

Hello, I am Briana Sullivan. I work in the Data Visualization Lab on the “Chart of the Future” related projects.

I’d like to discuss the Coast Pilot data structure as it is today and some insights on where I think it should be and some ideas on how to make it happen.

I’ll follow up with a brief synopsis of the work we are involved in with the various IHO Working Groups.



let me first explain, visually, how the data structure is based on the publication...

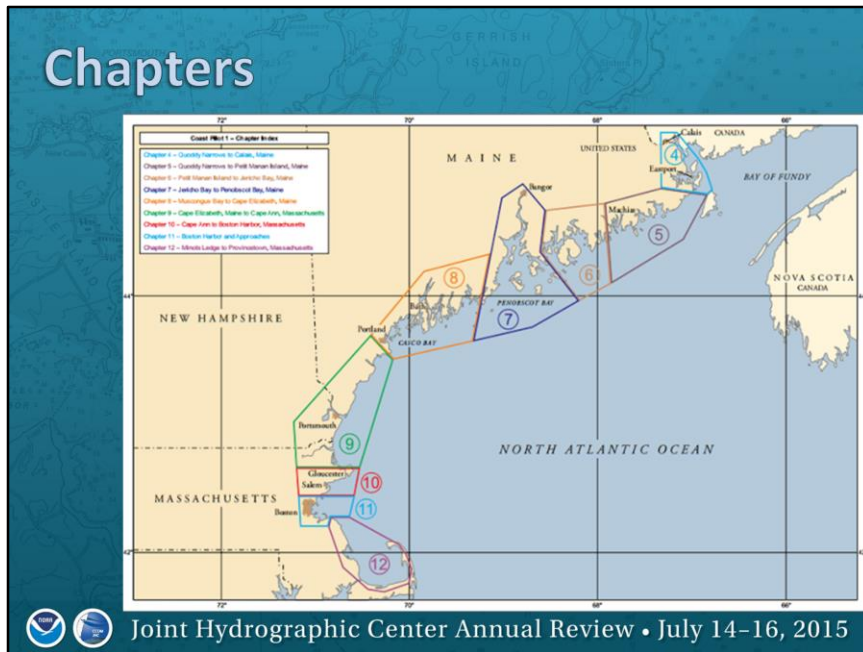
# Books



Joint Hydrographic Center Annual Review • July 14–16, 2015

The Coast Pilot is set up via Books. Books subdivide the working area into maintainable sections. So, to get local information you need to know the Book as the root element

# Chapters



The chapters unique to each book, subdivide the area represented by the book into manageable parts. NOTE: each chapter contains one or more nautical chart within it's limits (and due to this some information repeats where charts/chapters/books share a border/overlap)



Coast Pilot 1 - Chapter 4 - Edition 45, 2015

**Quoddy Narrows to Calais, Maine**

1. This chapter describes the Maine and New Brunswick coasts from Quoddy Narrows through Lubec Channel, Friar Roads, Western Passage, and the St. Croix River to the head of navigation at Calais. Included in the text are discussions of the Maine ports of Lubec, Eastport, and Calais; the Canadian ports of St. Stephen and St. Andrews; several small harbors on Campobello Island; and Head Harbour Passage.

2. COLREGS Demarcation Lines

3. The lines established for this part of the coast are described in 80.105 chapter 2.

4. Charts 13394, 13396, 13398

5. The approaches to St. Croix River include Quoddy Narrows, Lubec Channel, Friar Roads, Head Harbour Passage, Western Passage, and Passamaquoddy Bay. The principal entrance is around the northern end of Campobello Island through Head Harbour Passage. This passage is deep and generally clear of dangers. The channel through Lubec Narrows is also used, especially at high water. The tidal currents are strong in both passages.

6. West Quoddy Head the easternmost point of the United States, is bold and wooded. West Quoddy Head Light (44°43'31"N, 69°5'70"W), 83 feet above the water, is shown from a 49-foot red and white horizontally banded tower on the eastern edge of the headland. A sound signal is at the light. The abandoned Coast Guard lookout tower near the summit of the ridge westward of the light is the most conspicuous landmark in the approach to Quoddy Narrows from seaward.

7. Between West Quoddy Head and Calais, fluorescent red pyramidal markers define straight line segments and turning points of the United States-Canada boundary.

8. Quoddy Narrows (Quoddy Roads) between West Quoddy Head and Canada's Campobello Island, is the usual anchorage for vessels seeking shelter or waiting for a favorable tide to pass through Lubec Narrows. The entrance, between West Quoddy Head and The Boring Stone, is about 0.8 mile wide and has a depth of 28 feet near the middle. Winds from east to south generate rough seas in the entrance.

