Standardization, sharing, and ownership: key areas of concern for Great Lakes database integration

IAGLR
May 25-29, 2015
Burlington, Vermont

Norine E. Dobiesz
Workshop process

• Roundtable format
  – Focus on Great Lakes data integration issues
  – Participation from Lake Victoria and the Laurentian Great Lakes

• Created a list of obstacles
  – Challenges, issues, roadblocks, concerns

• Group into areas of concern
  – Which areas present the “biggest” obstacle?
Areas of concern

• Data
  – Types of data
  – Standardization
  – Data gaps
  – Data sharing
  – Data security
  – Technical issue

• Users
  – Usability

• Support
  – Ownership
  – Funding
Type of Data

• What types of data are needed?
  – Who determines this?

• Raw vs. summarized

• Spatial dimensions
  – Nearshore, offshore, etc.

• Temporal dimensions

• Biological and environmental
Standardization

• Address project or jurisdiction differences in data
• Need to standardize data on units and scales
  – Transparency (don't worry who collected what)
  – Interoperability (all data is comparable)
  – Reduces need and time to harmonize data later
  – Lack of this causes long lag time to get data to management
• Process for submitting data
Data Gaps

- Missing data
- Spatial issues
- Data not in database
  - Because it's not contributed (no mandate)
  - Because of lack of funding
Data Sharing

- Rapid access to new data
- Lack of sharing protocols
  - Data integration (at what level)
  - Technical problems can impact sharing
- What do agencies choose to share
- Not all project data shared (champion or mandate)
Data Security

• Can data be secured?
• Current technical limitations of securing data (separate from sharing)

“Somebody broke into your computer, but it looks like the work of an inexperienced hacker.”
Technical Issues

- Handling large databases and associated issues
  - Availability
  - Backup
- Methods for updating
  - Use of web services
  - Other methods of updating
  - Synching data
Usability

- Need user access to the data
- What services will be available to access and visualize data
- How easy is it to use
Ownership

• Lack of mandates
  – Impacts other concerns such as sharing and data gaps

• Who will do it?
  – Needs a curator or champion
  – Long term vision and goals
  – Manage large teams
Funding

• Lack of funding impacts data collection

• Ongoing funding for future updates and maintenance

• Long-term viability of the database and/or web site
# Examples of obstacles and ranking

<table>
<thead>
<tr>
<th>Item</th>
<th>Usability</th>
<th>Standardization</th>
<th>Funding</th>
<th>Security</th>
<th>Sharing</th>
<th>Tech</th>
<th>Type of data</th>
<th>DataGaps</th>
<th>Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>standardized data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no data sharing protocols</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>science to management lag time</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lack of nearshore data and lower trophic data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>all existing databases don't have web services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>end users in mind</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transparency</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>who curates data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

4 10 3 2 8 2 4 6 7
## Examples of obstacles and ranking

<table>
<thead>
<tr>
<th>Item</th>
<th>Usability</th>
<th>Standardization</th>
<th>Funding</th>
<th>Security</th>
<th>Sharing</th>
<th>Tech</th>
<th>Type of data</th>
<th>DataGaps</th>
<th>Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>standardized data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no data sharing protocols</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>science to management lag time</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lack of nearshore data and lower trophic data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all existing databases don't have web services</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>end users in mind</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>....etc....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transparency</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>who curates data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

|               | 4 | 10 | 3 | 2 | 8 | 2 | 4 | 6 | 7 |

- Transparency: X
- Who curates data: X
Results and more questions

• Most obstacles impact
  – Standardization, data sharing, ownership
• Data obstacles are about “what”
  – What data is collected and stored
  – What standards are used
  – What type of access is granted
• User obstacles are about “how”
  – How will users access the data
• Support obstacles are about “who”
  – Who will support this in the long term
  • And “how”?
Conclusions …. so far

• Big lakes need big data and that causes many challenges

• The most challenges identified are standardization, data sharing, and ownership

• More groups are working towards addressing these issues
  – But often not globally

• Needs political and economic “will” to support large integrated databases for the long run
Acknowledgements

• **East Africa**
  – Oliva Mkumbo, Lake Victoria Fisheries Organization
  – Anthony Munyaho, NFRR
  – Lewis Sitoki, Technical University of Kenya

• **Laurentian Great Lakes**
  – Catherine Riseng and Danielle Forsyth, GLAHF
  – Ted Treska, GLFC
  – Jennifer Read and Tad Slawecki, GLOS
  – Lizhiu Wang and Mark Burrows, IJC
  – Cynthia Hagley, Minnesota SeaGrant
  – Jim Bence, Michigan State University
  – Brent Lofgren and Heather Lucier, NOAA
  – Bill Taylor, Sea Grant Associate Director
  – David Allan and Lacey Mason, University of Michigan
  – Bo Bunnell, Kurt Newman, Norman G Grannemann and Scott Nelson, USGS

• **Project Researchers**
  – Rich Axler, George Host, Norm Will, and Terry Brown, University of Minnesota Duluth
  – Robert Hecky, University of Minnesota Duluth
Thank you!