# Qualark Species Composition Reproducible Data Product Governance Charter

Title: Qualark Species Composition Data System

Date Initiated: 2025-04-02

## 1. Purpose

To recreate the Qualark Species Composition Database that, Pacific Salmon Commission (PSC) Staff are prototyping in the PSC’s Azure environment, in the DFO cloud environment to fulfill a clear operational need using transparent, automated, and interoperable processes. The database aims to consolidate counts, recorded sonar lengths, test fishing catch and biological data, and the results of the Qualark species composition model into a single, accessible platform. This product will serve as a flagship case study, demonstrating end-to-end stewardship from data ingestion through public dissemination and support Fraser Interior Area Staff.

## 2. Problem Statement

* Data scattered in multiple Excel files & text exports (DIDSON/ARIS).
* Manual data entry → prone to errors, delays.
* Limited QA/QC checks before data is used for analysis.
* Data not centralized → difficulty consolidating for in in-season or post-season analysis.

## 3. Objectives

* Streamline data processing and storage, improve data management efficiency, enhance data analysis capabilities in order to support species composition assessments.
* **Development of a Centralized Database:** Create a centralized database for Qualark hydroacoustic and test fishing data, consolidating data from multiple sources into one unified repository.
* **Integration with R:** Enable seamless integration with R for data querying, analysis, and visualization.
* **Data Backup and Recovery Mechanisms:** Establish robust data backup procedures and recovery mechanisms to prevent data loss and ensure data availability.
* **Data Management Workflow:** Define workflows where administrators manage data uploads, receive updates from crew members through Excel files. This maintains familiar processes, ensuring smooth collaboration and minimal disruptions.
* **Potential for Future Integration:** Design the system to allow for future integration via APIs or with other databases.
* Use this case to **formalize governance structures**—assigning roles and decision authority for data formatting, pipelines, and publication.

## 4. Deliverables

* A working, documented ETL pipeline that implements QA/QC and serves the data in a SQL database
* A published code repository (e.g., GitHub or Azure Repos)
* Governance documentation: Data Owner, Trustee, Steward roles and escalation paths
* Data output from the SQL database that can be shared with the Pacific Salmon Commission staff with daily updates.

## 5. Team Roles

An important part of the long term success of a data product relates to the roles and responsibilities related to governance of the data product. The roles outline below reflect DFO’s Data Stewardship Initiative. Data roles and responsibilities should be defined and assigned for all data assets within the Department to ensure that departmental data is managed as a strategic asset.

| Role | Name | Responsibilities |
| --- | --- | --- |
| Task Team Leads | Brett Johnson, Eric Taylor, Catarina Wor (Brooke Davis) | Oversees project delivery, coordinates cross organizational collaboration and support |
| Data Trustee | Scott Decker or Kaitlynn Dionne | [Data Trustees](https://intranet.ent.dfo-mpo.ca/dfo/sites/dfo-mpo/files/en_-_dfo_data_trustee_persona_0.pdf) ensure the strategic management of assigned data assets as well as compliance with departmental and enterprise data-related strategies, regulations, policies, directives and standards. Chairs the Data Product Governance Team. Usually a section head or higher. Coordinates with Salmon Data Domain Governance Team and the Regional Data Committee. |
| Data Steward | Kaitlynn Dionne (Lead DS)  Zhipeng Wu (DFO DSU, Initial development and transfer)  Albury C (DFO DSU, Initial development and transfer) | [Data Stewards](https://intranet.ent.dfo-mpo.ca/dfo/sites/dfo-mpo/files/en_-_dfo_data_steward_persona_0.pdf) maximize the quality and reusability of their assigned data asset by enforcing data management business rules. Owns the product lifecycle and ensures documentation. |
| Data Custodians | Brian Smith (DFO, Long term maintenance)  Sai Chandra (Initial development and prototyping with PSC)  Daniel Doutaz (To troubleshoot pipelines in the office) | [Data Custodians](https://intranet.ent.dfo-mpo.ca/dfo/sites/dfo-mpo/files/en_-_dfo_data_custodian_persona_0.pdf) ensure the safe custody and integrity of hosted data, and safeguarding data repositories, including the design and implementation of technical solutions. |
| Data Contributors | Michael Gauthier | [Data Contributors](https://intranet.ent.dfo-mpo.ca/dfo/sites/dfo-mpo/files/en_-_dfo_data_contributor_persona_0.pdf) ensure that the data they provide to the Department (including data sourced from third-parties) aligns with all technical and business policies, procedures and standards, including those defined by data stewards. |
| Data Consumers |  | [Data Consumers](https://intranet.ent.dfo-mpo.ca/dfo/sites/dfo-mpo/files/en_-_dfo_data_consumer_personas.pdf) ensure that their usage of data supports departmental and government objectives and mandate, including communicating regularly with data stewards on their data needs. |

## 7. Timeline (Proposed)

| Phase | Duration | Milestone |
| --- | --- | --- |
| Kickoff and Scoping | 2 weeks (June 5th- June 15th) | Charter signed off, roles confirmed. |
| DFO Infrastructure Setup | 4 weeks (June 15th- July 15th) | DFO SQL Database Provisioned. |
| Replication of PSC Architecture in DFO Environment in Parallel | 8 weeks (July 15th - September 15th) | First reproducible version created |
| Governance, Maintenance and Standards Review | Concurrent. Now - August 1st | Infrastructure, governance, sustainability plan and roles approved by Trustee |
| User Testing and Refinement | 4-6 Weeks (Sept 15th-Nov 15th) | Final version ready for 2026 season |

## 8. Success Criteria

* Output is reproducible end-to-end using only the shared codebase and cloud resources
* Clear governance structure and sustainability plan established and documented
* Data Sharing Agreement Set up with Pacific Salmon Commission staff and they are able to access outputs in near real time

## 9. Dependencies

## 10. Risks & Mitigation

| Risk | Mitigation Strategy |
| --- | --- |
| Staff bandwidth | Scope narrowly and provide flexible contribution paths |
| Data access delays | Formalize data sharing agreements early |
| Cloud complexity |  |
| Governance ambiguity | Escalate unclear decisions to RDC or OCDS early |

## 11. Contact

Reach out to **Brett Johnson**, Data Stewardship Unit: [Brett.Johnson@dfo-mpo.gc.ca](mailto:Brett.Johnson@dfo-mpo.gc.ca)