9. The anchorage affords shelter from northerly and westerly winds in depths of 12 to 25 feet, but is open to winds from the east and south, and protection from northeast gales is reported poor. The northern and western parts of Quoddy Narrows between West Quoddy Head and Lubec are full of shoals which partly uncover.

10. Sail Rock and Little Sail Rock are two bare rocks on a ledge about 0.2 mile southeastward of West Quoddy Head Light. The ledge extends more than 100 yards east of the two rocks. As swirls form just southward and eastward of Sail Rock during the strength of the tidal current, the rock should be given a good berth. A lighted whistle buoy is about 0.4 mile southeastward of Sail Rock, about in line with the rock and West Quoddy Head Light. A fairway bell buoy, about 0.5 mile north-northeastward of the light, marks the entrance to Quoddy Narrows and the approach to Lubec Channel.

11. Round Rock which uncovers, and The Boring Stone 3 feet high and bare, are 500 yards southwest of Liberty Point a bold headland, which is the southern extremity of Campobello Island. Vessels should pass at

**Chapter Elements**

- Chapter Header
- Chart Header
- Paragraph Header
- Paragraph -> CP\_Tags
- Images

U.S. Coast Pilot®  
Color Legend

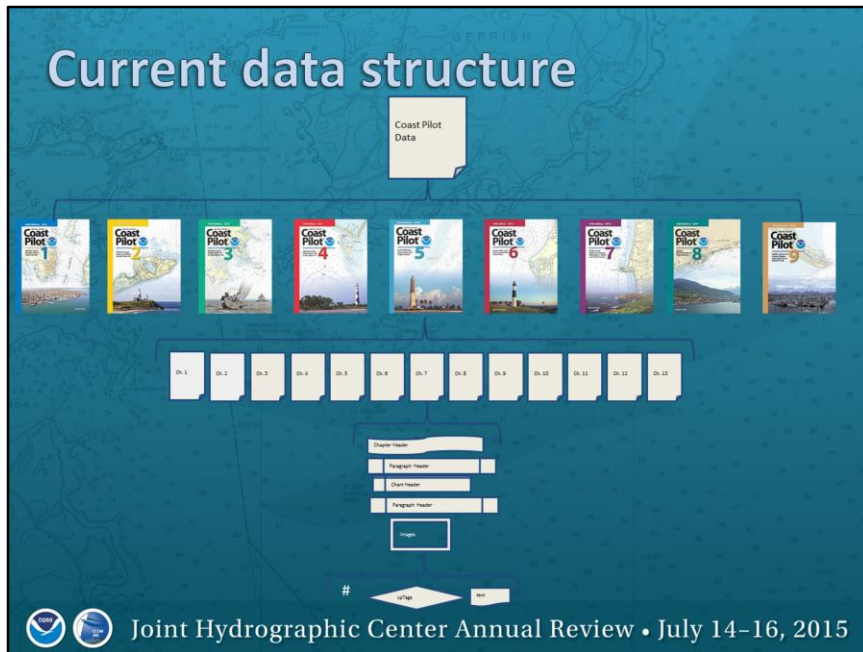
- Chart
- Demarc
- Code of Federal Regulations
- Updated Paragraph

Joint Hydrographic Center Annual Review • July 14–16, 2015

Within each chapter are “elements”. Each element is represented as a paragraph within the chapter and is numbered. But the elements have types associated with them such as:

- Chapter Header (Orange)
- Chart Header (Yellow)
- Paragraph Header (Aqua)
- Paragraph -> CP\_Tags (Blue)
- Image (Pink)

These items are typically described in geographic order. On the East Coast of the US, for example, it is from North to South.



This is a hierarchal data structure ....just like a tree, where the whole CP is the tree, each book is a branch, each chapter are leaves on the branch, etc.





"I honestly think we're seeing a more profound change, for map-making, than the switch from manuscript to print in the Renaissance,"


-University of London cartographic historian Jerry Brotton told the Sydney Morning Herald.

"That was huge. But this is bigger."

# Thinking Differently

"There are a couple of steps (to make a Google Map). You acquire data through partners. You do a bunch of engineering on that data to get it *into the right format* and conflate it with other sources of data, and then you do a bunch of operations, to hand massage the data. And out the other end pops something *that is higher quality than the sum of its parts.*"

-former NASA engineer Michael Weiss-Malik

 Joint Hydrographic Center Annual Review • July 14–16, 2015

With the S-100 standard in the works, technology use common place, and a need for more useable and accessible data it's "time to think differently".

