

DATA MANAGEMENT IN AQUACULTURE: ABIDING BY THE US FEDERAL MANDATES

Amanda A. Hyman*, Cole Ensminger, Stephanie Ichien, and Hillary Egna

The Feed the Future Innovation Lab for Collaborative Research on Aquaculture & Fisheries
Oregon State University, Corvallis, Oregon, USA | aquafish@oregonstate.edu | aquafish.oregonstate.edu

Abstract

Data and information are valuable national and global resources. Although data sharing has increased over the last decade partly because of government and/or peer-reviewed journal mandates, many scientists never publish their datasets to publicly available repositories and, furthermore, of those datasets that are published, many are incomplete or unusable. As AquaFish Innovation Lab (AquaFish) begins to publish its data in compliance with the U.S. mandate for increased access to federally funded scientific research, some of our greatest hurdles and lessons revolve around data management. As an interdisciplinary program working in 33 countries in Africa, Asia, and the Americas with nearly 300 partner institutions since 2006, AquaFish presents as a unique opportunity to explore the challenges of managing legacy datasets in international development along with the importance and potential impacts of successfully managing such data (e.g., augmented collaboration, decreased administrative costs, increased technology transfer to local communities).

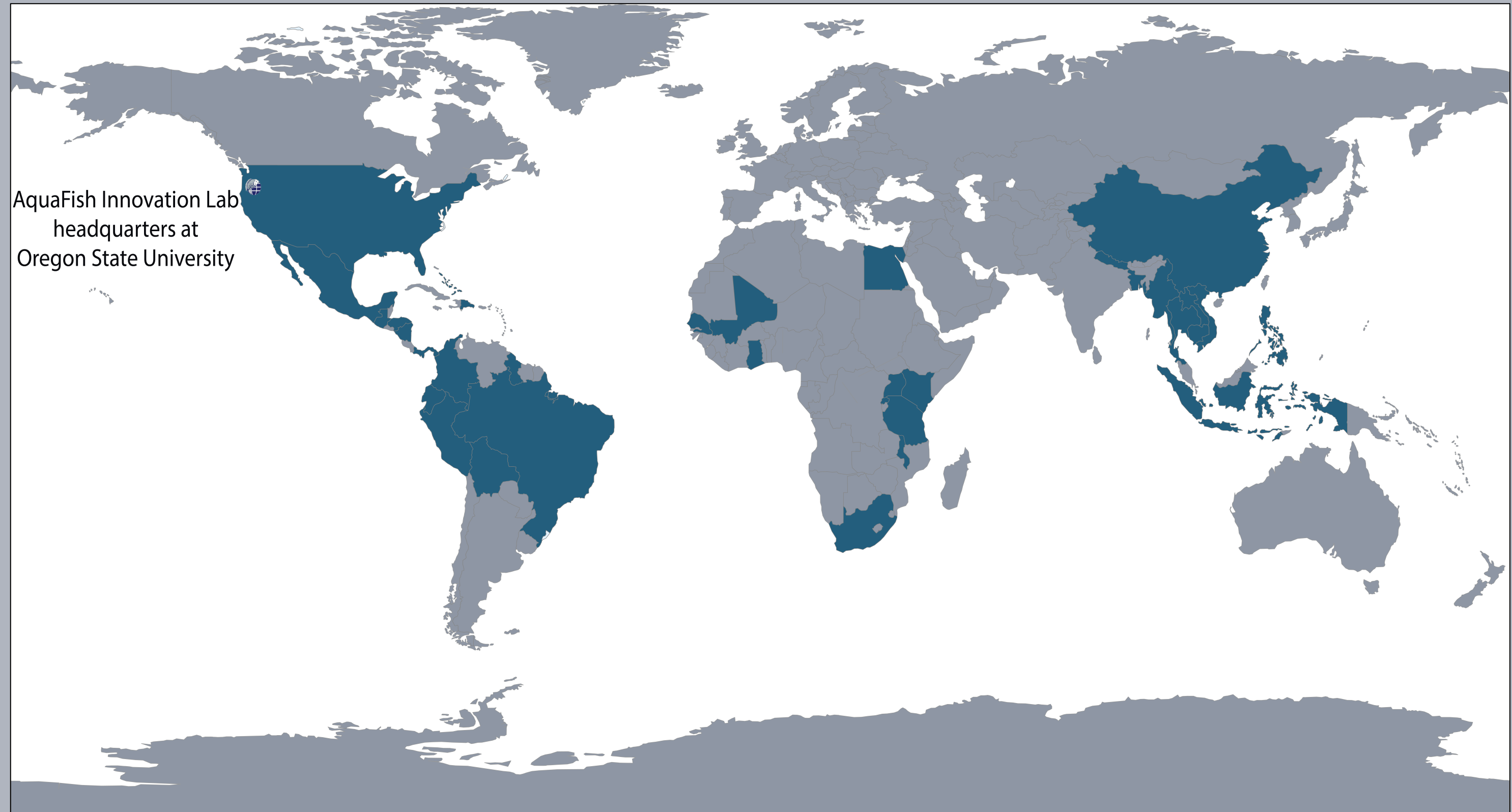


Figure 1. From 2006-2018, AquaFish has worked in 33 countries colored in blue.

What is Open Data?

According to the OMB Open Data Policy (M-13-13), “...the term “open data” refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users. In general, open data will be consistent with the following principles: public, accessible, described, reusable, complete, timely, managed post-release...”

Why Publish Your Data?

