Coast Pilot Workshop

DATA STRUCTURES, MRN AND INTEROPERABILITY

briana sullivan – UNH/CCOM

Tom Loeper – NOAA/NSD

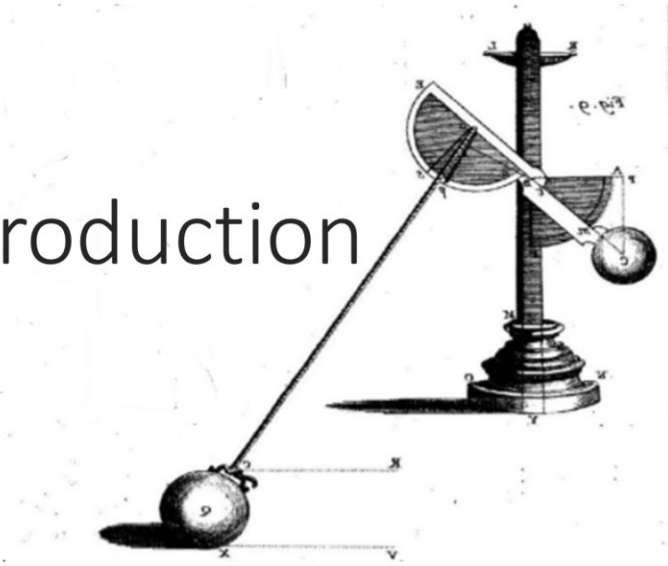
This presentation is a report from a workshop led by UNH and held at NPB in November 2018 to begin the process of implementation of a data-centric production system; it will attempt to highlight the current publication-centric production system, the mechanism moving forward to slowly transition into a data-centric system, as well as discussions and issues that came up relating to interoperability with the ENC data.

This presentation is an attempt to share ideas and challenges with the global community on transitioning to a new data services system based on S-100 standards as well as to document the journey for others to learn from or lend ideas to.