*It's probably better not to think of Google Maps (or Nautical Charts) as a thing like a paper map.*

*Geographic information systems represent a jump from paper maps like the abacus to the computer.*

A quote from London cartographic historian ...

(read quote at top of slide).

A former NASA engineer working at Google lays out the steps to the success of Google Maps...(read quote at bottom of slide)

Isn't that the holy grail? Getting something out of this that is "higher quality than the sum of its parts"?

source: <http://www.theatlantic.com/technology/archive/2012/09/how-google->

[builds-its-maps-and-what-it-means-for-the-future-of-everything/261913/](#)



We can accomplish this if we can move from a publication-centric data format to a data-centric one. What do I mean by this?

## Rewrite

*“(Coast Pilots) contain supplemental information that is  
difficult to portray on a nautical chart.”*

Coast Pilots *describe* the features  
found along the coast  
and  
how to get to / access / use / avoid them.



Joint Hydrographic Center Annual Review • July 14–16, 2015

It brings me back to that disclaimer I mentioned earlier...(read slide).

OCS is starting the process of thinking differently

## From the OCS website

The Coast Pilot project needs to  
**share its information** with other sources and  
offer a **customizable** product

**This degree of customization** requires **more control** than a traditional  
Desktop Publisher (DTP) can provide

...the Extensible Markup Language (XML)  
and associated tools have that capability.

*If done correctly!*



Joint Hydrographic Center Annual Review • July 14–16, 2015

...I recently found this on their web page...(read slide)

But remember...just tagging something with XML doesn't mean it will allow the kind of control mentioned here.

Thought and foresight needs to come into play to make sure the mark-up is being used to maximize data use.

source: <http://www.nauticalcharts.noaa.gov/nsd/RD-cpdb.html>

# Example

**Metadata relating to chapter details**

**Quoddy Narrows to Calais, Maine**

This chapter describes the Maine and New Brunswick coasts from Quoddy Narrows through Lubec Channel, Friar Roads, Western Passage, and the St. Croix River to the head of navigation at Calais. Included in the text are discussions of the Maine parts of Lubec, Eastport, and Calais; the Canadian parts of St. Stephen and St. Andrews; several small harbors on Campobello Island; and Head Harbour Passage.

**COLREGS Demarcation Lines**

The lines established for this part of the coast are described in 80.105 chapter.

**Charts 13394, 13396, 13398**

The approaches to St. Croix River include Quoddy Narrows, Lubec Channel, Friar Roads, Head Harbour Passage, Western Passage, and Passamaquoddy Bay. The principal entrance is around the northern end of Campobello Island through Head Harbour Passage. This passage is deep and generally clear of dangers. The channel through Lubec Narrows is also used, especially at high water. The tidal currents are strong in both passages.

West Quoddy Head the easternmost point of the United States, is bold and wooded. West Quoddy Head Light (44°44'34"N, 66°57'02"W), 83 feet above the water, is shown from a 49-foot red and white horizontally banded tower on the eastern edge of the headland. A sound signal is at the light. The abandoned Coast Guard lookout tower near the summit of the ridge westward of the light is the most conspicuous landmark in the approach to Quoddy Narrows from seaward.

[104] St. Croix River extends north-northwestward for 8 miles from the southern part of Passamaquoddy Bay, then turns westward between Devils Head and Todds Point. The channel is deep and comparatively clear as far as the turn, then is narrow and winding, and has a controlling depth of about 16 feet for some 3 miles to Hills Point (47°02'57"N, 67°12'23"W).

[105] A dredged channel leads from above Hills Point to Calais. In 1977, the midchannel controlling depth was 7 feet to Todd Point, about 4.2 miles above the mouth; thence 3 feet to Calais and St. Stephen on the Canadian side of the border, except for shoaling to 3 feet about 90 feet below the International Bridge at Calais. The channel is marked by lights and buoys, but is not charted and the two buoys on the north side of the channel at The Narrows opposite Whitelocks Mill Light 25 ton under during the strength of the tide. Local knowledge is necessary for the river above Whitelocks Mill.

[106] Small craft up to 40 feet in length can anchor in 14 feet on the west side of the channel just above Whitelocks Mill Light, but larger craft should anchor off Devils Head.