- To comply with government mandates and/or peer-reviewed journal requirements
- To collaborate with other universities, institutions, and agencies
- To increase the opportunity for reproducibility of a study
- To enable and to conduct longitudinal studies
- To augment the longevity of data by establishing an archive that will remain beyond funding and grant periods
- To promote interdisciplinary work by creating a standard set of metadata and increasing the exposure of the dataset to scientists who may not otherwise come across the data and project
- To improve the utility of existing data, especially in a field with limited and/or decreasing funding opportunities
- To reduce costs
 - According to T.C. Redman (2013) “...knowledge workers waste up to 50% of their time hunting for data, identifying and correcting errors, and seeking confirming sources for data they do not trust.”
- To enhance review of the data and, therefore, possibly increase quality
- To more easily transfer skills and knowledge to a wider audience, particularly outside the academic sphere

Source: Redman, TC. 2013. Data's credibility problem. Harvard Business Review.

AquaFish as a Case Study

Federal Mandates

AquaFish is partially funded by the United States Agency for International Development (USAID) under a cooperative agreement and, hence, must comply with federal mandates and agency regulations on data. According to USAID Development Data (ADS 579) Policy “*USAID staff, as well as contractors and recipients of USAID assistance awards (e.g. grants and cooperative agreements), must submit any Dataset created or collected with USAID funding to the Development Data Library (DDL) in accordance with the terms and conditions of their awards. This is in keeping with Executive Order 13642 (Making Open and Machine Readable the New Default for Government Information) and the OMB Open Data Policy (M-13-13).*”