Current Production System

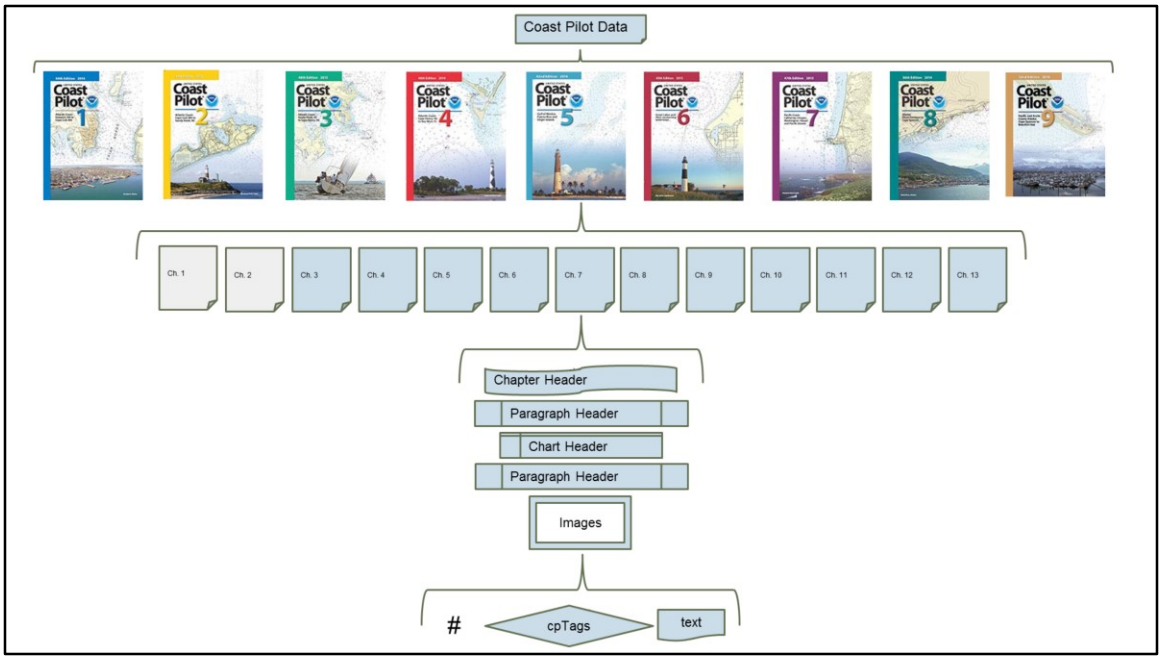
PUBLICATION-CENTRIC

LEAD-LINES



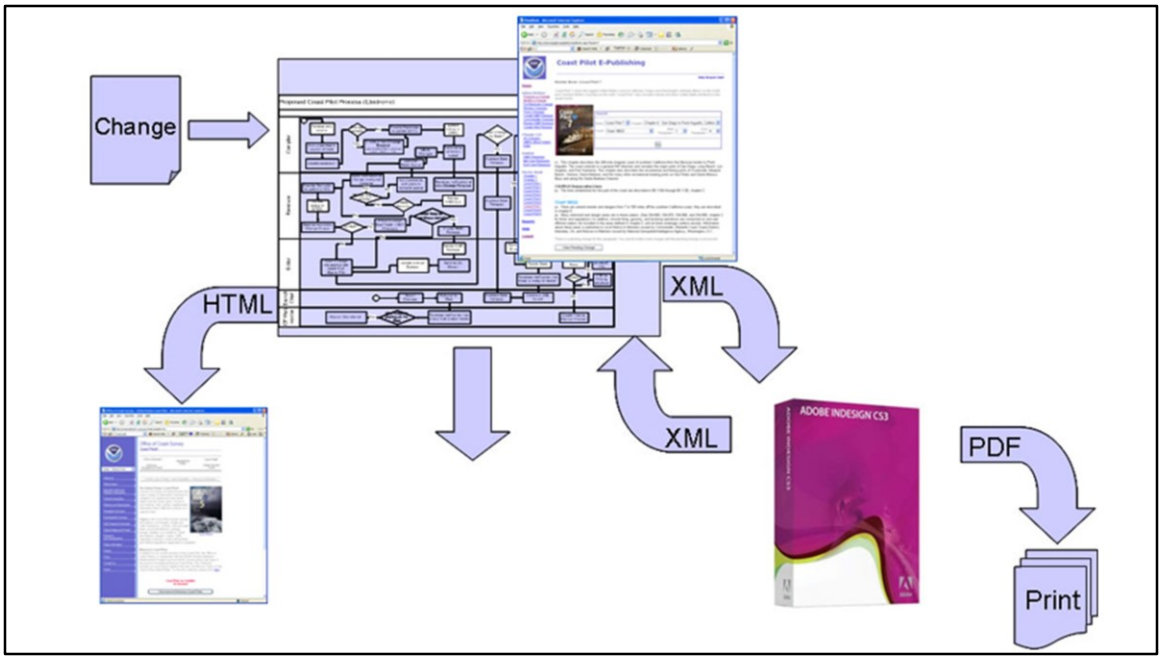
Making a comparison to the evolution of soundings on the chart, the publication-centric format of maritime documents are currently like “lead-lines” where a data-centric paradigm can transform them into “multi-beam”.

First some background to where the system is currently and what we had to work with.



The US Coast Pilot's (CP) current production system involves a database laid out in a structure of tables containing the information and metadata for books, chapters and chapter elements (consisting of headers, paragraphs and images – see NIPWG1-06.3 for details).

Simple database schema designed specifically for the production of a publication in book form.



Connecting the database to Adobe InDesign the Coast Pilot data can then be presented as a pdf, HTML or XML.

NOTE: the geo-tagged XML on the left is where the current production system is, still publication-centric.

The XML representation on the right is feature-centric where everything about the Feature in question becomes an attribute related to that feature. This is a more robust way to extract the data out of the text and make it more machine readable.

Features

DATA-CENTRIC

The transformation from publication-centric to data-centric paradigm means that we need to start with the features.

Features in the Coast Pilot – object name

(18) Many of the distances in this and later Chesapeake Bay chapters are given in nautical miles above the **Virginia Capes**, or "the **Capes**," which is a short way of referring to a line from Cape Charles Light to Cape Henry Light.

(19) **Cape Charles**, on the north side of the entrance, is low and bare, but the land back of it is high and wooded. **Wise Point** is the most southerly mainland tip of the cape. Low **Fishermans Island**, a National Wildlife Refuge, is 1 mile south of Wise Point.

(20) The southwest end of **Smith Island** is 2.4 miles eastward of Wise Point; the island is 6 miles long, low and sparsely wooded and awash at half tide midway along its length.

(21) **Cape Charles Light** (37°07'23"N., 75°54'23"W.), 180 feet above the water, is shown from an octagonal, pyramidal skeleton tower, upper part black and lower

ID	object name
1	Virginia Capes
2	The Capes
3	Cape Charles
4	Wise Point
5	Fishermans Island
6	Smith Island
7	Cape Charles Light

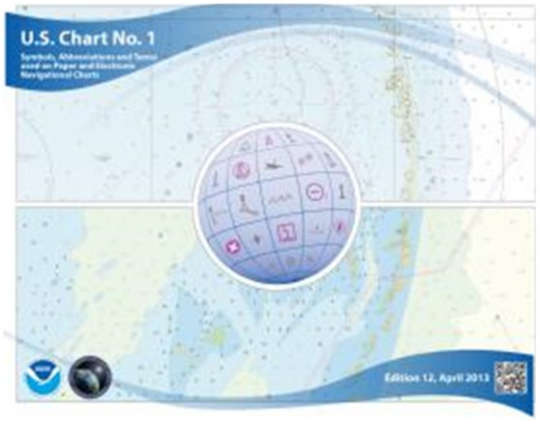
Features drive the production of the ENC as well as the content of the Coast Pilot and all other chart supporting data.