[107] The scattered remains of an old breakwater, which uncover 12 feet in spots, extend southward across the mudflats on the south side of St. Croix River for about 300 yards from near channel buoy 19. The mudflats, which uncover 11 feet, are opposite The Ledge, a village on the north side of the river about 9.7 miles above the mouth; caution is advised in this area.

[108] Ice

[109] St. Croix River above Robbinston is reported to be closed by ice for about one or two weeks in February. The channel to the of what is Calais is usually kept open by the tug and barge bound there. Quoddy Narrows and Eastport Harbor are never closed by ice.

```

<feature name="St. Croix River">
  <extent>
    <direction>north-northwestward</direction>
    <dist units="miles">8</dist>
    <from>the southern part of <place>Passamaquoddy
    Bay</place></from>
    <to>then turns westward between <place>Devils Head</place>
    and <place>Todds Point</place>
  </extent>
  <characteristics>
    <characteristic>narrow</characteristic>
    <characteristic>windy</char
  </characteristics>
  <channelDepth>deep</channelDepth>
  <controllingDepth units="feet">16</controllingDepth>
  <controllingLength units="miles">3</controllingLength>
  <approaches>
    <approach>Quoddy Narrows</approach>
    <approach entrance="secondary">Lubec Narrows</approach>
    <approach entrance="principal">Friar Roads</approach>
    <approach entrance="principal">Head Harbour
    Passage</approach>
    <approach>Western Passage</approach>
    <approach>Passamaquoddy Bay</approach>
  </approaches>
  <ice>
    <closurePlace>Robbinston</closurePlace>
    <closureDuration>one or two weeks</closureDuration>
    <closureMonth>February</closureMonth>
  </ice>
</feature>

```

Joint Hydrographic Center Annual Review • July 14–16, 2015

So to demonstrate how “Thinking Differently” could work to maximize the data...here’s an example paragraph.

Notice how the info in the red outlines could be considered “metadata” to this chapter/book. The Chart headings would be irrelevant in a world of charts based on features. Now the line is blurred between charts and boundaries that were so prevalent in the past.

The XML colored in black is directly from paragraph 5 (the green highlight). Basically, I read the paragraph and attempted to take everything I could and “tag” and organize it. But, since I’m focused on the feature, I’m stopping at the first one I see, “St. Croix River”. Then I search the rest of the chapter for “St. Croix River”. My goal is to find and group all info about the “St. Croix River” that I can. The orange text is just that...info from the rest of the chapter about the St. Croix River. Wow, how great is it to be able to find out all about this river in one location!

Doing this exercise brought up quite a few other questions in quality control with consistency, formatting, and organization. Like:

1) Why isn’t the first instance of “St. Croix River” (5) in bold? Instead it is bold in

(104).

- 2) Principal entrance is listed in (5) as “Head Harbour Passage”, yet elsewhere in the chapter “Friar Roads” is also a principal entrance
- 3) what is the determining factor to having the geo-referenced links?
- 4) What aren't all the occurrences within the text (i.e. Whitlocks Mill) all geo-referenced? I should be able to click on any one of those feature names and get the same result.

# Example

```


<!-- symbol definitions -->
<Book Number="1" BookID="327" Edition="45" ChapterNo="4" Title="Coast Pil
<Chapter>
  <img_detail IS_NUM="0" ImageTag="ch4" DisplayFile="Images/CP1Chapter
  93a1-6b4e411cb518.jpg"/>
  <chapterTitle>Quoddy Narrows to Calais, Maine</chapterTitle>
  <paragraph>
    <paraIndex>(1)</paraIndex>
    <paraText>
      <spacers>/spacers
      This chapter describes the Maine and New Brunswick coasts from Quo
      Calais. Included in the text are discussions of the Maine ports of
      Island; and Head Harbour Passage.
    </paraText>
  </paragraph>
  <paragraphHeader>
    <paraIndex>(2)</paraIndex>
    COLREGS Demarcation Lines
  </paragraphHeader>
  <paragraph>
    <paraIndex>(3)</paraIndex>
    <paraText>
      <spacers>/spacers
      The lines established for this part of the coast are described in
      <CP_B>80,105,</CP_B>
      chapter 2.
    </paraText>
  </paragraph>
  <chartHeader>
    <paraIndex>(4)</paraIndex>
    Charts 13394, 13396, 13398
  </chartHeader>
  <paragraph>
    <paraIndex>(5)</paraIndex>
    <paraText>
      <spacers>/spacers
      The approaches to St. Croix River include Quoddy Narrows, Lubec Ch
      northern end of Campobello Island through Head Harbour Passage. Th
      water. The tidal currents are strong in both passages.
    </paraText>
  </paragraph>
  <paraIndex>(6)</paraIndex>

