Library (Pacific Salmon Foundation 2025) did over several years, illustrating that leveraging established public data infrastructures, rather than developing institution-specific ones, can substantially increase discoverability.

3. Make Data, People, Projects, and Outputs Discoverable, Linked and Citable with PIDs

Persistent identifiers (PIDs), including Digital Object Identifiers (DOI) are essential for tracking the provenance and reuse of data, and linking data, protocols, organizations and people. They allow for consistent referencing, integration across systems, and automated credit via data citation.

Practical Applications:

3.1 Encourage researchers to register for an Open Researcher and Contributor ID (ORCID) and include ORCIDs in metadata records and submission forms

3.2 Register your organization with the Research Organization Registry (ROR) and use ROR IDs to identify institutions involved in salmon science.

- Several salmon data holding institutions are already registered with ROR. As a result, those organizations can track and demonstrate their scholarly impact from data publications: DataCite Commons: Pacific Salmon Foundation

3.3 Assign DOIs to data packages, protocols, and reports using DataCite. Maintain version history for all metadata records and document the provenance of metadata creation, updates, and quality control processes to ensure accountability and traceability.

- The North Pacific Anadromous FIsh Commission (NPAFC) assigns DOIs to IYS-related data packages which are served by a CKAN catalogue at https://data.npafc.org. The Commission also assigns DOIs to NPAFC Technical Reports and Bulletins.

3.4 Embed DOIs in dashboards, figures, and metadata so they persist in derivative products.

4. Use Shared Data Models, Vocabularies and Metadata to Enable Integration

Standardizing metadata and terminology ensures data can be interpreted correctly and integrated across systems. Controlled vocabularies, community ontologies, and structured metadata schemas allow data to retain its full semantic meaning.

Practical Applications:

4.1 Configure data catalogues and metadata intake tools to accept internationally recognized metadata schemas such as ISO 19115, Ecological Metadata Language (EML), or DataCite. Implement automated validation against schema requirements and manual review processes to ensure metadata completeness, accuracy, and consistency before publication.

- The Pacific Salmon Foundation's data portal asks contributors to provide metadata in ISO 19115 or other standard formats. marinedata.psf.ca, ensuring consistent metadata structure
- The NPAFC uses ISO 19115 metadata standard in their data catalogue https://data.npafc.org

4.2 Model datasets and databases using the Darwin Core Standard

- The Hakai Institute Juvenile Salmon Program publishes their data to OBIS using Darwin Core: Hakai Institute Juvenile Salmon Program

- The International Year of the Salmon High Seas Expeditions data mobilization efforts [Johnson and Stap (2024)] published their data to OBIS: https://www.gbif.org/dataset/search?project_id=IYS

4.3 Re-use or publish data terms that are shared online using a persistent identifier in a controlled vocabulary or ontology

- DFO Salmon Ontology...

- State of Alaska Salmon and People…
- Measurement Types in OBIS…
- WDFW has definitions of all hatchery escapement data. Hatchery escapement reports | Washington Department of Fish & Wildlife
- Fish Passage Counts has defined metadata that can be used across OFDW and WDFW. https://www.fpc.org/111_sharedfiles/ColumbiaRiverBasinAdultFishPassageCountsMetadata.pdf

Best Practice 5: Store and Analyze Data in Ways That Others Can Easily Access, Use, and Trust

Making data easily accessible promotes its use in research and management, enabling seamless integration with tools and applications. Ensuring accessible, persistent data storage requires more than just file hosting. Data should be structured, accessible via API, and stored in repositories that support long-term preservation.

Practical Applications:

5.1 Provide Direct Data Access via Application Programming Interfaces (APIs) using tools such as FastAPI, Flask, or Django REST Framework that allows users to access, filter, and retrieve data programmatically, facilitating automation and integration into analytical tools and decision-support systems

- The Pacific States Marine Fisheries Commission make's their PIT Tag Information System data accessible via the PTAGIS API

5.2 Archive data in certified long-term, domain-specific repositories such as the Global Biodiversity Information Facility, the Federated Research Data Repository (FRDR), or NOAA's NCEI, USGS ScienceBase, or EMODnet

- TODO

5.3 Leverage the integration between GitHub and Zenodo to automate archiving and DOI assignment, ensuring long-term data preservation.