So we first need to find and track all of the features that are in the text.

The items in BOLD are already geo-tagged in the text of the database…these become the initial list, a new table in the database.
Using the S-100 standards nomenclature the column is named "object name".

The next step would be to categorize the features.

Categories – based on Chart 1 ontology



Symbol Sections

GENERAL
- A Chart Number, Title, Marginal Notes
- B Positions, Distances, Directions, Compass

TOPOGRAPHY
- C Natural Features
- D Cultural Features
- E Landmarks
- F Ports
- G (Not currently used)

HYDROGRAPHY
- H Tides, Currents
- I Depths
- J Nature of the Seabed
- K Rocks, Wrecks, Obstructions, Aquaculture
- L Offshore Installations
- M Tracks, Routes
- N Areas, Limits
- O (Not currently used)

NAVIGATION AIDS AND SERVICES
- P Lights
- Q Buoys, Beacons
- R Fog Signals
- S Radar, Radio, Satellite Navigation Systems
- T Services
- U Small Craft (Leisure) Facilities

The Chart 1 is ideal for

1. The ability to reuse a known ontology
2. To follow the current ENC database data table layout
3. Mapping to various product specifications

Features in the Coast Pilot - categories

(18) Many of the distances in this and later Chesapeake Bay chapters are given in nautical miles above the **Virginia Capes**, or "the **Capes**," which is a short way of referring to a line from Cape Charles Light to Cape Henry Light.

(19) **Cape Charles**, on the north side of the entrance, is low and bare, but the land back of it is high and wooded. **Wise Point** is the most southerly mainland tip of the cape. Low **Fishermans Island**, a National Wildlife Refuge, is 1 mile south of Wise Point.

(20) The southwest end of **Smith Island** is 2.4 miles eastward of Wise Point; the island is 6 miles long, low and sparsely wooded and awash at half tide midway along its length.

(21) **Cape Charles Light** (37°07'23"N., 75°54'23"W.), 180 feet above the water, is shown from an octagonal, pyramidal skeleton tower, upper part black and lower

ID	object name	category
1	Virginia Capes	NaturalFeaturesA
2	The Capes	NaturalFeaturesA
3	Cape Charles	NaturalFeaturesA
4	Wise Point	NaturalFeaturesA
5	Fishermans Island	NaturalFeaturesA
6	Smith Island	NaturalFeaturesA
7	Cape Charles Light	AidsToNavigationP

The categories of the features tell which ENC table to get the associated data.

Features in the Coast Pilot

(18) Many of the distances in this and later Chesapeake Bay chapters are given in nautical miles above the **Virginia Capes**, or "the **Capes**," which is a short way of referring to a line from Cape Charles Light to Cape Henry Light.

(19) **Cape Charles**, on the north side of the entrance, is low and bare, but the land back of it is high and wooded. **Wise Point** is the most southerly mainland tip of the cape. Low **Fishermans Island**, a National Wildlife Refuge, is 1 mile south of Wise Point.

(20) The southwest end of **Smith Island** is 2.4 miles eastward of Wise Point; the island is 6 miles long, low and sparsely wooded and awash at half tide midway along its length.

(21) **Cape Charles Light** (37°07'23"N., 75°54'23"W.), 180 feet above the water, is shown from an octagonal, pyramidal skeleton tower, upper part black and lower

ID	GUID/MRN	OBJNAM	Category
1	?	Virginia Capes	NaturalFeaturesA
2	?	The Capes	NaturalFeaturesA
3	{BA862B7F-24D0-4125-B1CB-B490D71065D2}	Cape Charles	NaturalFeaturesA
4	?	Wise Point	NaturalFeaturesA
5	{1093005A-3CFE-4AEB-9A99-FE0C1970D76F}	Fishermans Island	NaturalFeaturesA
6	{A53D79BC-7196-457F-900D-909F51CB3FF3}	Smith Island	NaturalFeaturesA
7	{2A240B06-413A-43B8-8A1E-90F74D0DBAA9}	Cape Charles Light	AidsToNavigationP

And since we are looking into the ENC database, might as well save the unique identifier for it….GUID for now…eventually the IHO MRN (Marine Resource Name)..I'll talk about that soon.

