NEO4J SCHEMA

The Neo4j graph schema is depicted in Figure 1, Figure 2, and Figure 3.

Neo4j's Cypher query language syntax is used in this document. The syntax is:

- Neo4j nodes are written as camel-case expressions that begin with an upper-case character, are preceded by a colon, and embedded in parentheses, for example:

  (:Bird)
- Neo4j relationships as written as upper-case expressions that are preceded by a colon and embedded in square brackets, for example:

  [:HAS_GEOGRAPHY]

The schema is based largely on the North Carolina Partners in Flight wiki database. We took advantage of Neo4j capabilities when designing the schema.

Multiple labels for each graph node

Neo4j allows each node to be associated with one or multiple labels. Examples of labels in Figure 1 include ":Habitat" (a singular label) and ":Species:Bird" (two labels). The ability to attach multiple labels to a given node makes it possible to include (:Species:Amphibian), (:Species:Fish), (:Species:Mammals), or (:Species:Reptile) nodes in the future.

Unidirectional relationships

Consider the following example:

(:Bird:Species)-[:HAS_GEOGRAPHY]->(:Geography:State)

The Neo4j database currently includes symmetrical relationships in the form of:

(:Geography:State)-[:HAS_GEOGRAPHY]->(:Bird:Species)

The latter symmetrical relationship is unnecessary and not depicted in the following pages, even though the symmetrical relationships exist in the Neo4j database. Symmetrical relationships need not be explicitly asserted in this particular (and other analogous) case because:

(1) Given any (:Geography:State) node, the Neo4j Cypher query is capable of retrieving a (:Bird:Species) node. This is despite the cardinality (direction) of the [:HAS_GEOGRAPHY] relationship being explicitly asserted that originates from (:Bird:Species) to (:Geography:State). The symmetrical relationship is therefore unnecessary and makes it harder to ensure that the database exhibits relational integrity. That is, over time as the database evolves, it is harder to ascertain that there is a [:HAS_GEOGRAPHY] relationship going in both directions for every pair of nodes that is related by the [:HAS_GEOGRAPHY] relationship.
(2) Conceptually, one might state, as a narrative, that "a bird has a certain geography that is this US state), but not "this US state has a geography that is this bird".

The symmetrical relationships are not currently used in Neo4j Cypher queries. To ease maintenance overheads, these relationships will be removed in the next major version of the Neo4j database.

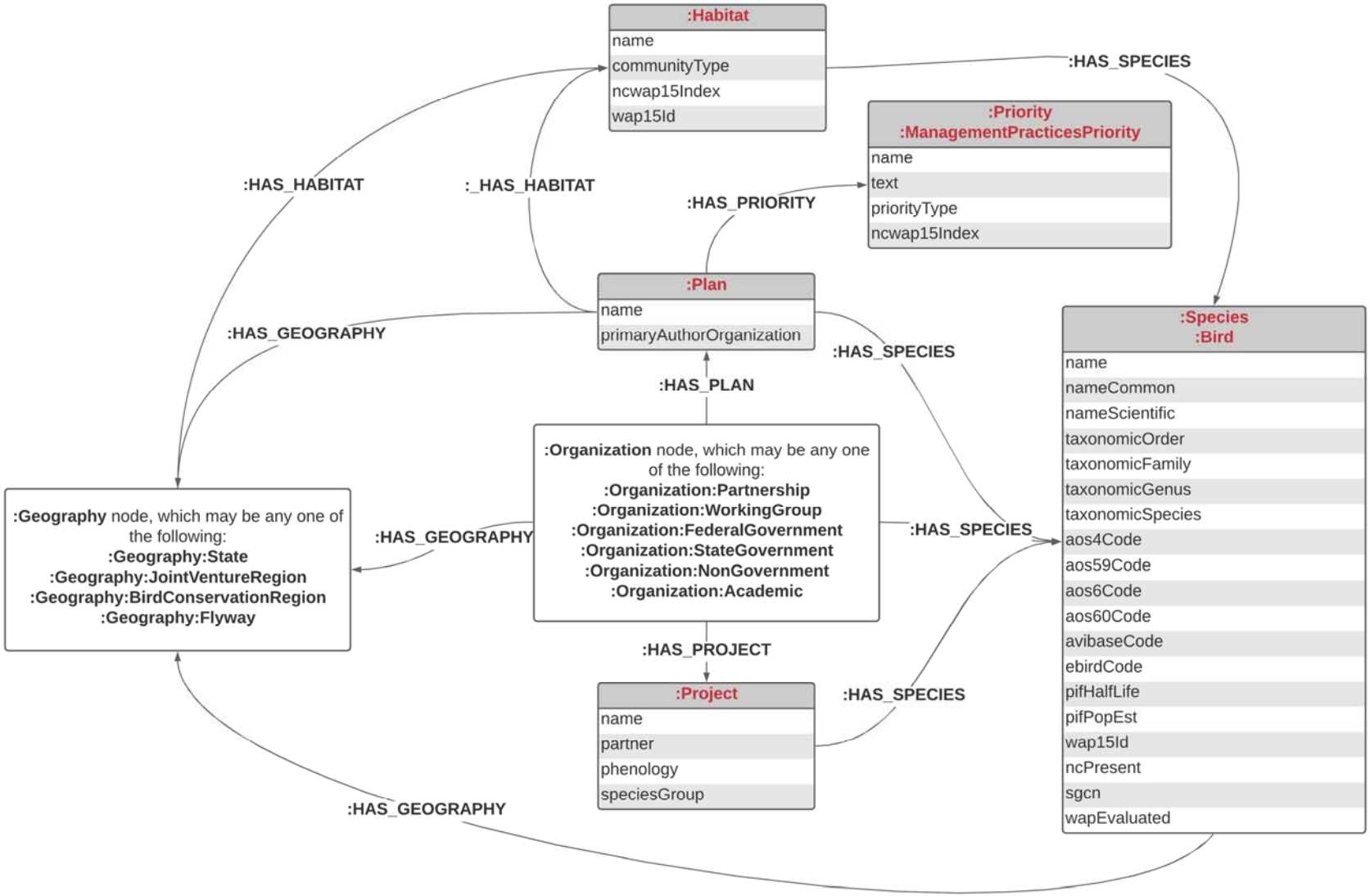

Figure 1: Top level Neo4j schema.