6. Incentivize and Track Data Sharing and Reuse

The currency of research lies in recognition—credit, citations, and opportunities for collaboration or co-authorship. Promoting data sharing requires both cultural and technical infrastructure. The cultural infrastructure requires a shift towards viewing data publication as equal in importance to article publication. The infrastructure put in place needs to support the process of generating citation records that give credit to all First Nations, Tribes, agencies, and organizations. By recognizing contributions, tracking reuse, and supporting citation, data stewards can create a system where sharing is rewarded.

Practical Applications:

6.1 License data for reuse using liberal licenses

- All data accessible through the NPAFC data catalogue is licenced as Creative Commons Attribution 4.0 International

6.2 Provide recommended citation text and visible credit fields in metadata

6.3 Create summary dashboards that highlight reuse using COUNTER Code of Practice compliant metrics to track dataset views/downloads and the DataCite APIs

6.4 Ensure that datasets are properly cited in journal articles using in text citations and the recommended citation in the articles list of references, not just in a Data Availability statement

- In late 2024, the NPAFC began citing data sets using in-text citations and the recommended citation in the list of references with the publication of NPAFC Bulletin 7 titled, *Highlights of the 2022 International Year of the Salmon Pan–Pacific Winter Expedition*.

6.5 Promote the view that well documented data publications are primary research outputs and are significant contributions to the field

7. Build Community Through Co-Development and Mutual Benefit

Creating an infrastructure that standardizes and provides cross-border and cross-ecosystem data integration is only effective if there's community engagement. Standards and tools must be co-developed with their intended users using user-centred design principles (citation required) to be effective. Engaging biologists, Indigenous stewards, and data managers ensures relevance, usability, and long-term participation.

Practical Applications:

7.1 Participate in salmon data focused communities such as the Research Data Alliance's Salmon Research and Monitoring Interest Group

7.2 Run participatory workshops for metadata mapping and vocabulary alignment

- American Fisheries Society 2025 WA-BC Chapter Annual Meeting workshop. 'Fishing for Clarity: Knowledge Modelling to Support Cross-organizational Collaboration and Data Sharing about Salmon Escapement

7.3 Support and follow through on Community Engaged Research (e.g. The Salmon Prize Project) that provides tangible value to the communities in which research or monitoring was conducted.

Source: Appendix 1. Real-world Example Applications of the Best Practices

Appendix 2: Training Roadmap for Salmon Biologists Transitioning to Data Stewardship

This roadmap outlines essential topics, resources, and learning materials salmon biologists should engage with to effectively transition into data stewardship roles. The roadmap follows a structured progression similar to roadmap.sh.

1. Foundations of Data Stewardship

- **Principles:**
  - FAIR Data Principles
  - CARE Principles for Indigenous Data Governance
- **Seminal Papers:**
  - Wilkinson et al. 2016 FAIR Guiding Principles
  - Carroll et al. 2019 Indigenous Data Governance
- **Courses and Tutorials:**
  - FAIR Principles Explained (GO-FAIR)

2. Data Management & Governance

- **Seminal Papers and Reports:**
  - Plotkin, 2014 Data Stewardship
  - NOAA, 2007 Environmental Data Management
- **Practical Tools:**
  - Data Management Plan Templates (DMPTool)
  - DACI and RACI Frameworks (Atlassian DACI Guide)

3. Metadata Standards and Ontologies

- **Standards to Master:**
  - ISO 19115 Metadata Standard
  - Ecological Metadata Language (EML)
  - Darwin Core Standard
- **Case Studies & Examples:**
  - Pacific Salmon Foundation Metadata Standards
  - Hakai Institute Juvenile Salmon Program

4. Controlled Vocabularies & Persistent Identifiers (PIDs)

- **PIDs to Implement:**
  - DOIs via DataCite
  - ORCID IDs
  - Research Organization Registry (ROR)
- **Practical Guides:**
  - Introduction to PIDs (DataCite Blog)

5. Data Integration & Interoperability

- **Seminal Papers:**
  - Johnson & Stap, 2024 Salmon Ocean Observing System
  - Bull et al. 2022 Likely Suspects Framework
- **Technical Skills & Tools:**
  - APIs with FastAPI, Flask, Django REST Framework
  - Zenodo-GitHub Integration

6. Data Sharing, Citation & Metrics

- **Best Practices:**
  - Creative Commons Attribution 4.0 International License
- **Tracking & Metrics:**
  - COUNTER Metrics
  - DataCite API

7. Community Engagement & Co-Development

- **Communities & Groups:**
  - Research Data Alliance Salmon Research Group
- **Approaches & Frameworks:**
  - User-centered Design (Interaction Design Foundation)
  - Community Engaged Research (University of Victoria Guide)

Additional Resources

- **Free Courses:**
  - Introduction to Open Science (FOSTER)
  - Research Data Management and Sharing (Coursera)
- **Blogs & Websites:**
  - Open Knowledge Foundation Blog
  - DataONE Data Management Resources

This roadmap serves as a structured guide to equip salmon biologists with the practical and theoretical skills required to excel in data stewardship roles.