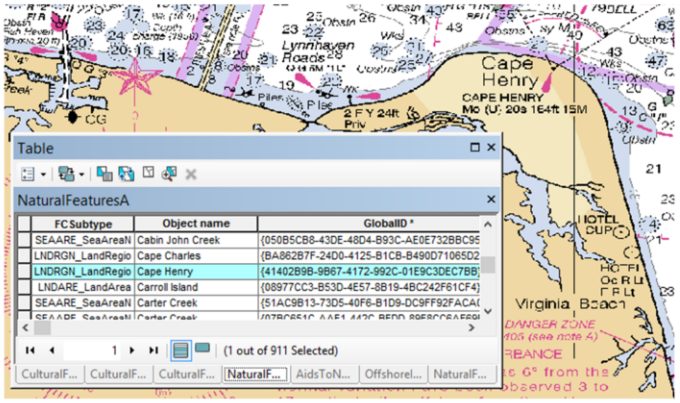
Data Structures

REVERSE ENGINEERING

Making the data more machine readable is the goal…this will be the transition to making the data in the publications more “multi-beam”-esque.

CP features to ENC features and XML

(26) **Cape Henry**, on the south side of the entrance, has a range of sand hills about 80 feet high.




```
<paragraph no="26">
   <LNDRGN type="naturalFeature"
class="area" guid="{41402B9B-9B67-4172-
992C-01E9C3DEC7BB}">Cape
Henry</LNDRGN>,
            on the  south side of the entrance,
has a range of sand hills about 80 feet high.
</paragraph>
```

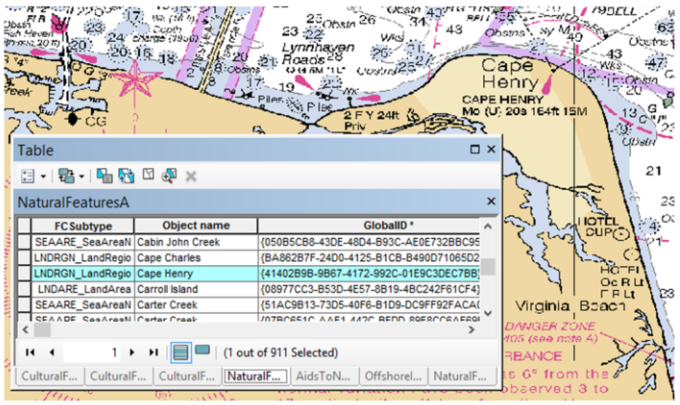
A SIDE NOTE: Using the ARCMap interface with the ENC data it is simple to find the feature in question that matches the text.

To begin the process of creating CP data structures it helps to tag the feature with the same system that is used in the database so in this example the Feature Catalog (FC) subtype is used as the XML tag and the type and class combined would be the table name in the database. Then the guid is stored to connect to the ENC data.

Coding Note: these attributes could also be elements in XML….but this is just an iteration on the data structure, it will be refined each iteration.

CP features to ENC features and XML

(26) **Cape Henry**, on the south side of the entrance, has a range of sand hills about 80 feet high.




```
<feature para_no="26">
  <LNDRGN
      type="naturalFeature"
      class="area"
      guid="{41402B9B-9B67-4172-992C-
            01E9C3DEC7BB}">Cape Henry
  </LNDRGN>,
  <spatialRelationship>on the
      <CARDIR id="9">south</CARDIR>
      side of
  <featureAssociation>
      the entrance</featureAssociation>
  </spatialRelationship>,
  <landFeature type="sandHill">
      <height unit="feet"
          accuracy="about">80
      </height>
      <quanity>a range of</quanity>
  </landFeature>
</feature>
```

In the XML workshop we also discussed the possibility of finding all the *atomic elements,* that would be data that is easy to define with a tag.
Iterations on this paragraph yielded a Land Region, a cardinal direction and a distance.

Note: everything in the first iteration is attempting to find overlap in other domains to enable reuse and see where elements might need to be extended from the original domain, or added to an appropriate NIPWG PS.

A second pass would result with more complex attributes and identifying core data structures for features as well as establishing feature associations.

Remember, we are still trying to maintain a layout that can be displayed in paragraph format…if it is possible, either way we need to see where it is and isn't feasible based on the data structure.

CP features to ENC features and XML

Steps to build data structure

1. Find the atomic elements (aka simple attributes)
2. Identify main theme of the paragraph
3. Refine the details within the structures (aka complex attributes)
4. Create feature associations
5. Normalize the tags used (attempting to reuse what is already used elsewhere)

