



Image Data Work Group Meeting #3

June 5, 2026 | Meeting Summary

The Image Data Work Group is one of two Work Groups supporting ROSA's Data Governance Program. The Image Data Work Group is focused on all image data (video and still) from any marine source.

Work Group Goals

- Where should ROSA recommend people publish fisheries-related data?
- How should data be organized to maximize discovery and reusability?
- What else can ROSA do to support data governance for this data type?

Potential topics for recommendations

- Documentation (including metadata)
- Metadata standards and recommended repositories
- How to organize and structure datasets within a 'project'
- Any other specific info to be included (i.e. special fields, tags)
- Catalog of experimental designs used
- How to be good data managers for our future selves
- How to preserve and publish data to answer regional or cumulative impact questions

Meeting Summary

After a brief recap of the group's earlier findings, including rapid evolution of analytical methods, the high and often unfunded cost of file storage, and a lack of standardized file-level metadata across institutions, the group spent the bulk of the meeting on a roundtable discussion of the current image data landscape. Participants shared what tools, storage strategies, and workflows they're using across organizations, revealing a wide range of approaches: some rely on institutional cloud storage covered by overhead, others cycle data to cheaper storage tiers over time or offload from the cloud periodically to manage costs, and some are constrained by funders who don't budget for long-term storage and management. Several annotation platforms were discussed and compared (including CVAT, VIAME, Roboflow, and Tator), along with emerging options like Squidle+, a collaborative platform for managing, discovering, and annotating marine imagery. There was general agreement that annotation tools are plentiful, but most groups are running them to feed AI/ML models rather than to publish or share underlying images or annotations.

The group also discussed regional efforts already underway to organize and expose image-derived data, including RWSC's work linking data across regional ocean partnerships (NROC, MARCO) and its build-out of a seafloor data repository on DataONE, as well as federally-driven efforts like the Marine Geoscience Data System (MGDS), to which many federally funded programs must submit image data. Participants raised the idea of scripting connections to existing APIs (MGDS, USGS) to support discovery without requiring data transfer or duplication, and noted that even limited, targeted exposure of a dataset via an API can meaningfully support discoverability. Data licensing, institutional risk, and the practical burden of standardizing older datasets remained open concerns, particularly for offshore wind developers managing data-sharing obligations across multiple partners and projects.

The remainder of the meeting turned to the future of the group itself. Participants discussed narrowing focus going forward, potentially through sub-groups organized around individual image data types (e.g., a BRUV-specific data standard), and prioritizing file-level over field-level metadata. Using large language models to help standardize metadata, both within and across organizations, was raised as a promising area to explore, along with continuing to evaluate tools like [Squidle+](#).

Next steps

- Continue discussions on publishing image metadata in future WG meetings
- Develop possible timelines, work tasks, and support needs for the group's next phase
- Identify and reach out to other ongoing, complementary image data efforts to explore collaboration and coordination (e.g., Northeast and Mid-Atlantic Ocean Data Portal teams)
- Explore reviving relevant ideas from the 2025 Oceanic/NLR environmental data sharing meeting
- Catalogue existing best practices and workflows for cloud-based image data storage and organization, including identification of barriers to image data sharing
- Explore use of TACC's offer of free cloud storage (5TB, with up to 1PB potentially available), which will require an experienced data manager to take advantage of