Truth or Consequences

Don Katnik
Maine Inland Fisheries & Wildlife

OFWIM 2010 - Cadiz, KY

Data Integrity in the Information Age
### The Honeymoon is Over

<table>
<thead>
<tr>
<th></th>
<th>THEN</th>
<th>NOW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXPECT</strong></td>
<td>• Modest</td>
<td>• High</td>
</tr>
<tr>
<td></td>
<td>• Having GIS data huge &amp; novel</td>
<td>• NOT having GIS data unheard of</td>
</tr>
<tr>
<td></td>
<td>• Data limitations understood</td>
<td>• Assume all data is current &amp; accurate</td>
</tr>
<tr>
<td><strong>USERS</strong></td>
<td>Mostly “us” (internal)</td>
<td>Mostly “them” (external)</td>
</tr>
<tr>
<td><strong>USES</strong></td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>Broad-scale</td>
<td>Fine-scale</td>
</tr>
</tbody>
</table>
MDIFW’s Observations Database

Microsoft Access Relational Database

ESRI GeoDatabase

1-to-1
1-to-∞
∞-to-∞
Conduct attribute/spatial queries to find errors

“Complicated” because of many-to-many relationship between Access records and ESRI polygons

Step 1. Attribute query to find all Access records coded for export
Manual QC Part II

Step 2. “Push” selection through RELATE to polygon feature class (Table Options > Related Tables)
Step 3. Select from selected polygons those that are coded for export to determine mismatch

Not really “complicated,” just multiple steps – tedious way to check database before every export
Automated QC – Birth of a Tool

ArcObjects
- StandaloneTable = Access Observations Table
- Cursor = query result from table
- Feature Class = ESRI polygons
- Cursor = query result from feature class

VBA Subroutines
- GetTable
- QueryTable (returns Cursor)
- Loop through Cursor
- GetFeatureLayer
- QueryFeatureClass (returns Cursor)
- Test values
If I Can Do That, Why Not...?

- All Features (points & polygons) have valid attributes
  - Not NULL
  - Match domain/list
- No “orphaned” features (not linked to an observation)
- No “orphaned” links in Spatial Lookup Table
- Features/observations linked correctly (many variations besides just for export)
- No cross-links (features linked to same observation coincide)
- “Add” metadata record for every feature
- Mapping protocol current for every polygon

Collection of Tests = Database Integrity Tool
Error Report

***CRITICAL: FEATURES LINKED TO SAME OBSID COINCIDE*** 91.1%

ObsID 205 Primary Habitat Poly does not contain Observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in open water adjacent to saltmarsh at southern end of extensive polygon
> MOVED POINT JUST ENOUGH TO BE INSIDE HABITAT

ObsID 206 Primary Habitat Poly does not contain Observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in open water adjacent to saltmarsh
> MOVED POINT TO APPROXIMATE CENTER OF HABITAT POLYGON

ObsID 207 Primary Habitat Poly does not contain Observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in open water adjacent to saltmarsh
> MOVED POINT TO CENTROID OF HABITAT POLYGON

ObsID 216 Primary Habitat Poly does not contain Observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in forest adjacent to saltmarsh
> MOVED POINT TO CENTROID OF HABITAT POLYGON

ObsID 218 Primary Habitat Poly does not contain Observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in forest adjacent to saltmarsh
> MOVED POINT TO APPROXIMATE CENTER OF HABITAT POLYGON

ObsID 222 Primary Habitat Poly does not contain Observation Point
> Linked to multiple habitat polygons (see Error_Report_2009-11-17)
> WAITING FOR INPUT FROM AM

ObsID 222 Environmental Review Poly does not contain Primary Habitat Poly
> Linked to multiple habitat polygons (see Error_Report_2009-11-17)
> WAITING FOR INPUT FROM AM

ObsID 225 Primary Habitat Poly does not contain observation Point
> Point from Biotics - unclear how accurate it is
> Currently mapped in open water adjacent to extensive saltmarsh polygon
> MOVED POINT JUST ENOUGH TO BE INSIDE HABITAT

ObsID 234 Primary Habitat Poly does not contain Observation Point
> Habitat polygon is Wells Barren
> Biotics point was mapped east of Barren near farm house
> Point represents multiple observations across years
> WAITING FOR INPUT FROM TH

ObsID 234 Environmental Review Poly does not contain Observation Point
A Database Within a Database!

