Modeling data and information needs for avian conservation using Neo4j (2020 ESIP Lab Project)

Brian Wee[1], Scott Anderson[2], Jessica Burnett[3], Steven Aulenbach[3], William Teng[4], Rustem Albayrak[4]

The Challenge

Conservation organizations set priorities as a balance of their overall mission and the larger conservation needs. As we craft conservation strategies across ever-larger landscape scales, we need to understand and incorporate conservation priorities that span different organizations, geographies, management plans, and jurisdictional boundaries.

An integrative assessment of the linkages among organizational conservation priorities provides opportunities for effective planning to increase the impact of conservation investments. This project is an experiment in that type of integrative assessment.

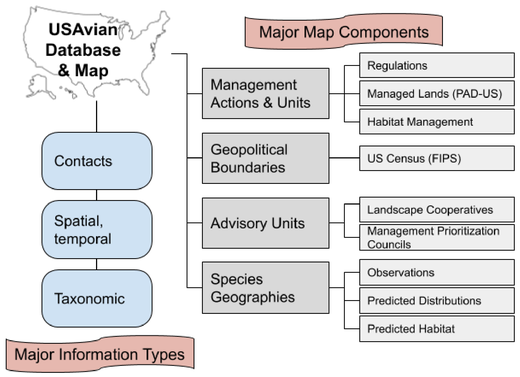
ESIP Lab Project Objectives

Graph databases are increasingly used by the private-sector to analyze and act on the connections between customers, their needs, and products. Here, we use the same approach to better link actors and information within the bird conservation community. The objectives include:

- Explore the use of graph databases for connecting data, information, and knowledge relevant to conservation planning and management activities.
- Deliver a prototype that provides a way for both data scientists and conservation practitioners to work together in real-time and in a collaborative environment, where management decisions are supported by data-driven visualization.

The USAvian Concept

Born from unsuccessful attempts to build a comprehensive and intuitive mental map of how various organizations, research groups, working groups, and state and federal agencies are implementing conservation science or delivery, USAvian is an attempt to develop a data-driven resource for 'mapping' the complex ecosystem of entities in the North American bird conservation community. Using the concept of USAvian, this ESIP lab project employs graph database technology to connect bird conservation priorities and taxonomic information across management scales.



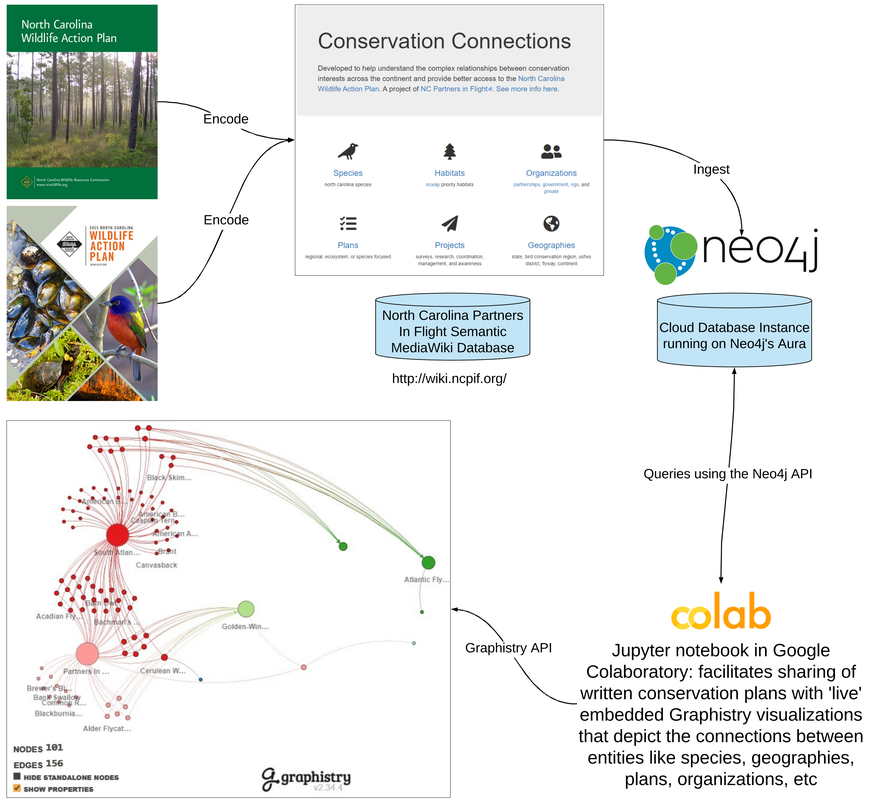
Long-term Objectives

- Provide a framework for connecting conservation priorities within and across conservation-relevant scales.
- Streamline and improve the FAIRness (**F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets) of the processes for populating and writing SWAPs (**S**tate **W**ildlife **A**ction **P**lan)
- Synthesize data and information sources to populate locally- and contextually-relevant tools, like fact sheets, or species summary documents.
- Use the database as an exploratory research tool for identifying temporal and spatial patterns in conservation management and planning activities