```

```

<feature name="St. Croix River">
  <extent>
    <direction>north-northwestward</direction>
    <dist unit="miles">8</dist>
    <from>the southern part of <place>Passamaquoddy
    Bay</place></from>
    <to>then turns westward between <place>Devils
    Head</place> and <place>Todd's Point</place>
  </extent>
  <characteristics>
    <characteristic>narrows</characteristic>
    <characteristic>windy</char
  </characteristics>
  <channelDepth>deep</channelDepth>
  <controllingDepth units="feet">16</controllingDepth>
  <controllingLength units="miles">3</controllingLength>
  <approaches>
    <approach>Quoddy Narrows</approach>
    <approach entrance="secondary">Lubec
    Narrows</approach>
    <approach entrance="principal">Friar Roads</approach>
    <approach entrance="principal">Head Harbour
    Passage</approach>
    <approach>Western Passage</approach>
    <approach>Passamaquoddy
    Bay</approach>
  </approaches>
  <ice>
    <closurePlace>Robbinston</closurePlace>
    <closureDuration>one or two
    weeks</closureDuration>
    <closureMonth>February</closureMonth>
  </ice>
</feature>

```


 Joint Hydrographic Center Annual Review • July 14–16, 2015

the current XML (publication-centric) mark-up...comparing this time against the data-centric mark-up example on the right.

Note: the St. Croix River (the green box) is inside the <paraText> tag hidden amongst a narrative string of words. Not easy to quickly find or extract.

Granted...it's a LOT of work! But the uses for how the data is marked up on the right is many times more useful than the original.



# Example

Coast Pilot Data - 2015

Paragraph Number:

Paragraph Text:

The approaches to St. Croix River include Quoddy Narrows, Lubec Channel, Friar Roads, Head Harbour Passage, Western Passage, and Passamaquoddy Bay. The principal entrance is around the northern end of Campobello Island through Head Harbour Passage. This passage is deep and generally clear of dangers. The channel through Lubec Narrows is also used, especially at high water. The tidal currents are strong in both passages.

Coast Pilot Data - the future

Feature:

**Characteristics**

description:

description:

description:

**Approaches**

Principal Entrance:

Secondary Entrance:

Approach:

**Extent**

Distance (miles):

**Bounding Box**

lat/long polygon:

**Orientation/Direction**

**Ice**

Closure Location:

Closure Duration (max weeks):

Closure Month:



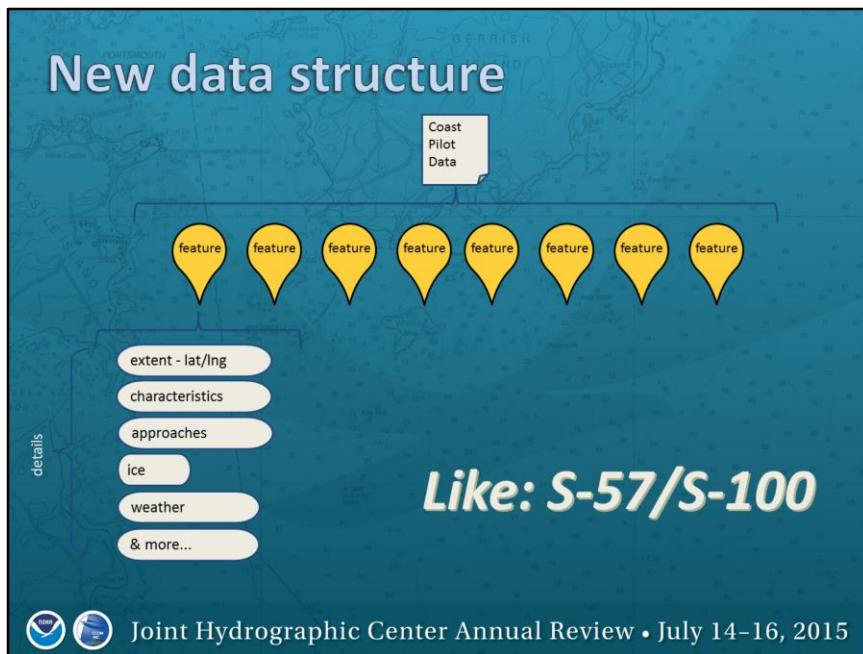
Joint Hydrographic Center Annual Review • July 14–16, 2015

For those of you who don't like to look at code, another way to view and understand this would be like filling out a form online...