ETSC Observation Database

**ERROR Table**
- Unique ID
- Test Date
- Obs ID
- Feature ID
- Test Code
- Error Code
- Fix Status (Y/N)
- Fix Date
- Fix Code

**REPORT Table**
- Date
- Test 1 Accuracy
- Test 2 Accuracy

**Test Lookup Table**
- Test ID
- Description

**Error Lookup Table**
- Error ID
- Description

**Fix Lookup Table**
- Fix ID
- Description

**ACCESS**
- Obs Table
  - Obs ID
  - Species
  - Date
  - ...

**Point FC**
- Point ID
- Type
  - ...

**Polygon FC**
- Poly ID
- Type
  - ...
Investigate/Fix Errors Tool

**ETSC Database Errors Viewer**

**SEARCH**

<table>
<thead>
<tr>
<th>Feature ID:</th>
<th>Error</th>
<th>ObsID</th>
<th>Feature ID</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly_1569</td>
<td>62</td>
<td>222</td>
<td>Poly_1569</td>
<td>2009-11-17</td>
</tr>
<tr>
<td></td>
<td>307</td>
<td>222</td>
<td>Poly_1569</td>
<td>2009-12-01</td>
</tr>
<tr>
<td></td>
<td>308</td>
<td>222</td>
<td>Poly_1569</td>
<td>2009-12-01</td>
</tr>
</tbody>
</table>

**Search** button

**ERROR DETAILS**

**Test:**
Features linked to the same ObsID coincide

**Error:**
Habitat polygon does not contain observation point

**Notes:**
HAB Poly_1569 incorrectly linked to ObsID 222

**Fix:** 2010-01-21
Corrected link in tblSpatialLU
Database Archive

Allows testing for changes in features:

• Feature no longer exists – “Delete” metadata record?
• Feature has moved/changed shape -- “Modify” metadata record?

Fixing “Change” errors requires investigating feature history:

• When was it created or deleted?
• What observation(s) was it linked to?
• When did it move/change shape?
• What metadata exists?
## Archive Tools

### ObsID Feature History

<table>
<thead>
<tr>
<th>ObsID: 4026</th>
<th>Show ObsID History</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Archive Date</th>
<th>Polygons</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-09-21</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-17</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-16</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-15</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-14</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-13</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-12</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-11</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-09-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-31</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-30</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-29</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-28</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-27</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-08-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-07-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (ES)</td>
</tr>
<tr>
<td>2010-06-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (EO)</td>
</tr>
<tr>
<td>2010-05-04</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (EO)</td>
</tr>
<tr>
<td>2010-04-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (EO)</td>
</tr>
<tr>
<td>2010-03-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (EO)</td>
</tr>
<tr>
<td>2010-02-01</td>
<td>Point_4155 (Animal Location) Poly_4323 (ER), Poly_4324 (EO)</td>
</tr>
<tr>
<td>2010-01-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-12-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-11-02</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-10-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-09-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-08-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-07-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-04-01</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-03-04</td>
<td>(EO)</td>
</tr>
<tr>
<td>2009-02-24</td>
<td>(EO)</td>
</tr>
</tbody>
</table>
The Future

- Convert all VBA code to VB.NET
- Opportunity to generalize tool?

Common elements to each test:
- Objective – what aspect is being tested?
- Test Subjects (e.g., table records, polygons or subset of those)
- Conditional Statement evaluated for each subject
  One or multiple nested “if/then” queries evaluate to T/F
- Output – meaningful report based on objective, subject, and outcome of conditional statement(s)
The Future

- Generic tool = create a set of tests for any database
- Form(s) to build/modify each test by defining its elements:
  - Objective
  - Subjects
  - Conditional statement(s)
  - Error output
- Instead of hard-coding specifics, **store element details in a Table**
- Conditional statements difficult to capture
  - Can involve multiple objects, fields, & values
  - Different classes of statements? Examples:
    - Field value is valid
    - ID exists in a related table/feature class
    - Geometry has changed
  - Different subform Wizard for each class?
The Future
The Future

• Does such an “integrity testing” tool already exist?

• If not, would it be worth the effort to build one (or easier just for everyone to hard-code the specifics themselves?)

• If worth building, could/should it be an OFWIM collaboration?
In the truth is out there?

Questions?