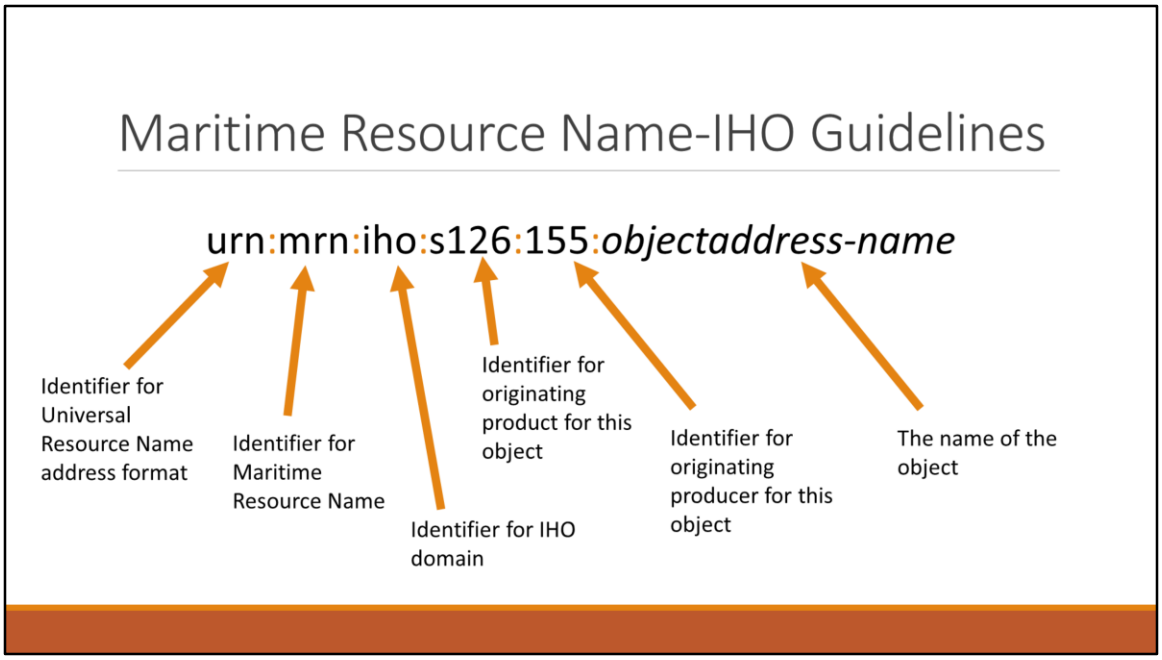
Reminders:

1. The Atomic elements are simple items to identify (regex): distances, times, months, seasons, addresses, urls, phone/fax numbers, etc.
2. The main themes should follow the Chart 1 ontology to link with the ENC
3. Refinement of the details includes attempting to tag everything, linking with ENC data via a unique persistent ID
4. Feature association examples can be seen in the S-122. This allows one feature to show its relation to another feature, so far typically this relationship is spatial for situational awareness.
5. This step ensures that all the like structures follow the same pattern and is necessary for consistency…Ideally, these are checked against the IHO register to prepare for compatibility with S-100 feature catalogs.

MRN

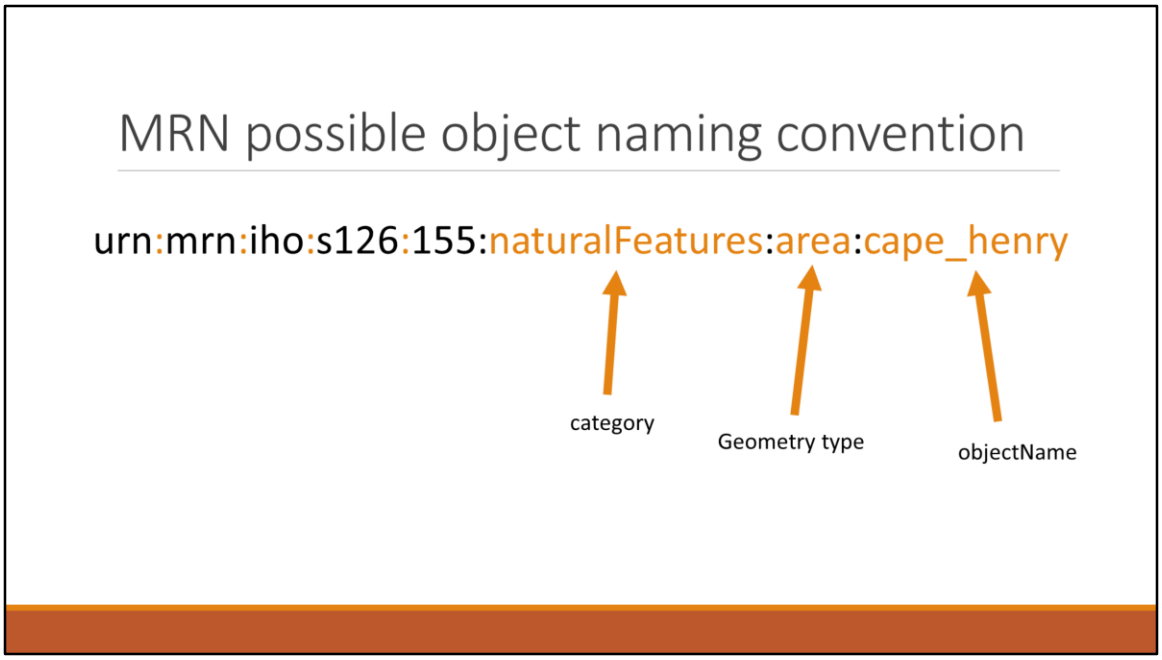
HOW TO BUILD IT AND USE IT

Now to dive a bit deeper into the MRN issue that came up.