The example on the left allows for a narrative - a rich textual description. Easy for input purposes, but just not easy to process, discover and use on the output.

The example on the right, sure it's more complicated to set up, it takes longer to fill out...BUT, it forces the input to be formatted, tagged, organized, and verified on submit, which means there is no more/very little work to do on the output...

This is a great example of how to standardize the data content.



Yet another way to view this data structure (for those of you that like pictures!)

This is the direction of the S-100...Feature-based data...

it can work for the Coast Pilot/Sailing Directions too...in fact it should work in conjunction with S-101 (the ENC) data.

Everything related to a specific physical feature (S-126 - the physical environment) should be able to reference the S-101 ENC utilizing the overlap and simply linking the features together. Remember the goal to reduce redundant work and work towards “harmonization” with other related specifications.

For the most part the S-126 is basically describing the S-101 feature....the textual description of the physical feature as it relates to navigation.

## Why? – the end result

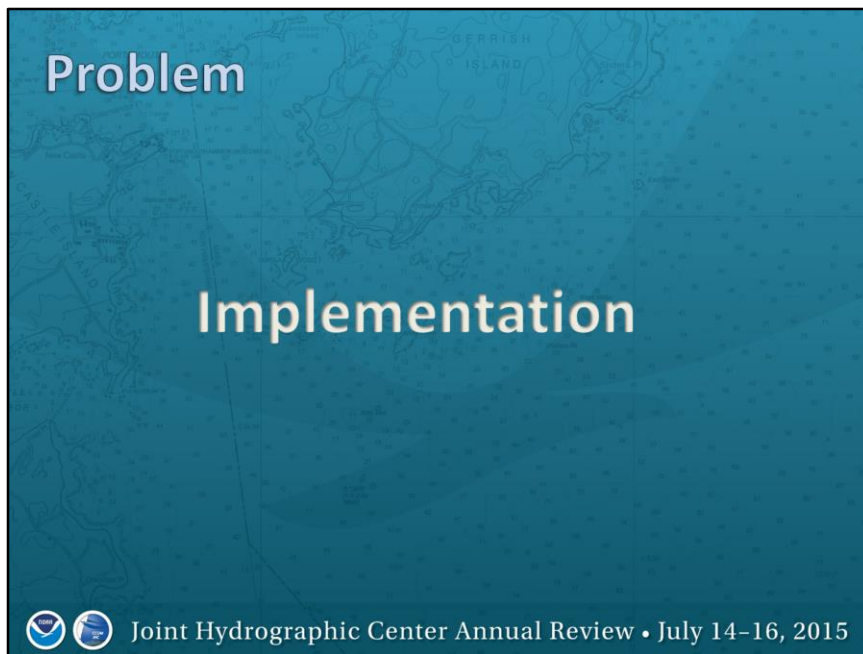
1. A mariner of a small private vessel wants to get rid of all text relating to pilots and large ships. (and vice-versa)
2. Correlating the weather data in the text with historical, current and predictions of the same area.
3. A researcher wants to test AUV in a specific area and wants a quick way to search for all rivers that have a “windy” characteristic.
4. Filter out anchorages along a coastline to help a mariner decide which to head for. (using S101 feature catalogue rules of anchorage types relevant to her - code 8: small craft mooring area).
5. Get drawbridge information for a route quickly...all in one place.
6. Text reader for a audio description of area



Joint Hydrographic Center Annual Review • July 14–16, 2015

Remember, we need to keep in mind the end result, how will the data be used? It is (or should) be there to serve us...not us be slave to it.

1. (or better would be “winding” since “windy” in the dictionary is “wind swept” not “twisty”...but that’s another presentation!)



Of course, the biggest problem in this whole story is HOW do you get from the current situation to the dream?

Or even something on the way to the dream? I don't think the presentation should end here (even though you may want it to!)



## Intermediate steps

1. *Normalize/clean the database*
2. Geo-reference data (use GML)
3. Add Chart numbers to all elements (for backward compatibility)
4. Better Tags:
  1. Replace cp\_index tags to ref/extend feature tags - *standardize*
  2. Replace “formatting” tags (cp\_bold, cp\_italic) with feature tags
5. Add categories for filterable data
6. Recode tags with fewer attributes

