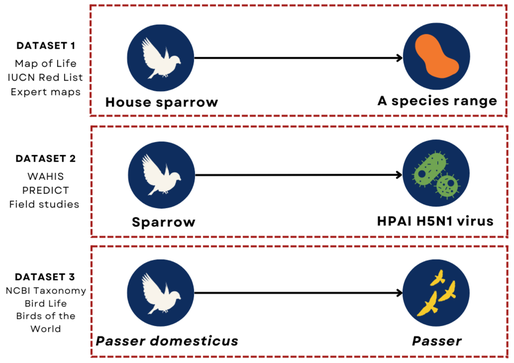
# Knowledge Graphs for Scalable Data Integration: A Case Study of Highly Pathogenic Avian Influenza (HPAI)

Hailey Robertson, Ellie Graeden[1], Adrian Castellanos, David Rosado, John Drake, Barbara Han
Correspondence: ellie.graeden@georgetown.edu

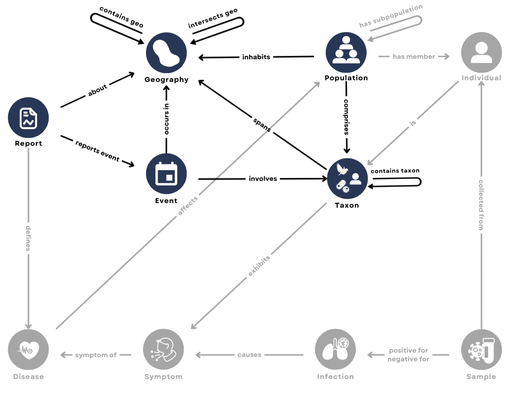
## Introduction

The growing occurrence of zoonotic disease outbreaks underscores the interconnectedness of animal and human networks. Researchers have developed numerous mechanistic, statistical, and machine learning models to capture the multifaceted and dynamic nature of these systems; however, there remains a gap in methods to bridge data and models across disciplines, in part due to the lack of interoperability between datasets. In this context, knowledge graphs offer a powerful solution for reconciliation, integration, and synthesis of diverse data that can support end-to-end modeling between human behavior and animal zoonotic risk systems. **We demonstrate the utility of a knowledge graph to integrate laboratory, field, and historical data about highly pathogenic avian influenza (HPAI) and generate new hypotheses upon which future disease mitigation will depend.**

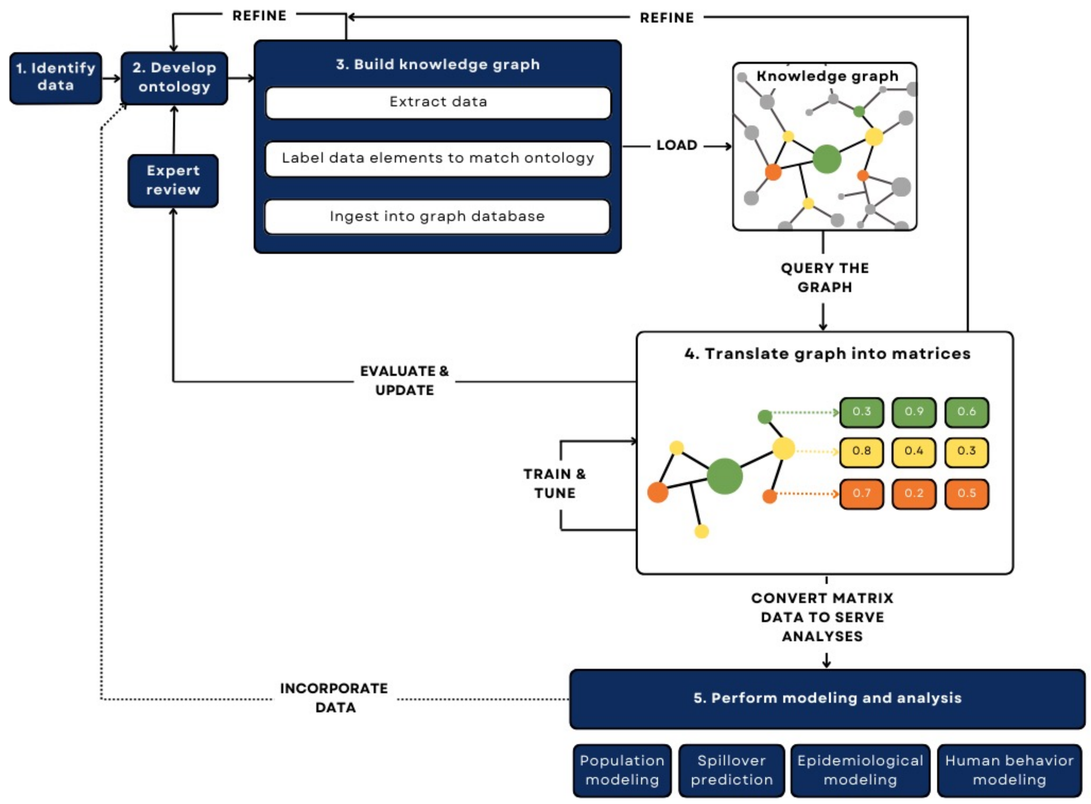
## Traditional data systems are not interoperable across sources