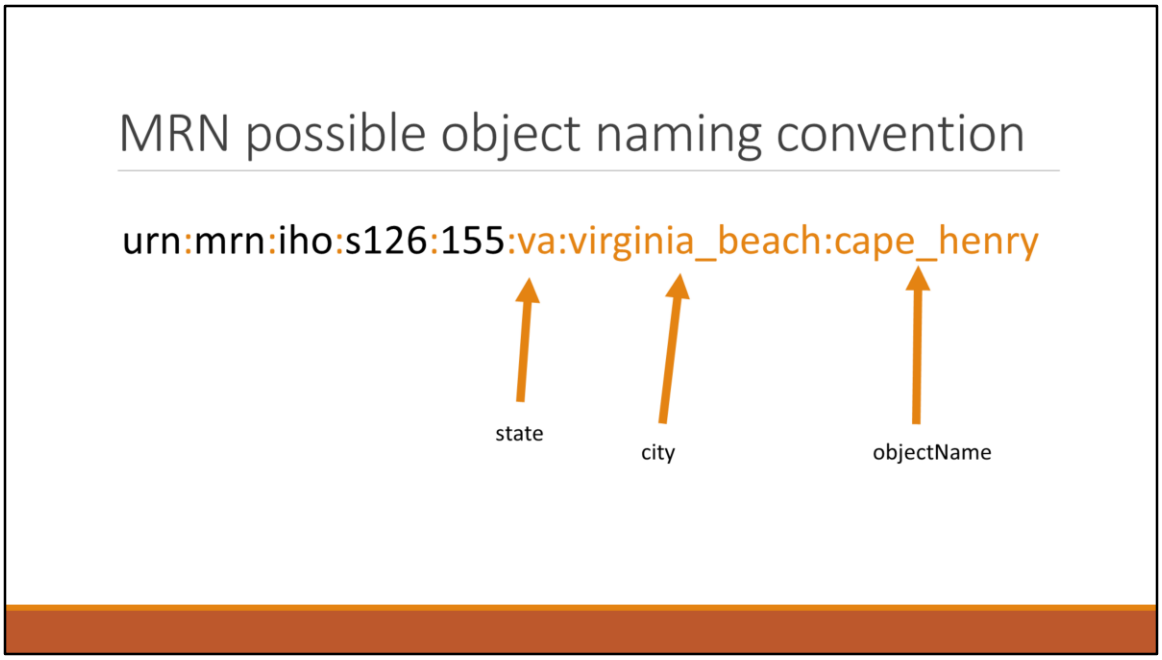
Note:
155 is the producer code for the Office of Coast Survey in the US therefore US is implied in this address.

It was thought that the object address name would be easily defined by it's location in the database…



So Cape Henry, for example, would be something like: category/table, geometry type (to distinguish between the tables of same category), object name.

However, in changing environments with standards and organization name changes and the like, It was thought that maybe the object address name would be defined by it's real world location instead….



So Cape Henry, for example, would be something like: state, city, object name

This pattern was founded on the idea to keep it as AGNOSTIC to changing standards as much as possible.
We realized there would be a few areas that might pose problems as to actual boundaries but felt this was a promising way forward, at least for FEATURES.

It was also thought that perhaps scale independent features (e.g., buoys) the naming should be “real-world based”, whereas for the scale dependent (meaning features that change from scale to scale and thus from chart to chart, e.g. coastlines and depth curves/areas) could –and probably- should be “product based”.

…but this was just the beginning of a long discussion.