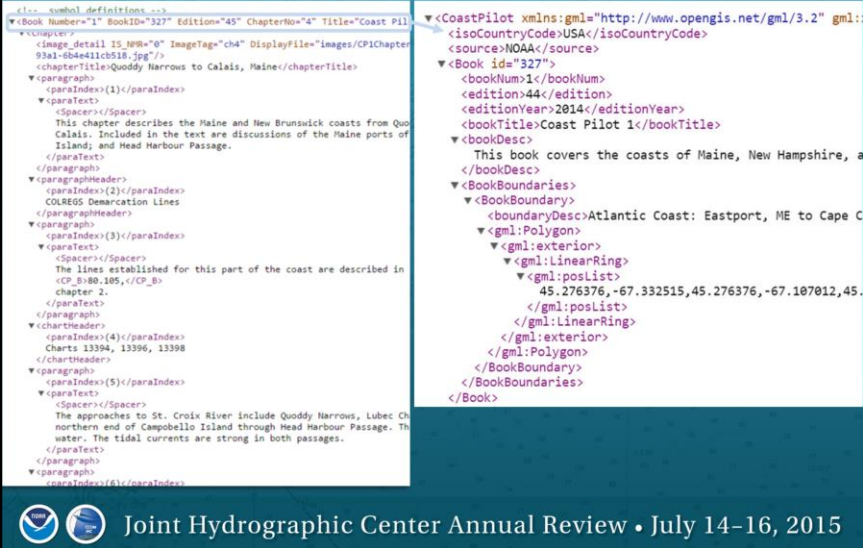


Joint Hydrographic Center Annual Review • July 14–16, 2015

1. Process of reducing redundancy in the DB.
  - a. Use query string to find distinct Paragraph Header type of elements, then clean the list manually
    - i. Get rid of tags, duplicates, make things generic (get rid of state and regional specific locations in text) - this could be a starter list for common headers in all of the world's sailing directions -- This is part of the STANDARDIZATION process! (example: Pilotage for Coast of New Hampshire, just Pilotage since it's already in the section about New Hampshire)
      1. duplicates in the database reduce headers by 318 entries (from 2258 to 1940),
    - ii. This is a great exercise to see how consistent the document is written, how terms/phrases/headers are similar but not the same.(i.e. using MA vs. Mass, VA vs. Va., NY vs. N.Y., under way vs. underway, use of caps, use of spaces/punctuation, plural vs. singular, cross current vs. crosscurrent, small craft vs. small-craft vs. small-boat)
  - b. Get rid of added tags and attach as a new field “type” or “paragraph header type” (Appendix, CFR, sect, CP\_INDEX, CP\_BOLD, etc.)

- C. Punctuation within tags isn't necessary.
- 1. Geo-reference data - how is Tom's group doing it (regex automation to start)  
According to publication "Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia"
  - a. Checking for potential matches: find all possible matches (either use the xml version of each document or make new version from DB) - "feature catalogue"
  - b. match with GNIS table
- 2. Isolate the the chart numbers using regex, set up parent/child relationships via query and create new table of chart numbers to elements.  
This set will give everything an associated area even if it doesn't have a geo-reference.
- 3. Compare against the ENC feature catalogue so the tagging can be synced up/reused
- 4. Such as Anchorage, Pilotage, etc. Later these can be changed into feature categories that will comply with S-100 standards.
- 5. recode without attributes

## Example: fewer attributes



The image shows a side-by-side comparison of XML code for a Coast Pilot. The left side shows a verbose version with many attributes, while the right side shows a simplified version with fewer attributes.

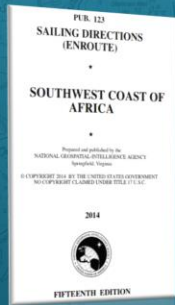
```
<!-- symbol definitions -->
<Book Number="1" BookID="327" Edition="45" ChapterNo="4" Title="Coast Pil
<img alt="Image of Quoddy Narrows" data-bbox="200 176 480 200"/>
<chapterTitle>Quoddy Narrows to Calais, Maine</chapterTitle>
<paraIndex>{1}</paraIndex>
<paraText>
  This chapter describes the Maine and New Brunswick coasts from Quo
  Calais. Included in the text are discussions of the Maine ports of
  Island; and Head Harbour Passage.
</paraText>
<paraIndex>{2}</paraIndex>
COLREGS Demarcation Lines
<paraIndex>{3}</paraIndex>
The lines established for this part of the coast are described in
chapter 2.
<paraIndex>{4}</paraIndex>
Charts 13394, 13396, 13398
<paraIndex>{5}</paraIndex>
The approaches to St. Croix River include Quoddy Narrows, Lubec Ch
northern end of Campobello Island through Head Harbour Passage. Th
water. The tidal currents are strong in both passages.
<paraIndex>{6}</paraIndex>
```

```
<CoastPilot xmlns:gml="http://www.opengis.net/gml/3.2" gml:
  <isoCountryCode>USA</isoCountryCode>
  <source>NOAA</source>
  <Book id="327">
    <bookNum>1</bookNum>
    <edition>44</edition>
    <editionYear>2014</editionYear>
    <bookTitle>Coast Pilot 1</bookTitle>
  </Book>
  <BookDesc>
    This book covers the coasts of Maine, New Hampshire, a
  </BookDesc>
  <BookBoundaries>
    <BookBoundary>
      <boundaryDesc>Atlantic Coast: Eastport, ME to Cape C
    </BookBoundary>
    <gml:Polygon>
      <gml:exterior>
        <gml:LinearRing>
          <gml:posList>
            45.276376,-67.332515,45.276376,-67.107012,45.
          </gml:posList>
        </gml:LinearRing>
      </gml:exterior>
    </gml:Polygon>
  </BookBoundaries>
</Book>
```