Source: Appendix 2: Training Roadmap for Salmon Biologists Transitioning to Data Stewardship

Appendix 3: Getting Started Checklist

Getting Started Checklist for Salmon Data Stewardship

Use this practical checklist to assess how well your project, program, or organization aligns to the seven Best Practices. Start at the Project level, then scale to Program and Organization. Check off items you've completed and note gaps to prioritize.

Tip: For each item, capture a link to the living source (e.g., repository, shared drive, policy page) and the responsible owner.

Practice 1 — Make Data Governance Explicit

Project

```
- [ ] Do you have a Data Management Plan (DMP) covering scope, sensitive data, retention, an
- [ ] Is there a RACI (Responsible, Accountable, Consulted, Informed) table for key tasks? (
- [ ] Are Indigenous knoweledge holders or community members involved in the project?
- [ ] Are Indigenous Data Sovereignty (IDS) requirements identified and documented (who to c
- [ ] Is a data product charter written for each dataset or analysis product with purpose, a
```

Program

```
- [ ] Are DMP and charter templates standardized across projects and stored centrally?
- [ ] Are role definitions for Data Steward, Product Owner, and Maintainer explicit and assi
- [ ] Is this 'community-engaged' research that provides tangible benefit to communities?
- [ ] Are data sharing agreements/MOUs and ethical review pathways documented and reusable?
```

Organization

```
- [ ] Does a governance policy exist that sets minimum requirements for DMPs, RACI, retentio
- [ ] Is there a standing review forum (e.g., monthly data governance check-in) and a regist
```

Evidence to collect: DMP link, data product charter(s), RACI, IDS guidance, sharing agreements registry.

Practice 2 — Reuse Proven Infrastructure

Project

```
- [ ] Have you researched the existing data sharing infrastructure and data storage options s
- [ ] Is your code in version control (e.g., Git) with an issue tracker and releases?
- [ ] Are you using an approved repository or data store rather than creating a new silo? (wh
- [ ] Do you use existing organization authentication/authorization and backup processes?
```

Program

```
- [ ] Is there a preferred stack list (storage, metadata catalog, workflow runner, packaging
- [ ] Do projects consistently deposit finalized data in approved repositories with clear in
```

Organization

```
- [ ] Are enterprise services available and documented (data lake, object store, catalog/por
- [ ] Is there a deprecation pathway for legacy systems and a migration plan for priority da
```

Evidence to collect: repository URLs, infrastructure inventory, intake criteria, backup/DR documentation.

Practice 3 — Use Persistent Identifiers (PIDs) for People, Projects, Data, and Methods

Project

```
- [ ] Do all contributors have ORCID IDs recorded in metadata?
- [ ] Does the project have a resolvable PID (e.g., DOI for a project page or protocol, inte
- [ ] Are datasets assigned DOIs (or other PIDs) at publication, and are versions tracked?
- [ ] Are methods/protocols published and citable (e.g., protocol DOI) and linked from datase
```

Program

```
- [ ] Is there guidance on when to mint PIDs, by whom, and where they resolve?
- [ ] Are projects linked to organizational identifiers (e.g., ROR for institutions) in metad
```

Organization

```
- [ ] Is there a PID policy and a provider/registrar configured (e.g., DataCite) with a docu
- [ ] Are PID linkages automated in the catalog (people  projects  datasets  publications)?
```

Evidence to collect: ORCID list, PID policy, DOI records, resolver links in the catalog.