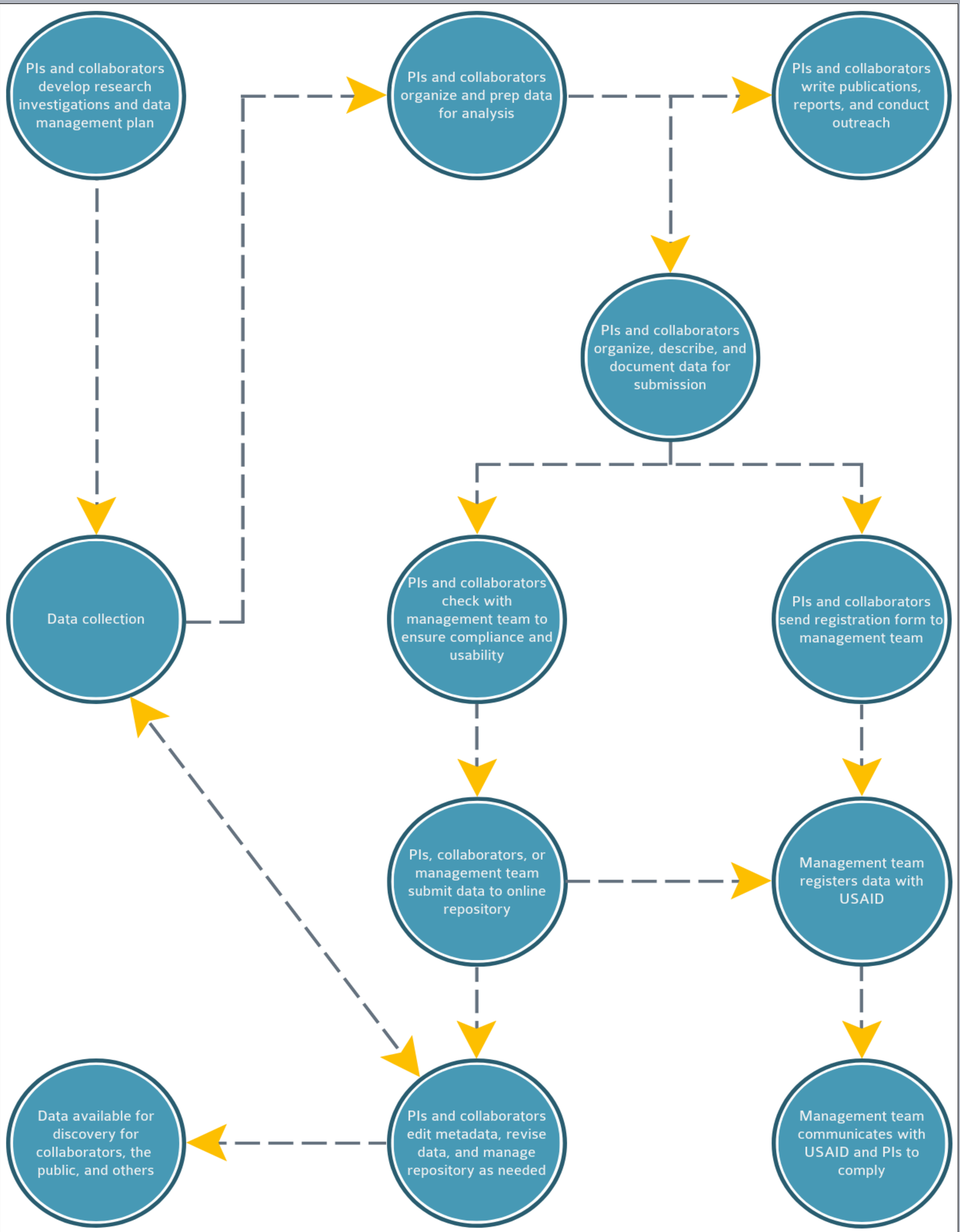


Figure 2. Flow of data and information at AquaFish with Principal Investigators (PIs), USAID, and the AquaFish management team.

Towards Meeting the Mandate

- Learn best practices and options for various forms of data from university librarians at the beginning of the process and check in with them as data are entered
- Attend workshops and seminars by USAID, Oregon State University, and other similar institutions
- Research relevant and reputable repositories
 - For AquaFish, Harvard's Dataverse was selected to accommodate all datasets.
- Persistence with Principal Investigators and collaborators for compliance
- Improved Data Management Plan for utility of Principal Investigators and collaborators

Challenges to Comply

- Unfunded mandate given on short notice
 - USAID mandate occurred retroactively without any additional funds or time allotted and USAID implemented a mandatory structure for data management plans, which was untenable for the planning process in the field of aquaculture.
- Unable to repurpose collected information from a variety of stakeholders
 - Collaborators submit reports, presentations, research tools, metadata, and other such documents, which were frequently created for stakeholders other than the US government, in languages and characters other than English (Figure 3, for example).
- Many surveys and data collection forms are non-digital data and host-country collaborators do not always have the means to digitize data efficiently or effectively.
- Data collected before the mandate are not easily recaptured for the new format.
 - Throughout the retroactive and current time periods, AquaFish has had several collaborators, working in different countries and varying levels of organization, data quality control, and retention procedures prior to the USAID mandate.
 - AquaFish attempts to engage undergraduate and graduate students. When these students graduate, frequently they do not leave their data in usable formats for other scientists.
- In a few AquaFish host countries, records have been lost or destroyed because of such things as disasters and political unrest.
- Changing and ambiguous forms, guidelines, and procedures from USAID make compliance difficult.

