DRAFT-Best practices and practical applications for salmon data mobilization—a guide for salmon biologists assuming responsibility for data stewardship-DRAFT

Brett Johnson Scott Akenhead Jennifer Bayer Tomas Bird
Samuel Cimino Graeme Diack Lara Erikson Nancy Leonard
Catherine Michielsens Fiona Martens Emily Lescak
Gottfried Pestal Matt Jones Shirly Stephen-Ambrose

Keywords: Salmon data stewardship, Data interoperability, FAIR principles, Persistent identifiers (PIDs), Controlled vocabularies, Metadata standards, Application Programming Interface (API), Data citation, Ontology development

Abstract

Salmon research, management and conservation generates increasingly vast and diverse data crucial for effective decision making in resource management. Yet, these resources remain largely fragmented across jurisdictions, disciplines and outdated infrastructures, limiting their use in responsive fisheries management. Biologists are increasingly assuming the responsibilities of data stewards to address these challenges, yet often lack clear guidance or institutional support to do so. To address this, we distill seven best practices for salmon data stewardship and demonstrate their application through a case study. We provide practical guidance for those transitioning into these essential stewardship roles, outlining real world examples, tools, and templates specific to the salmon research and management domain. We argue that effective salmon management hinges upon formally establishing data stewardship as a dedicated, institutionally supported professional role. We outline key best practices including both sociocultural and technical solutions that collectively ensure salmon data meet modern open science principles and respect Indigenous Data Sovereignty. Through an illustrative case study involving sockeye salmon productivity analyses across Pacific Coast jurisdictions, we highlight how clearly defined stewardship practices can enhance data reproducibility, integration, and management efficacy. With a foundation of shared best practices, salmon data stewards will enable

faster, more transparent decisions that draw from broader, cross-jurisdictional datasets, and support development of tools that leverage recent advances in artificial intelligence—ultimately strengthening the management and conservation of salmon populations and the ecosystems upon which they depend.

Introduction

Salmon biologists generate vast amounts of data on abundance, health, and environmental conditions, yet these data remain fragmented, inconsistently measured, and often incomplete across time, space, and life history stages—limiting their value for robust research, hypothesis evaluation, and management decision-making (Marmorek et al. 2011; Inman et al. 2021; Diack et al. 2024). Salmon traverse multiple ecological regions and jurisdictional boundaries, resulting in data collections managed by diverse agencies and institutions, often in isolation. This fragmented data landscape undermines timely, integrated analyses necessary for effective management and conservation decisions. Additionally, the lack of institutional support and dedicated roles for data management frequently relegates critical data stewardship tasks to an ad hoc status—something performed off the side of a biologist's desk. Institutional neglect of formal data stewardship has become a bottleneck in adaptive salmon management and conservation efforts.

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). Yet salmon biologists transitioning into data stewardship roles typically receive insufficient guidance or institutional support. We argue that fisheries management agencies must formally acknowledge and fund dedicated data stewardship roles to effectively mobilize and leverage salmon data. Without this institutional commitment, data remain inaccessible and fragmented, severely constraining the responsiveness and adaptability of management actions. To address these critical shortcomings, we outline practical steps that biologists, agencies and organizations can adopt.

The Issue

Effective integration and mobilization of salmon data mirrors the complexity of salmon biology itself: these fish traverse freshwater, estuarine, and marine ecosystems, crossing provincial, state, tribal, federal, and international management boundaries (Groot and Margolis 1991). 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 rare and prohibitively expensive for all but the most pressing challenges. 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. As a result, long-term datasets critical to stock assessment

and environmental monitoring frequently become inaccessible, poorly understood, or effectively lost once original data holders retire or move on.

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. Modernizing these systems will require coordinated investment, grounded in shared international data standards and stewardship practices that accommodate the full disciplinary and geographic diversity of salmon science.

The consequences of inaction are already visible. When critical datasets are hard to find, access, or interpret, biologists and analysts lose valuable time trying to reconstruct or harmonize them. This reduces transparency, increases the risk of errors, and delays urgent conservation or management responses. Without clear accountability for data stewardship, the system continues to rely on improvised, inconsistent, and ultimately unsustainable practices.

The Need for Coorindated Action

For fisheries managers, modernizing data systems and workflows is essential to improve the quality, speed, and interoperability of operational data assets. These systems must support an increasingly complex decision-making landscape that now depends on integrating broader types and sources of data, often in real time. At the same time, researchers face pressure to generate insights on future salmon abundance, the impacts of changing environmental conditions, and the effectiveness of restoration strategies across all salmon life stages. Yet the current scattered and siloed data landscape remains unfit for purpose—both for science and for management.

Despite operating under different mandates, 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 Authority, 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 socio-cultural importance of salmon to the Indigenous communities of the Northern Pacific and Trans-Atlantic (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.

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 (Johnson and Stap 2024). 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.

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 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.

Specifically, 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 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.

A user-centered design approach to salmon data stewardship is critical and focuses on creating tools that align with biologists' needs. Data stewards play a critical role as business analysts, bridging the gap between biologists and IT by translating data needs into application or data system features. 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 1) 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 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.
- **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.

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 gray 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 5: Store Data in Ways That Others Can Easily Access and Use** 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.
- **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.

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.
  Practical application: Use a DACI or RACI framework to 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, 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 outlined above, agencies and researchers can transform legacy datasets into living resources, enabling reproducibility, easing collaboration, and accelerating insight across the salmon research and management community.

Table 1: Best practices and practical applications of salmon data stewardship

Best Practice	**Summary**	**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 Indigenous data rights. - Form a governance or oversight committee to review data structures, timelines, and agreements.
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, Vocabularies 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. - Model datasets using the Darwin Core Data Package. - Use controlled vocabularies or ontologies with PIDs.

Best Practice	Summary	Practical Applications
5. Store Data in Ways That Others Can Easily Access and Use	Structured and accessible data formats ensure usability, reduce wrangling, and support integration with analytical tools and applications.	- Provide APIs using FastAPI, Flask, or Django REST. - Archive in trusted repositories (e.g., GBIF, FRDR, USGS). - Use GitHub-Zenodo for DOI assignment and preservation.
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. 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—leading to missed opportunities for synthesis, insight, and action. This fragmentation hampers our ability to understand the drivers of change across 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 gov-

ernance, 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.

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 (**roche2020Roche?**), 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 best 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.

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.

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

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/.

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.

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.

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, and 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

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

-

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

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

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

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.

3.3 Assign DOIs to data packages, protocols, and reports using DataCite.

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.

- 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

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 moblization efforts (Johnson and Stap 2024) published their data to OBIS: https://www.gbif.org/dataset/search?project_id=I

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

- State of Alaska Salmon and People…
- Measurement Types in OBIS…

5. Store Data in Ways That Others Can Easily Access and Use

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. 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

- 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

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 focussed communities such as theResearch Data Alliance's Salmon Research and Monitoring Interest Group

7.2 Run participatory workshops for metadata mapping and vocabulary alignment

7.3 Support and follow through on Community Engaged Research (citation required) that provides tangible value to the communities in which research or monitoring was conducted