Fewer Attributes: This is an easy interim step that could be done now to help improve the way the Coast Pilot XML is currently being done.



# Nautical Information Provision WG (NIPWG)



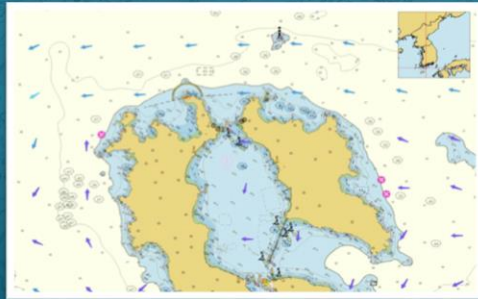
- S-122 – Marine Protected Areas
- S-123 – Radio Services
- S-125 – Navigational Services
- S-126 – Physical Environment
- S-127 – Traffic Management
- S-128 – Marine Services
- S-xxx – Digital Mariners' Routing Guide
- S-xxx – Harbor Infrastructure



Joint Hydrographic Center Annual Review • July 14–16, 2015

The data structure of the Coast Pilot

# Tides, Water level, Currents WG (TWCWG)

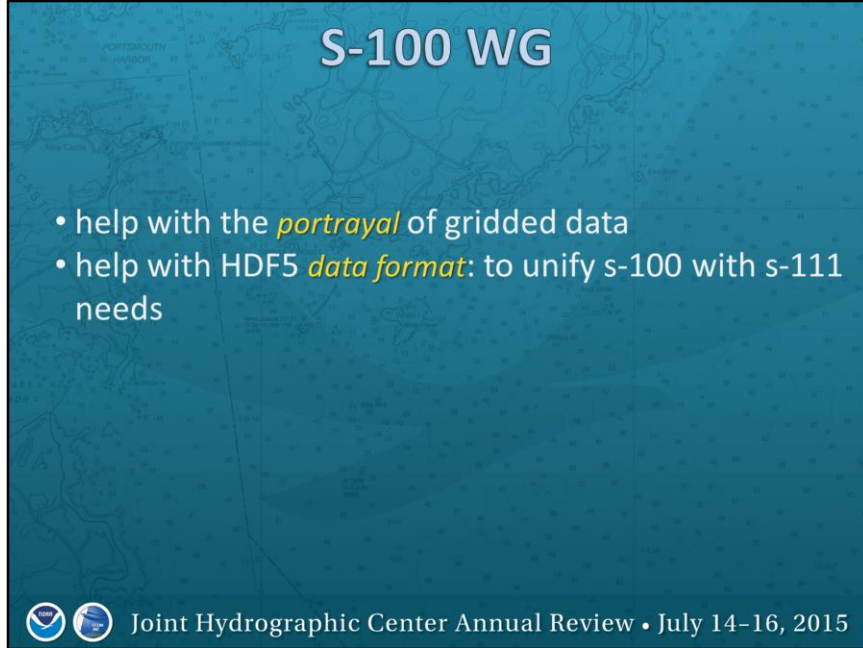


- 2014-2015 Develop initial draft of S-111 PS that includes *portrayal*, schema, and encoding
- 2014-2015 Test the *portrayal* of surface current vectors and real time current data





Joint Hydrographic Center Annual Review • July 14–16, 2015

The data structure of the Coast Pilot



## S-100 WG

- help with the *portrayal* of gridded data
- help with HDF5 *data format*: to unify s-100 with s-111 needs

  Joint Hydrographic Center Annual Review • July 14–16, 2015

The data structure of the Coast Pilot