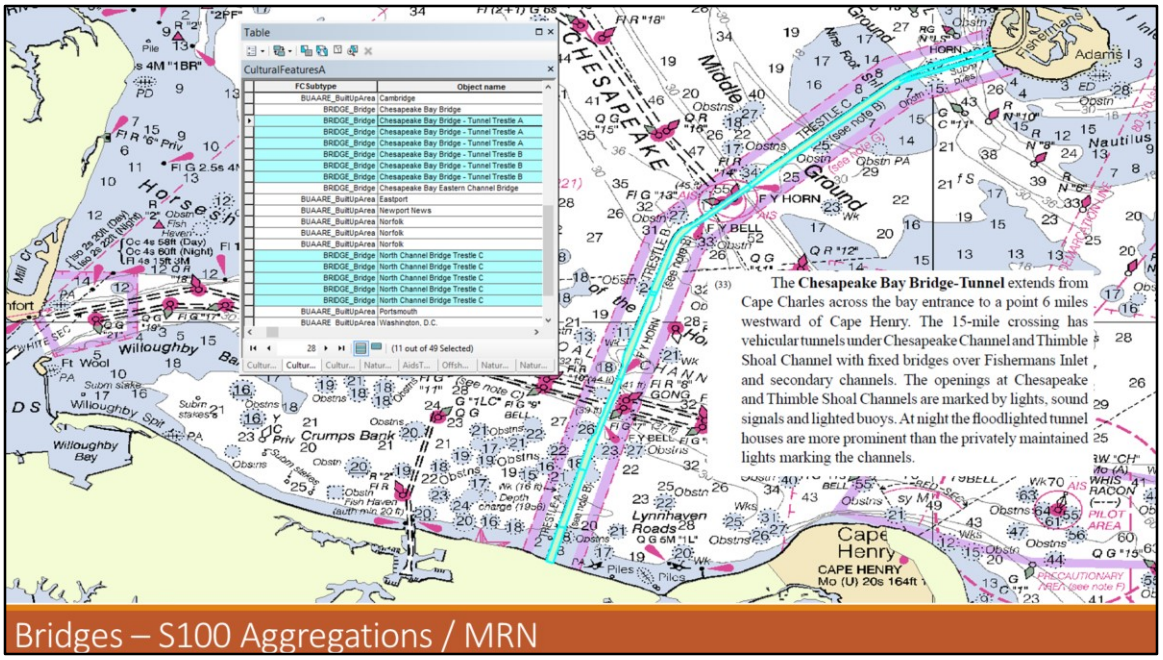
Of course, there are other objects like NavAids and bridges that pose a different set of issues related to location, name and ownership

Bridges Issue

KEEPING TRACK OF AGGREGATIONS

During the workshop it was evident the marine-related structures along the waterways already had a solid data structure for imparting the information because they were in TABLE format. …A “low-hanging” fruit, we all thought!

However, diving deeper into how bridges are stored on the ENC side of things opened the publication groups eyes to some possible logistical problems.



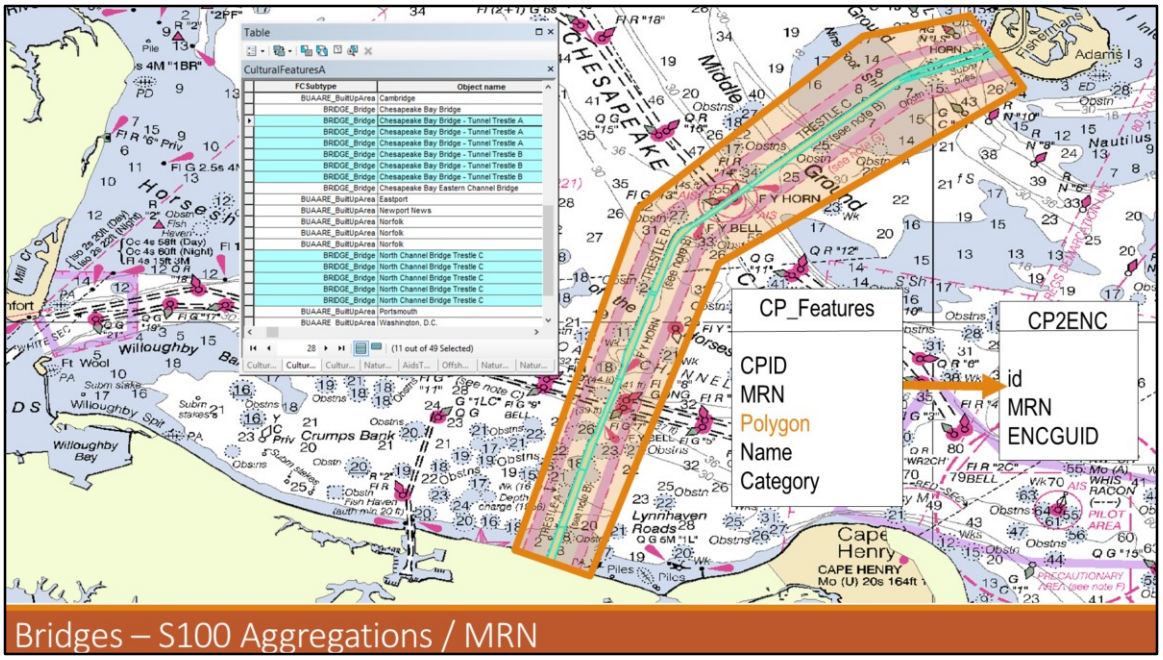
Bridges – S100 Aggregations / MRN

Bridges are composed of spans in the S101.
Even though the US ENC database is based on the S57, these “spans” can clearly be seen as making up this one long bridge across the channel.