Conceptual Approach

North Carolina's Wildlife Action Plans (for the years 2005 and 2015), already encoded in an existing database, were ported to the Neo4j labeled property graph database. A Jupyter notebook was used to execute queries to the Neo4j database. Query results, embedded within the notebook, were visualized through a Graphistry cloud service API.

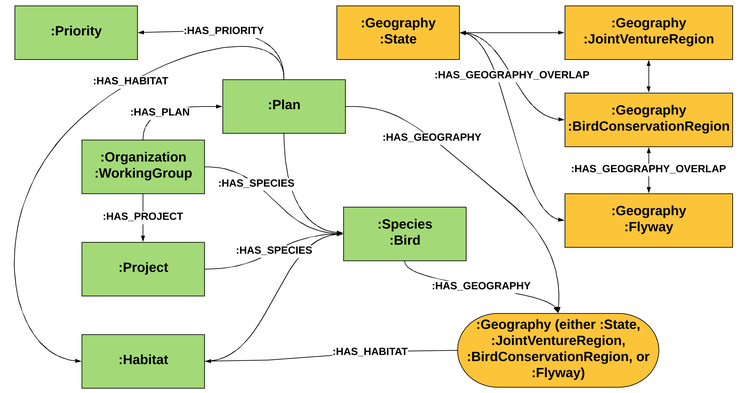
The vision is to use Jupyter notebooks as intermediate work-products that inform the formulation of SWAPs. Our approach uses Google Colaboratory (Colab), which is a free, cloud-based resource that hosts Jupyter notebooks. Colab notebooks can be shared between members of the team: some focused on drafting notebook sections that focus on strategy and policy for conservation action, while others focused on adjacent sections that visualize connections between conservation entities.



Graph Database Schema

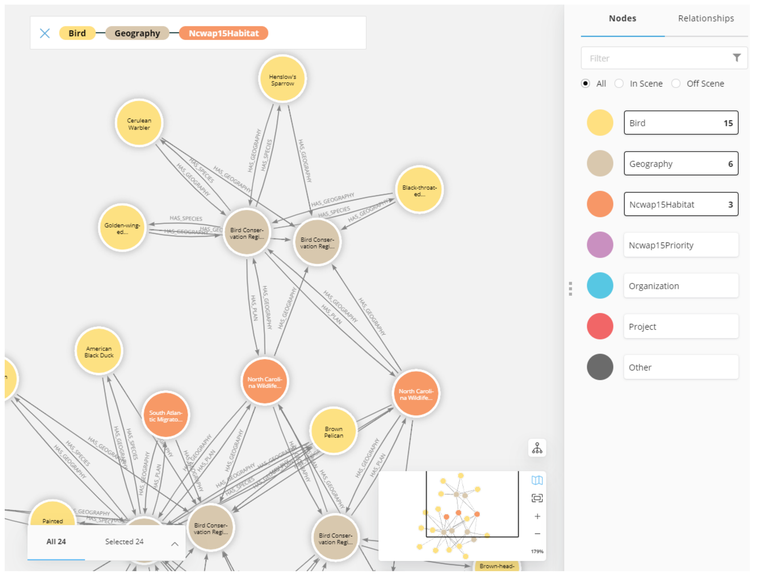
The high-level graph schema below includes selected graph database nodes and relationships. Neo4j nodes and relationships can be populated with properties (i.e. key-value pairs: not shown). In Neo4j, each node can be associated with one or more label(s). Examples of labels are ":Species", ":Priority", and ":Geography".

The graph database can be queried to determine if there is a connection between any two nodes in the graph database, regardless of the number of intermediate nodes connected through relationships. The database engine traverses just as easily 'forward' in the direction of the asserted relationship as it does in the opposite direction. Connectivity between two nodes can be ascertained through a simple database query.



Visualizing Graphs in the Database

The visualization below was produced using the Neo4j Bloom app, which is packaged with the free desktop version of the database. The graph below shows the connection between bird species, geography, and habitats.



[1] Massive Connections, LLC | [2] NC Wildlife Resources Commission | [3] United States Geological Survey | [4] NASA Goddard Space Flight Center (ADNET Systems)

This work is based on funding provided by the ESIP Lab with support from the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and the United States Geologic Survey (USGS).