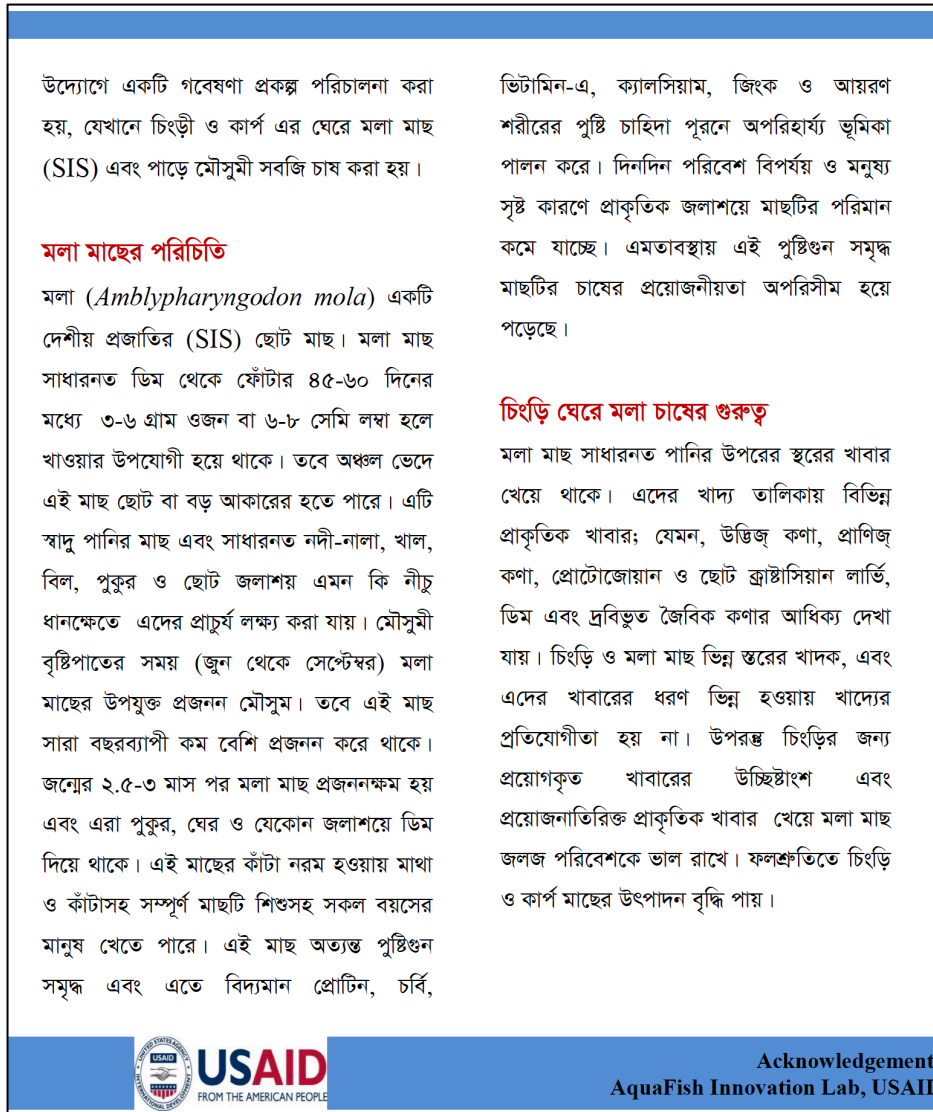


Figure 3. Example of outreach document created for local audiences in characters other than English.

Acknowledgements

The Feed the Future Innovation Lab for Collaborative Research on Aquaculture & Fisheries (AquaFish Innovation Lab) is funded under USAID Leader with Associates Cooperative Agreement No. EPP-A-00-06-00012-00 and by the participating US and Host Country partners. This work was made possible by the generous support of the American people through USAID. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

All photos and illustrations courtesy of AquaFish Innovation Lab.

We thank AquaFish Innovation Lab partners and collaborators for their contributions and their efforts.