An enabling semantic pipeline for fisheries observational data to improve sustainable management of fisheries resources

Aileen Kennedy[(2)] *David Currie*[(1)], *Enda Howley*[(2)] *and Jim Duggan*[(2)]
(1) Marine Institute, Ireland
(2) School of Computer Science, National University of Ireland Galway, Ireland
Presenter contact details: **a.kennedy28@nuigalway.ie**

This project explores the creation of a semantic pipeline for fisheries observational data. As part of the pipeline an ontology is implemented which utilises the Observation and Measurement ontology [1]. The ontology provides a standardised framework that allows interoperability between different data sources.

Introduction
In Ireland, the Marine Institute generates and consumes data from a number of different sources. In particular, Fisheries Ecosystem Advisory Services (FEAS) maintains databases containing information on:

- catch value and volume, including commercial landings, fishing effort and fleet capacity;
- biological data – which includes data on variables such as the number, length, weight, sex and age of fish species in a given location.

Fine-scale spatio-temporal data about vessel position is also stored. The fisheries data typically becomes available in batches rather than being streamed – these batches can range in size from an individual fishing trip, up to all vessel positions in a calendar year. These databases are typically siloed, so querying for information that are stored across many databases can be a significant challenge.

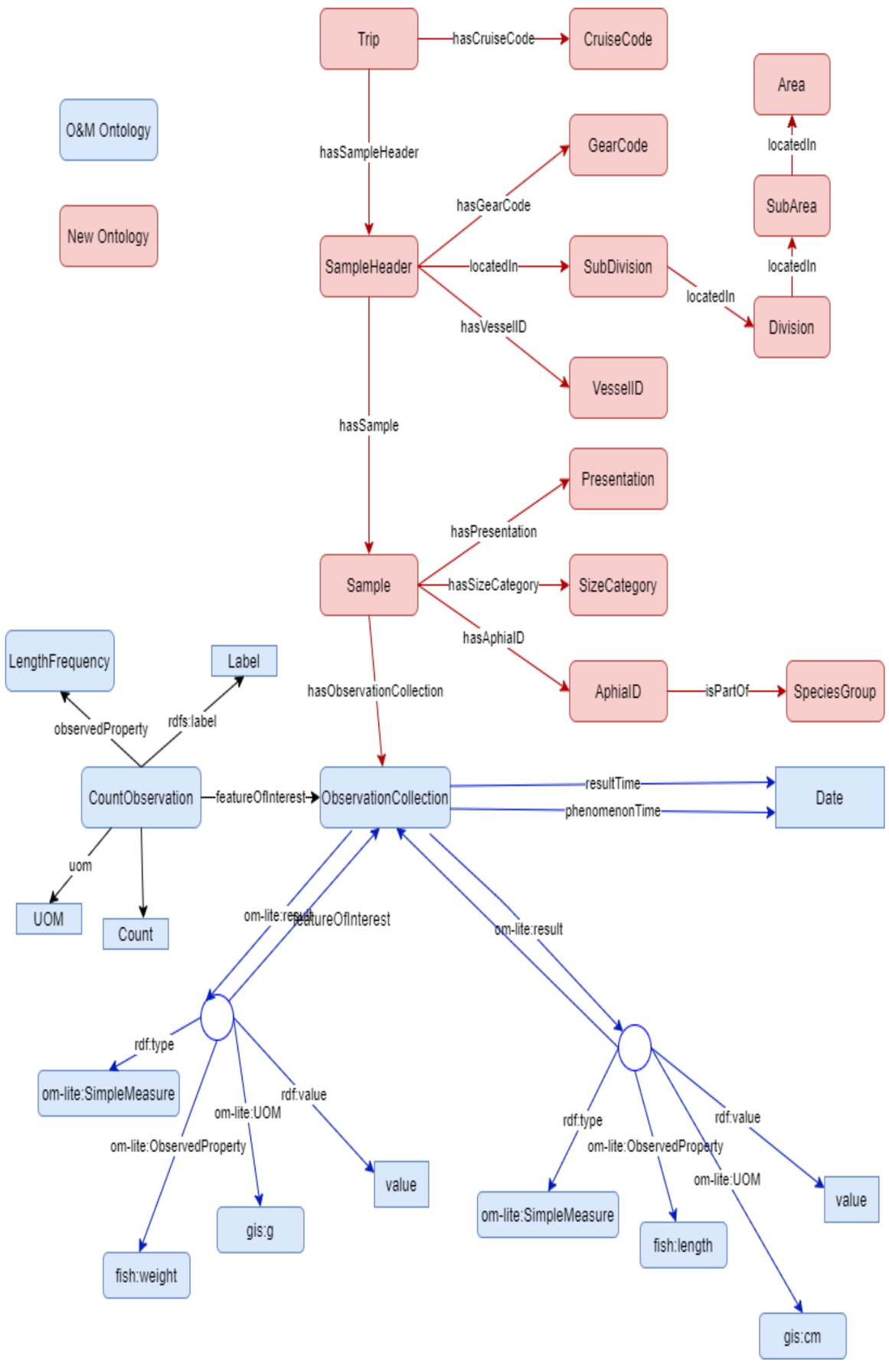

Figure 2 Ontology, blue represents classes and properties taken from the O&M ontology, while red represents new classes and properties

