A metadata-based approach to documenting hydrographic surveys

Jeremy McHugh <jeremy.mchugh@noaa.gov>
John Tucker <john.tucker@noaa.gov>
Dan Neumann <dan.neumann@noaa.gov>
James Hiebert <james.hiebert@noaa.gov>

7 May 2008
Outline

1. Introduction
2. Motivation
3. Goal
4. Development
5. Implementation
6. Future Work
7. Acknowledgement

McHugh, Tucker, Neumann, Hiebert
A metadata-based approach to documenting hydrographic surveys
Documentation (metadata) is tedious
Hydrographic data *and metadata* are a precious resource

- $10's of thousands per square nautical mile of coverage
- widespread utility
- growing client base
- good metadata enhances the value of the data
Status quo: loosely structured narrative reports
Motivational factors

- public demand for new data products
- U.S. federal government
  - National Academies’ Ocean Studies Board
  - U.S. Commission on Ocean Policy
  - Bush Administration’s response: Ocean Action Plan
Human error

“To err is human—and to blame it on a computer is even more so.”
-Robert Orben
Cutting and pasting

McHugh, Tucker, Neumann, Hiebert
A metadata-based approach to documenting hydrographic surveys
Redundancy

McHugh, Tucker, Neumann, Hiebert
A metadata-based approach to documenting hydrographic surveys
Goal: A semi-automated report that writes itself to the greatest extent possible

It should enable us to:

- produce reports more efficiently
- reduce inconsistency and eliminate redundancy
- generate a standard reporting/metadata product
- meet external metadata requirements
- populate a repository for hydrographic data
- tap into existing data streams to harvest information
Why use the eXtensible Markup Language (XML)?

- read and understood by both humans and machines
- well-suited for hierarchical semi-structured information
- direct links between metadata and data
- national and international standards with XML implementations (FGDC’s CSDGM and ISO 19139)
Prerequisite to metadata formalization

Introspective analysis of content requirements:

- What bits of information do we want in an ideal report?
- Are there any traditionally documented bits that are not relevant?
- What bits do we need to add to satisfy external clients?
XML schema development (formalization)

XML schema
(excerpt from entire schema)

```xml
<element name="Vessel">
  <element name="VesselName">
    <text/>
  </element>
  <element name="VesselHullNumber">
    <text/>
  </element>
  <element name="LengthOverall">
    <attribute name="units">
      <data type="string"/>
    </attribute>
    <data type="integer"/>
  </element>
  ...
</element>
```

XML model

McHugh, Tucker, Neumann, Hiebert
A metadata-based approach to documenting hydrographic surveys
XML report process

**Operations Branch input**
- Project name + number, survey number, required coverage...
- **field unit input**
  - chief of party hydrographer
- **harvest**
  - AffectedCharts
  - AffectedENCs

**SurveyResults**
- **harvest**
  - features
    - side scan contact list
    - final survey outline
  - **field unit input**
    - supplemental records

**SurveyControls**
- **harvest**
  - bathymetry
  - imagery
  - horizontal control
  - vertical control

**SurveyHQProcessing**
- **AHB / PHB input**
  - review and approval details

**Vessel**
- **field unit input**
  - vessel name
  - hull number
  - length
  - beam
  - draft
  - cruising speed
  - survey speed
  - vessel type

**Measurement**
- **field unit input**
  - measurement device
  - manufacturer
  - year made
  - model
  - serial number
  - comments
  - QA tools
  - date of calibration

**Root**
- traditional metadata
- **XML report**
  - final XML files
    - archived at NGDC

**Multiple output formats via custom -ized stylesheets**
Input Wizard Development
XML stylesheet transformation

XML

```xml
<Vessel>
  <VesselName>NOAA Ship THOMAS JEFFERSON</VesselName>
  <VesselHullNumber>S-222</VesselHullNumber>
  <LengthOverall units="meters">63.4</LengthOverall>
  <Beam units="meters">13.7</Beam>
  <draft units="meters">4.3</draft>
  <CruisingSpeed units="knots">12</CruisingSpeed>
  <MaxSurveySpeed units="knots">12</MaxSurveySpeed>
  <TypicalOperationsType>multibeam + SSS</TypicalOperationsType>
</Vessel>
```

output

<table>
<thead>
<tr>
<th>vessel name</th>
<th>hull number</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA Ship THOMAS JEFFERSON</td>
<td>S-222</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>length</th>
<th>cruising speed</th>
<th>maximum survey speed</th>
<th>typical operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>63.4 m</td>
<td>12 knots</td>
<td>12 knots</td>
<td>multibeam + SSS</td>
</tr>
<tr>
<td>13.7 m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3 m</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Future work

- continue development of input wizard for manual entry
- harvest information from existing data streams
- populate internal and external (e.g., geodata.gov) metadata repositories
Acknowledgment

Thank you to coauthors, especially James Hiebert!