VAST 2024 MC2 Data Description

Data dictionary and nodes for MC2:

Graph Description:

- Directed multi-graph, allowing multiple edges between nodes
- 5637 nodes
- 271752 edges
- 1 (weakly) connected component
- The graph format is a JSON format generated by Python's [network.node_link_data()](#) function. It can likewise be loaded to a networkx object using the corresponding [node_link_graph()](#) function. The root-level JSON object consists of graph-level properties specifying that it is directed and a multigraph, a "nodes" key which holds the list of nodes, and a "links" key which holds the list of edges.

TA2 is a composite of graphs from several sources. Each source provides nodes and edges of different types (described below). There is overlap in entities and types between the sources, forming one connected graph.

Vessel Movements: Oceanus is outfitted with a transponder/ping system named the Oceanus Vessel Locator System (OVLS). Vessels are outfitted with a transponder and periodic 'pings' from base-stations results in a report of vessel locations at any time. The raw ping granularity is at the minute-level but post-processing has converted it into visit/dwell times. OVLS is generally reliable, though vessel records may be missing for a variety of reasons.

Node/Edge types and properties present

- Entity.Vessel: Description of the vessel
- Entity.Location: Description of a geographic location
- Event.TransponderPing: Links a vessel to a location

Harbor Reports: Harbor masters regularly report the vessels found in their purview anytime during the day. This data is derived from a different system than OVLS (see "Vessel Movements"), though the data overlaps. Harbor Reports are provided on a different schedule from different harbors. Since no harbor reports every day, this data has lower temporal granularity than vessel movement data. Additionally, the Harbor Master is also responsible for proximate navigational beacon(s), so this data has lower spatial granularity as well. However, the list of vessels observed is considered canonical.

 Node/edge types present:

- Entity.Vessel
- Entity.location
- Event.HarborReport

Harbor Import Records: Vessels deliver cargo to the ports, and that cargo is brought into Oceanus. These records reflect the goods that *leave* the harbor to go to businesses in Oceanus or to be exported. It was filtered pre-ingest to focus on the delivery of raw fish. Because it is raw, fish leave the

port quickly (generally one day after delivery). Due to clerical error, the records purchased by FishEye do not include the vessel that delivered the cargo.

Node/Edge types present:

- Entity.location
- Entity.Commodity.Fish
- Entity.Document.DeliveryReport
- Event.Transaction

Type descriptions with attributes:

Note: There are many different sub-types in this data. Some subtypes have different properties. The property lists below do not distinguish the subtypes. If a field is not present on a node/edge, it may not apply to that subtype.

Node types:

- Entity.Vessel: Description of the vessel
    - Name
    - company: Company that owns the vessel
    - flag_country: Country the vessel is licensed in
    - length_overall: meters length of the vessel
    - tonnage: Gross tonnage of the vessel
- Entity.Location: Description of a geographic location
    - Name
    - Activities: List of common activities in the region
    - Description: Textual description
    - fish_species_present: List of fish common found in that area
    - kind: High-level grouping (city/preserve/etc.)
- Entity.Commodity.Fish
    - Name: Type of fish
- Entity.Document.DeliveryReport: A report on the import itself

Edge Types:

- Event.Transaction**:** Links related two parts of a transaction (location & commodity)
- Event.HarborReport: Links a vessel to a location via a harbor report
- Event.TransponderPing: Links a vessel to a location via OVLS
    - time: Start time of the visit
    - dwell: How long was the vessel in this location

Metadata attributes: Nodes and edges from all data sources may also have these metadata fields related to information handling at FishEye.

- **type** – Indicates the node/edge type (as described above)
- **_last_edited_by** – The name of the user who last modified this node/edge
- **_last_edited_date** – The last time this node/edge was modified
- **_date_added** – The day on which this node/edge was first added to the graph
- **_raw_source** – The source from which the information was originally obtained
- **_algorithm** – The method in which the information was obtained (for this mini-challenge, either automatically imported from pre-existing databases or manually updated by FishEye analysts)