Practice 4 — Shared Data Models, Vocabularies, and Metadata

Project

```
- [ ] Which metadata profile is used (e.g., ISO 19115, EML)? Is the minimum profile complete
- [ ] Are core entities modeled consistently (stock/population IDs, locations, temporal cove
- [ ] Are controlled vocabularies/code lists applied for key fields (e.g., species codes, ge
- [ ] Is a data dictionary included with definitions, units, allowed values, and provenance
```

Program

```
- [ ] Do projects use a shared schema and code lists across datasets to enable easy joins?
- [ ] Are validation checks in CI (schema validation, vocab checks) standardized across repo
```

Organization

```
- [ ] Is there an endorsed salmon domain profile and shared code lists with owners and change
- [ ] Are mappings to external standards maintained (e.g., taxonomic, geospatial, hydrologic
```

Evidence to collect: metadata profiles, data dictionary, code lists, schema validators, mapping documentation.

Practice 5 — Store and Analyze Data for Easy Access, Use, and Trust

Project

```
- [ ] Is raw data immutable and separated from processed/analysis outputs?
- [ ] Is there a fully reproducible workflow (scripts/notebooks + environment + parameters)
- [ ] Is the computational environment captured (lockfile/conda env, container image) and ve
- [ ] Are QA/QC checks automated with logs and thresholds documented?
- [ ] Are access controls and sensitive data handling documented and implemented?
```

Program

```
- [ ] Do projects follow a common repo layout and release process (tags, changelog, signed a
- [ ] Are standard storage classes, lifecycle policies, and archival rules applied?
```

Organization

```
- [ ] Are security, backup/retention, and audit requirements defined and routinely verified?
- [ ] Is there a trusted long-term archive with fixity checking and preservation metadata?
```

Evidence to collect: workflow definition, environment files, container references, QA/QC reports, storage/backup settings.

Practice 6 — Incentivize and Track Sharing and Reuse

Project

```
- [ ] Is a clear citation and license statement included in metadata and README?
- [ ] Are reuse metrics collected (downloads, citations, API hits) and reviewed?
- [ ] Do release notes document what changed and implications for users?
```

Program

```
- [ ] Are common metrics dashboards available for priority datasets and updated automatically
- [ ] Are data citations tracked in assessments, reports, and staff evaluations?
```

Organization

```
- [ ] Do policies require citation guidance and permissive, appropriate licensing where possi
- [ ] Are automated reports of reuse (e.g., via DOI provider APIs) delivered to product owne
```

Evidence to collect: LICENSE, CITATION, reuse dashboard link, policy excerpts, sample citations in reports.

Practice 7 — Build Community Through Co-Development and Mutual Benefit

Project

```
- [ ] Are stakeholders identified, including First Nations/Tribes/Indigenous partners, and en
- [ ] Have you held at least one co-design session to validate user needs and success criteri
- [ ] Is there an open feedback channel (issues form, contact) and a published roadmap?
```

Program

```
- [ ] Do cross-project working groups exist for models, vocabularies, and tooling with regula
- [ ] Are community contributions recognized (authorship, acknowledgements, meeting time, fun
```

Organization

```
- [ ] Is there an endorsed governance body or community of practice with decision records?
- [ ] Are procurement/funding mechanisms available to support shared components and Indigenou
```

Evidence to collect: stakeholder map, engagement records, roadmap, working group notes, decision log.

Quick Start: 30/60/90-Day Plan

- First 30 days
  - ☐ Create/standardize DMP + RACI; draft data product charters for top 1–2 datasets.
  - ☐ Move code to version control; document repo structure; capture environment file.
  - ☐ Choose metadata profile and draft a minimal data dictionary; list code lists in use.
- By 60 days
  - ☐ Mint/plan PIDs (project page/protocols), add ORCIDs to metadata, prepare DOI for first dataset.
  - ☐ Add schema + vocab validation to CI; separate raw/processed; automate QA/QC checks.
  - ☐ Stand up a reuse dashboard or basic metrics capture; add citation/license to README and metadata.

- By 90 days
  - □ Publish first governed release to approved repository with DOI and complete metadata.
  - □ Formalize cross-project working group and change control for vocabularies.
  - □ Document archival/retention path and verify backups; schedule governance reviews.

Minimal Artifacts Checklist (Project Level)

□ DMP (with IDS considerations) and RACI
□ Data product charter(s) for priority dataset(s)
□ Versioned repository with releases and changelog
□ Metadata profile file + data dictionary + code lists
□ Reproducible workflow + environment file or container
□ QA/QC checks and results log
□ Citation and license statements
□ Plan for PIDs (ORCID list, dataset/protocol DOIs)
□ Evidence of stakeholder engagement and roadmap

Maintain this list as a living issue in your repository and review quarterly.

Source: Appendix 3: Getting Started Checklist