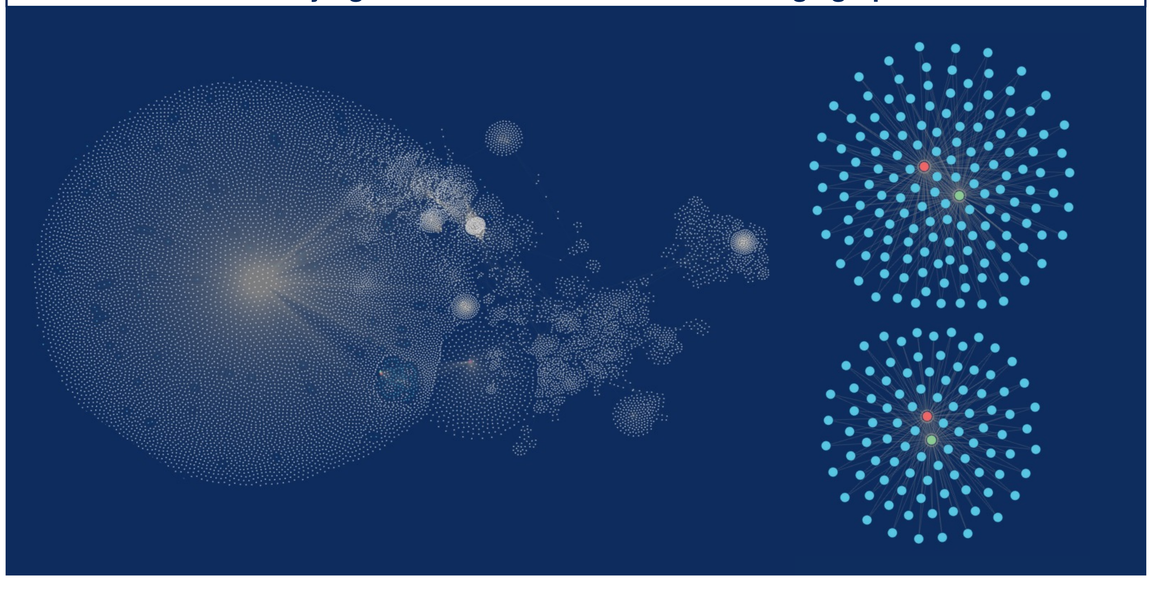

## A zoonotic disease ontology aligns data across domains based on shared elements for integration in the knowledge graph



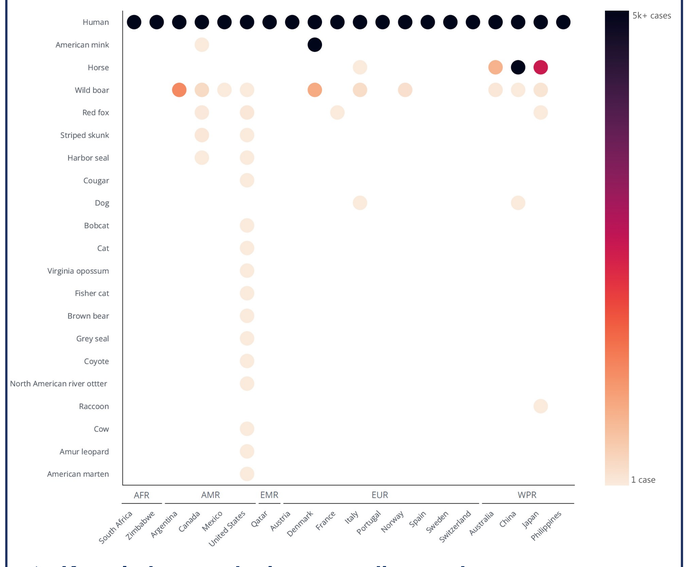

## Using knowledge graphs for broad-scale data integration




## Querying HPAI outbreaks within a knowledge graph



## Applying knowledge graphs to HPAI




- **Knowledge graphs integrate diverse data sources to identify potential sources of spillover transmission.** The knowledge graph links animal case reports (WAHIS), human epidemiological data (FluNet), geographical information (GeoNames), species distribution ranges (MDD), and taxonomic lineages (NCBI Taxonomy).
- **When applied to HPAI, we find that the knowledge graph accurately describes mammalian host range and identifies geographic variation in cross-species transmission risk.** Among the 20 countries with the highest human influenza A burden, wild boar and red fox were the top two non-domesticated mammalian species that most frequently test positive for HPAI. Graph-based summaries efficiently highlight geographic and species-specific surveillance priorities to maximize early detection and spillover response.

## ACKNOWLEDGEMENTS

This project is supported by NSF 21-590, *Heterogenous Model Integration for Infectious Disease Intelligence*

Cary Institute of Ecosystem Studies

GEORGETOWN UNIVERSITY | CENTER for GLOBAL HEALTH SCIENCE & SECURITY

Odum School of Ecology
UNIVERSITY OF GEORGIA