This old system needs to be converted to the new system eventually, but dealing with an interim solution now lays the foundation for the future.

This is important to the supporting textual data in the Coast Pilot because the CP doesn’t know of aggregations, it only has a common name identifying the length of the structure.
So how do we get one MRN to represent all the spans before an aggregation is implemented in the system?

Another BIG problem with this system is that there are no PERSISTENT unique id’s for these spans. The spans can be split, deleted or added to the structure. In this case, the associated GUID is deleted along with the span.



Bridges – S100 Aggregations / MRN

The solution decided upon is to use a polygon to encapsulate all the elements that make up the object add that to the database table discussed earlier.

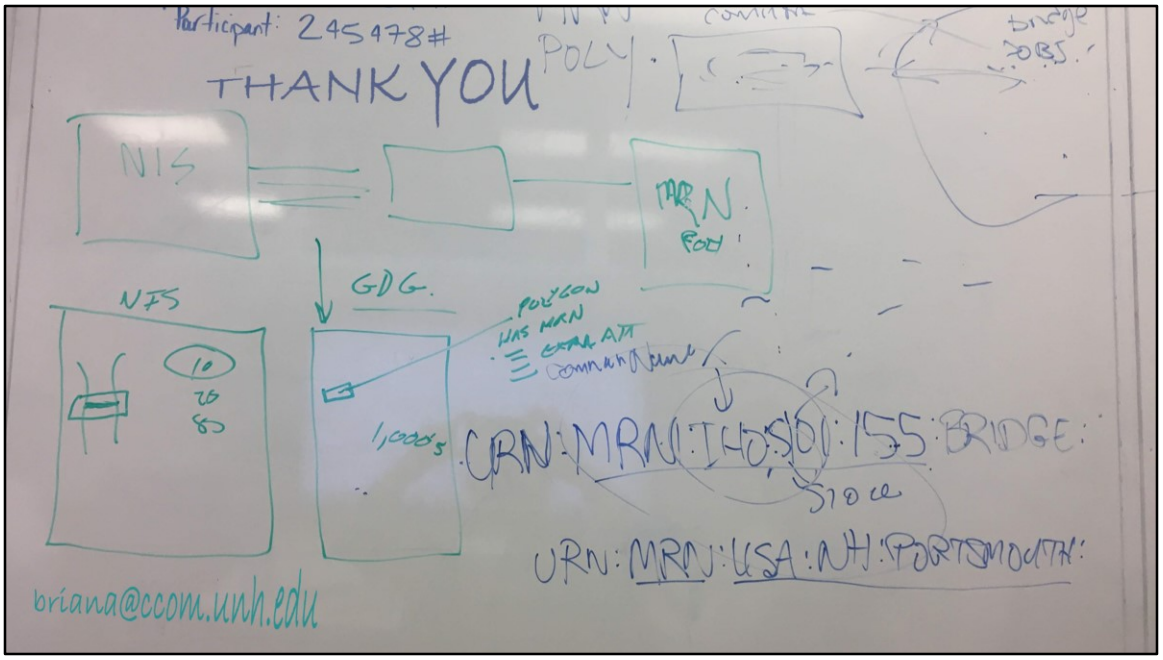
Then create another (many-to-many) table that would store the MRN associated to that polygon and match it up with all the Global Unique Identifiers (GUID's) from the database.

This solution takes care of the issue of deleting or adding spans to the structure, by then using the polygon to find all span objects within it and consider them all the same MRN. (updating the associated table) Changes would also be recorded in this system.

Workshop deliverables by 2020

1. Feature identification and storage for all text
2. System to automatically generate feature polygons
3. Test created to find specific feature types within polygons
4. Chart One Database Table – to assign categories features
5. MRN auto generation

1. A mechanism for identifying all geo-referenced features in the CP and collating them into a database table that contains a MRN and location/polygon bounds for that feature.
2. An interface to allow a human-in-the-loop to create or verify
3. Access from NSD to the ENC database will be granted and used for a hit test to determine feature object within the ENC that coincide with the CP features, which will then store feature id's from the ENC into the common CP features to ENC features table
4. A table to represent the elements and hierarchy of the Chart One publication to enable machine readable chart one data files as well as to assist in categorizing the CP data/features.
5. A convention for automatically generating persistent ID's that will be used as link to any new identifying system in the future, e.g. MRN's (once the IHO has stabilized it's recommended structure and uses)



briana sullivan
Univeristy of New Hampshire, Durham, USA
briana@ccom.unh.edu
vislab-ccom.unh.edu/~briana