**VAST 2023 MC1 Data Notes**

**Data dictionary and nodes for MC1:**

**Graph Description:** The challenge problem KG is generated using Natural Language Processing (NLP) tools extracting entities and relationships from a set of text articles.

* 3721 nodes
* 7422 edges
* Possible Node types are: {person, organization, company, political\_organization, location, vessel, event, movement}
* Possible Edge types are: {membership, partnership, ownership, family\_relationship}
* This is a directed multi-graph, so multiple edges between the same two nodes are possible
* The graph format is a json format intended to match d3's node-link format and be compatible with networkx.node\_link\_graph. At the root-level, it is a dictionary with graph-level properties specified as keys ( directed , mulitgraph , graph ). The nodes and links keys each provide a dictionary of the nodes and links respectively. The nodes entries that must include an id key that is unique for each node. The links entries include source and target keys that refer to node id values. All other keys provided in node and link dictionaries are attributes for that node or link.

**Node Attributes:**

* Id – Identifier of the node is also usually the name of the entity. Some nodes have a numeric ID and do not have a name even if they are a person, company, or organization.
* type—Type of the node as defined above. It is an optional attribute
* country—Country associated with the entity. It is an optional attribute.
* dataset – Always ‘MC1’

**Edge Attributes:**

* source – ID of the source node
* target – ID of the target node
* type – Type of the edge as defined above. It is an optional attribute
* dataset – Always ‘MC1’

**Note: all names included in this dataset are fictional. Any resemblance to real persons, companies, or locations are purely coincidental.**