Ontology
Observations & Measurements (O&M) is a domain neutral international standard information model. The scope of O&M includes in-situ observations, remote sensing, ex-situ observations, numerical models and simulations, forecasts. It can include any action whose result is an estimate of the property value. The INSPIRE European standards include a Species Distribution theme in which the O&M standards have been identified as being relevant however its purpose is not to directly record observations rather aggregations of such [3]. Our approach is to implement O&M to allow the integration and interoperability of fisheries data at a detailed level [2],[4]. We also use appropriate controlled vocabularies from sources such as ICES and NERC Vocabulary Server for this reason

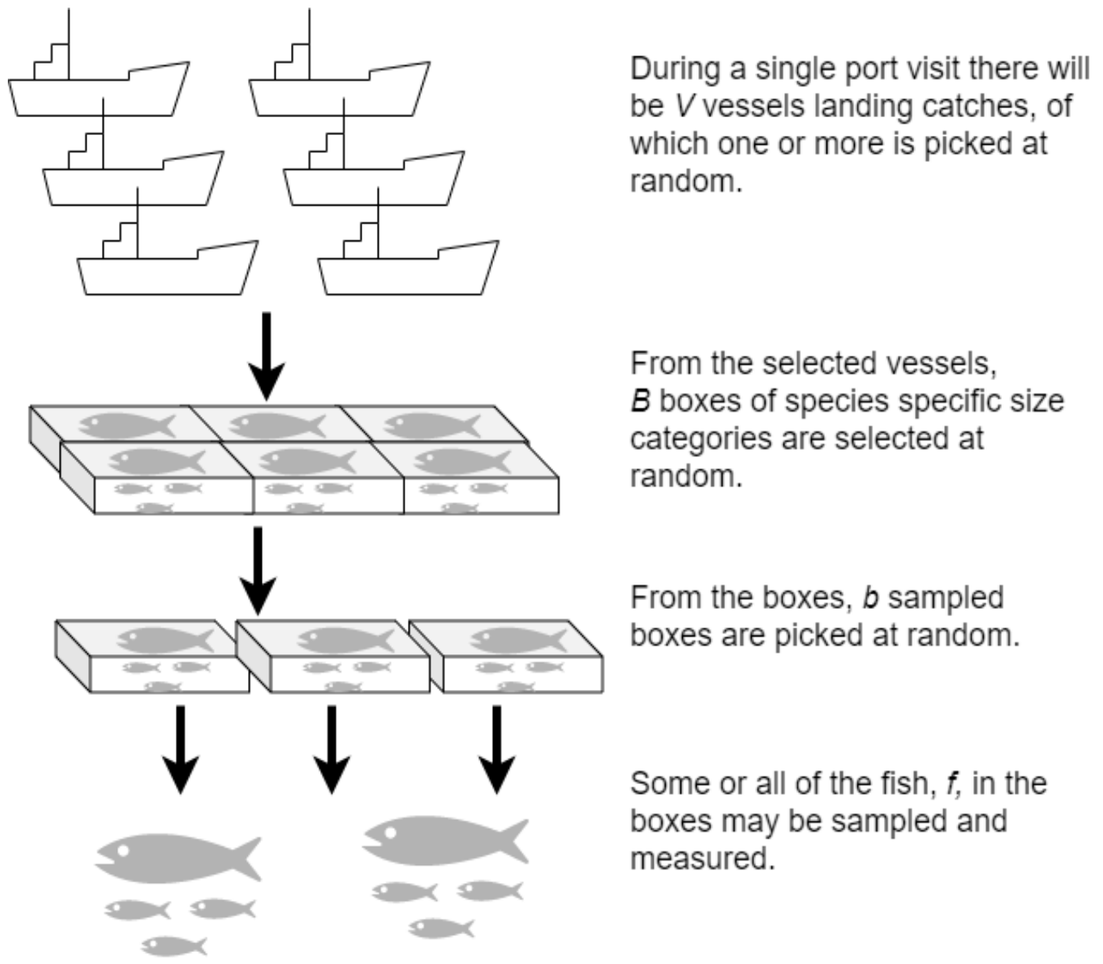

Figure 1 Sampling Process

Data
The existing data set, which is gathered according to a sampling scheme (Figure 1), is stored on a number of heterogeneous relational databases and is composed of both fine grain spatial data and coarser grained biological sampling data. The data is typically siloed which then results in querying across different databases, and this can be quite cumbersome and challenging.
Our system extracts the data from the appropriate databases using SQL. The extracted data is transformed to RDF triple format, with each triple composed of a subject, predicate and object. The structure of the triples is determined by the structure of the ontology.

