

# 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

# Documentation (metadata) is tedious

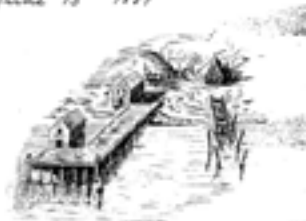


# Hydrographic data *and metadata* are a precious resource

- \$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

*Tide Gauge, Chatham Wharf, Stage Harbor, Mass.  
Established June 13 1857*



## *Description and Data.*

*The gauge is upon the eastern side of the easternmost fish wharf in Stage Harbor. It is nailed to one of the piles of the wharf and is about 10 yards from the outer end. The T.M. is a copper tack driven in the pile to which the gauge is attached.*

# 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





# Redundancy



# 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 heirarchical 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)

```

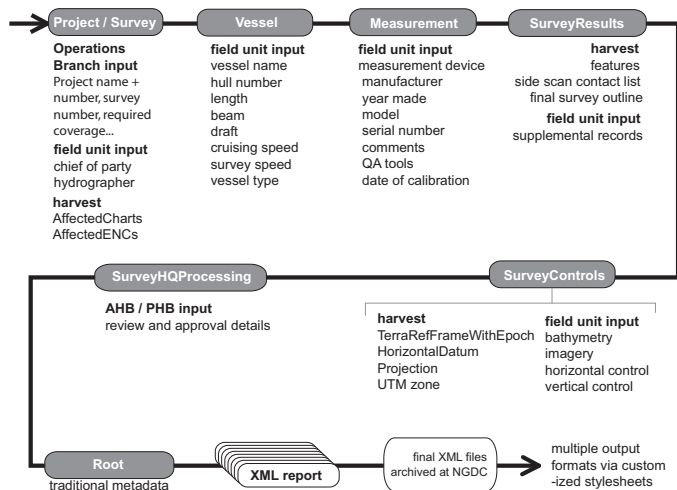
<element name="Vessel">
  <element name="VesselName">
    <text/>
  </element>
  <element name="VesselHullNumber">
    <text/>
  </element>
  <element name="LengthOverall">
    <attribute name="units" >
      <data type="string"/>
    </attribute>
    <data type="integer"/>
  </element>
  ...
</element>

```

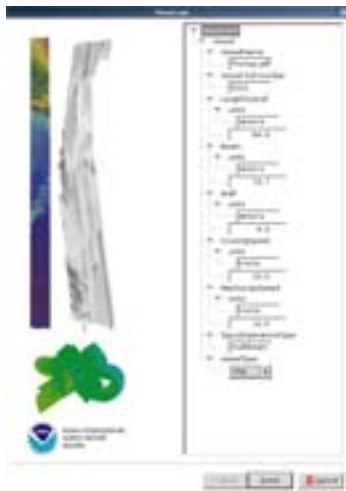
## XML model



# XML report process



# Input Wizard Development



# XML stylesheet transformation

## 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>
  
```



XML stylesheet

## output

vessel name	<b>NOAA Ship THOMAS JEFFERSON</b>	hull number	<b>S-222</b>
length	<b>63.4 m</b>	cruising speed	<b>12 knots</b>
beam	<b>13.7 m</b>	maximum survey speed	<b>12 knots</b>
draft	<b>4.3 m</b>	typical operations	<b>multibeam + SSS</b>



# Future work

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

# Acknowledgment

Thank you to coauthors, especially James Hiebert!