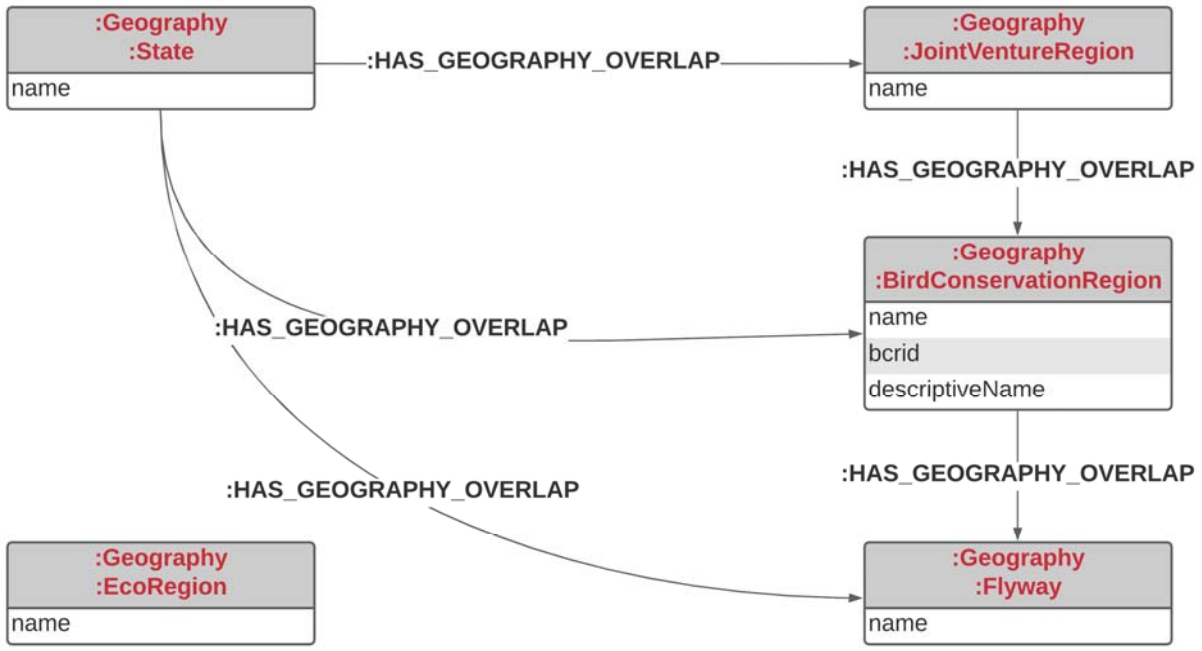

Figure 2: Neo4j schema for nodes with the :Geography label.

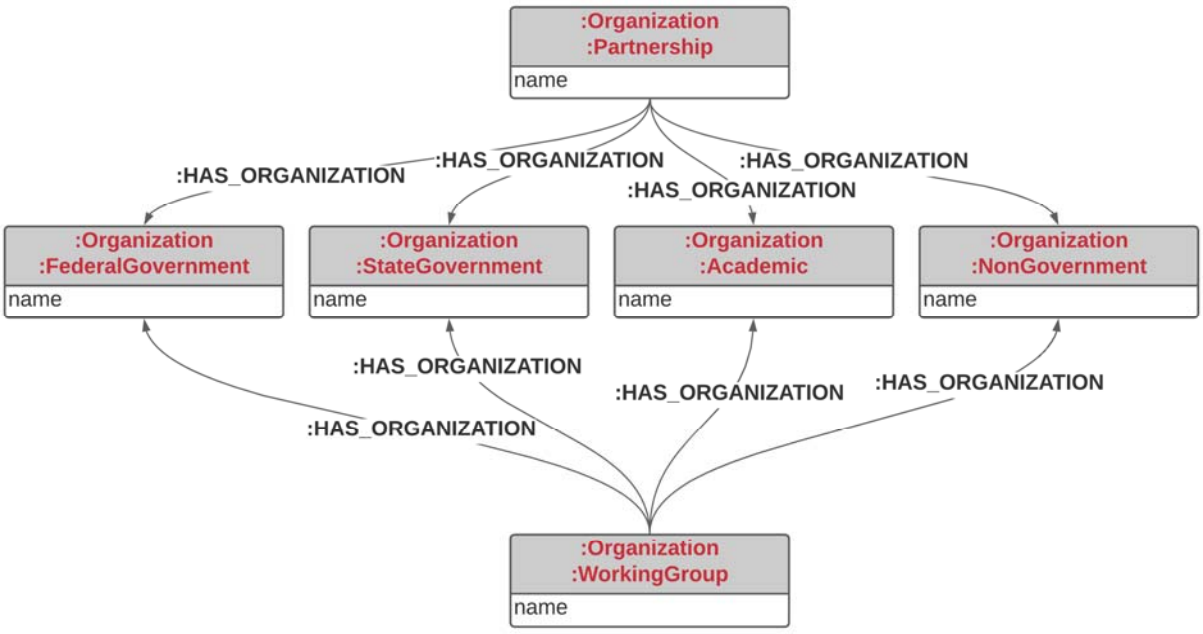

Figure 3: Neo4j schema for nodes with the :Organization label.