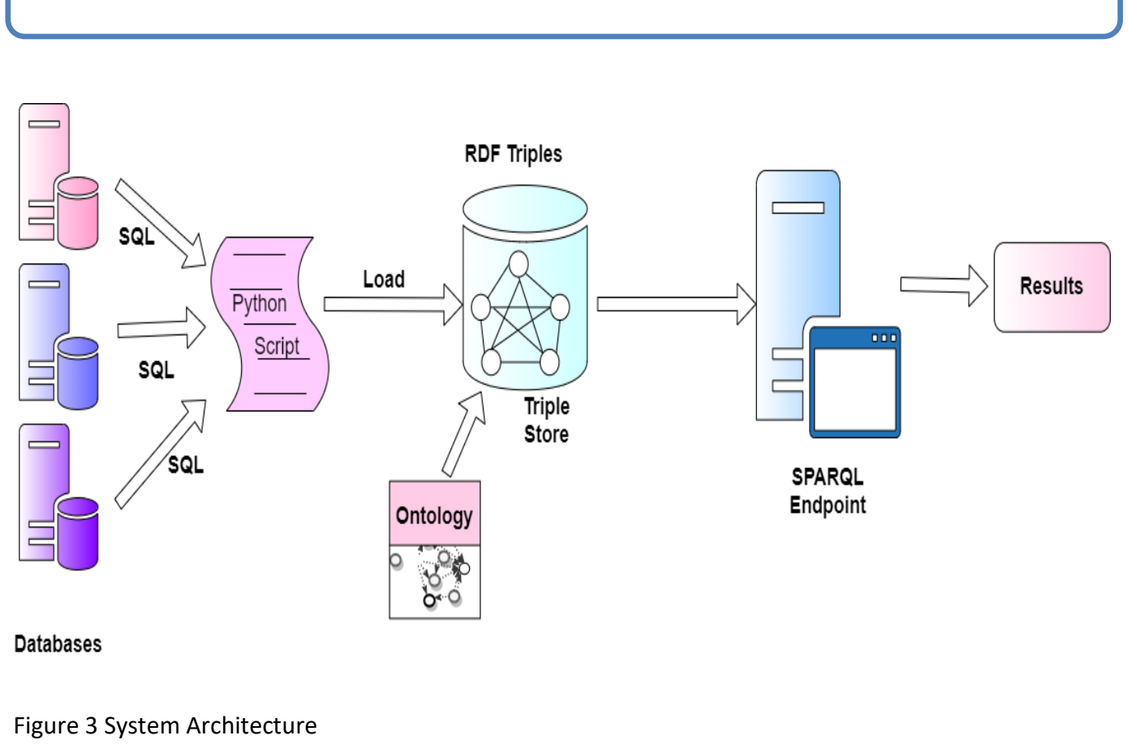

Figure 3 System Architecture

System Architecture
The technical architecture of the pipeline (figure 3) includes a number of different elements. The original data is stored in relational databases. This data is extracted and transformed to RDF using a Python script. The resulting serialised RDF data is stored in a triple store (Blazegraph). The ontology is written in OWL using Protégé and also loaded into the triple store. The triple store is queried using SPARQL, and RDF query language.

Conclusion
Future work include an investigation into the insights that can be gained through the use of inferencing on the transformed data. This may provide results that would not be possible with the underlying relational databases.
This pipeline integrates the data in a more standardised and interoperable format. This then opens opportunities for further analytics and application of machine learning techniques.

References

[1] Cox, S. J. (2017). Ontology for observations and sampling features, with alignments to existing models. Semantic Web

[2] Currie D., Howley E., and Duggan J. (2016) A Data Analytics Framework for Ecosystem-Based Fisheries Management. ICES Annual Science Conference 2016

[3]INSPIRE Thematic Working Group Species Distribution, 2013, D2.8.III.19 INSPIRE Data Specification on Species Distribution – Technical Guidelines

[4]Kennedy A., Currie D., Howley E., and Duggan J. (2018) Semantic Fisheries Data Integration

Acknowledgments

This research is funded by the Cullen Fellowship, The Marine Institute, Ireland.