Tuning the Citizen Science “Instrument” for Gathering Biodiversity Data and Maintaining Data Quality

Rob Stevenson, Yurong He, HeeJun Kim, Todd Suomela
Cornell Laboratory of Ornithology  Bird Team
Reasons Why Citizen Science Could Fail

- Limited knowledge or expertise of participants
- Lack of or limited training of participants
- Lack of consistency from the use of a large number of different observers
- Non-standardized and poorly designed methods of data collection
- Lack of commitment from volunteers can lead to gaps in space and time
- Costs to participate are preventative
<table>
<thead>
<tr>
<th></th>
<th>Stewhart 1986</th>
<th>Wiggins et al 2011</th>
<th>Our Summer Team</th>
<th>DataONE lifecycle language</th>
<th>Data Quality Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specification</strong></td>
<td></td>
<td></td>
<td>Plan</td>
<td>Plan</td>
<td>Plan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td>Before</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data collection and/or data analysis</td>
<td>Collect</td>
<td>Collect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inspection</strong></td>
<td>After</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data cleaning and non-domain specific filter</td>
<td>Assure</td>
<td>Inspect &amp; Clean</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Domain specific filter</td>
<td>Describe</td>
<td>Describe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data storage</td>
<td>Preserve</td>
<td>Preserve</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Sharing</td>
<td>Discover</td>
<td>Discover</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Planning Phase

- Quality assurance project plan
- Design task (e.g., Digital vouchers, Repeated samples/tasks)
- Participants
  - Participant motivation
  - Participant training
  - Participant testing
  - Self-report participant expertise/skills
Data cleaning and non-domain specific filter

Automatic recognition techniques

Error recognition (when: time, where: location, what: data type, missing data...)
Greater Reliance on Participant Judgment

Confidence in the Scientific Value of the Project
Greater Reliance on Participant Judgment

Confidence in the Scientific Value of the Project
Greater Reliance on Participant Judgment

Confidence in the Scientific Value of the Project

Greater Reliance on Participant Judgment
Data Types

Greater Reliance on Participant Judgment
Data Type

People Invent Solutions
• Fold-It Project to understand how proteins fold

People carry sensors with smart phone
• CommonSense Project – Air Quality Monitoring
Data type

No Data → Data from Scientists → Greater Reliance on Participant Judgment
Data Types

People score or translate data from scientists

- Zooniverse

Greater Reliance on Participant Judgment
Data Types

No Data  
Data from Scientists  
Collect Specimens

Greater Reliance on Participant Judgment
Data Types

People provide a specimen to be analyzed back at the lab

- Water sample
- Archeology – Chards of Pottery
- Biodiversity - Ants

Greater Reliance on Participant Judgment
Data Types

No Data → Data from Scientists → Collect Specimens → Digital Data

Greater Reliance on Participant Judgment
Data Types

People submit digital images, sounds recording or video

Many biodiversity recording scheme use digital images

Greater Reliance on Participant Judgment

No Data

Data from Scientists

Collect specimens

Digital Data
Data Types

- No Data
- Data from Scientists
- Collect Specimens
- Digital Data

Greater Reliance on Participant Judgment
Data Types

People report what they see or hear or smell, etc.
Greater Reliance on Participant Judgment

Confidence in the Scientific Value of the Project

Greater Reliance on Participant Judgment
Greater Reliance on Participant Judgment

Confidence in the Scientific Value of the Project

Greater Reliance on Participant Judgment
Greater Reliance on Participant Judgment

Probability Observation is Acceptable

Professional Science Projects

Citizen Scientist Projects

Greater Reliance on Participant Judgment
Greater Reliance on Participant Judgment

Probability observation is Acceptable

Professional Science Projects

Citizen Scientist Projects
Biological Integrity Indices
Professionals vs Citizen Scientists

Genus and family IBIs calculated from the same dataset (MBSS data, 2000 - 2004)
Variability of Indices

Genus and family IBIs calculated from the same dataset (MBSS data, 2000-2004)
Best Practices

1. Publish a Data Management Plan or Quality Assurance Project Plan
2. Think about the kinds of data you collect
3. Understand how your “pixels” work
Best Practices

4. Develop “business rules” to clean the data

5. Interact with people about questionable observations

6. Measure data accuracy
Citizen Science as an Instrument

• Building the Instrument
• Then Calibrate It
REINVENTING DISCOVERY
The New Era of Networked Science
Michael Nielsen