MARIS: 20 Years and Counting Developing Data Standards

Presented at Symposium on
Developing a National Fisheries Data Exchange Standard
August 18, 2014
144th Annual Meeting of the American Fisheries Society
Quebec, Canada

Andrew J. Loftus
Nancy Nate
Aloftus@andrewloftus.com

Multistate Aquatic Resources Information System (MARIS)

- A system for sharing information from multiple data providers through a common query portal.
- Initially started as a distributed system and ideally would evolve back into this.
- Development of MARIS began with identifying common variables and standardizing their representation in MARIS,

First Step: Identify Reasons for Development

- Need for assessing status and trends of fish populations over larger geographic regions.
- Desire to tap cumulative wealth of state fisheries data collection efforts.
- Desire by some state agencies to alleviate multiple requests for data.

Development Philosophy

- "Conservative inclusiveness" – include variables that facilitated diverse datasets to be added without burdening MARIS with too many extraneous variables.
- Don't require changes to state data providers' internal information systems.
- Initially focused on variables that could be used for quantitative assessment of status and trends.

Standardized Mandatory Data Categories

- Data Provider (state, agency)
- Water ID, Station ID, or other agency joining variable
- Water Type (e.g., lake, stream, reservoir)
- Sample date (begin and end)
- Species code (state specific)
- One or more variables related to quantity collected (e.g., count, population estimate, CPUE, etc.)

- Note: Should require a “date of data export” for each database.

Optional Data Categories

Main tables

- ~15 location variables (spatial coordinates, accuracy, NHD etc.)
- ~ 6 collection variables (gear, method description, etc.)
- ~13 catch variables (catch, CPUE, PE, Biomass, etc.)
- ~4 effort variables (time, area, units of measure)

Secondary Tables

- ~21 Water characteristics/quality variables (that aid in interpreting collection event)
- ~13 size and age variables

Automatically Added by MARIS

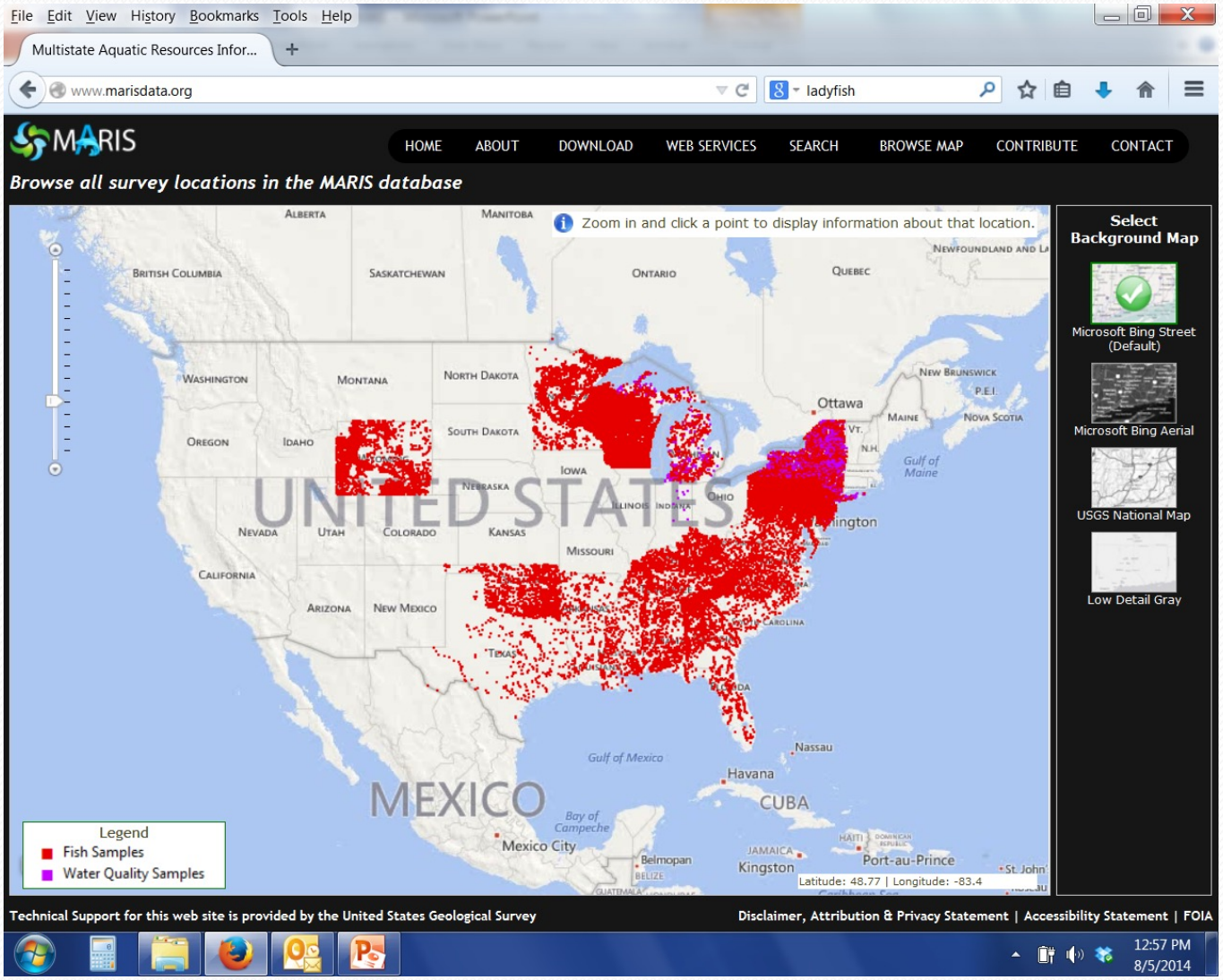
- HUC 8, 10, 12 if lat/lon is included
- ITIS code
- Consistent scientific and common names maintained by MARIS master species table.
- Dataset ID, Originator Name, dataset Name.

Current Content

# States/Data sets	# Species Occurrence	Date Range	Species & Species Groups	# Individual Lakes/ Reservoirs	# Individual Streams
24/41	1.2 million	1916-2014	1,163	10,753	22,316

**Most lake and stream locations have multiple survey dates and/or locations.

Current Content



Key Lessons: Development

1. Decide on the primary purpose of data sharing (and end users) first.
2. Prioritize questions to be asked of the data.
3. Decide whether you are developing standards for a newly implemented collection program or for historic data or both.
4. Be flexible if working with multiple organizations.
5. Beware (but don't discount) mission creep. Too much or too little flexibility can unravel the mission.
6. Build from what has been developed previously.
7. Be patient and persistent.

Key Messages: Technical

- Descriptors of sampling methods should be incorporated and standardized into table structures rather than meta-data.
- Descriptions of location accuracy are critical.
- A fully automated data exchange system should include file conversion to a standard.
- Include a check for standardization before uploading;
  - prevent file duplication;
  - back-up and log uploads with date stamp.

Questions?

MARIS: 20 Years and Counting Developing Data Standards

Andrew J. Loftus
Nancy Nate
Aloftus@andrewloftus.com