Semantic Fisheries Data Integration and Analytics

Aileen Kennedy[2] *David Currie*[1], *Enda Howley*[2] *and Jim Duggan*[2]
(1) Marine Institute, Ireland
(2) Discipline of Information Technology, National University of Ireland Galway, Ireland

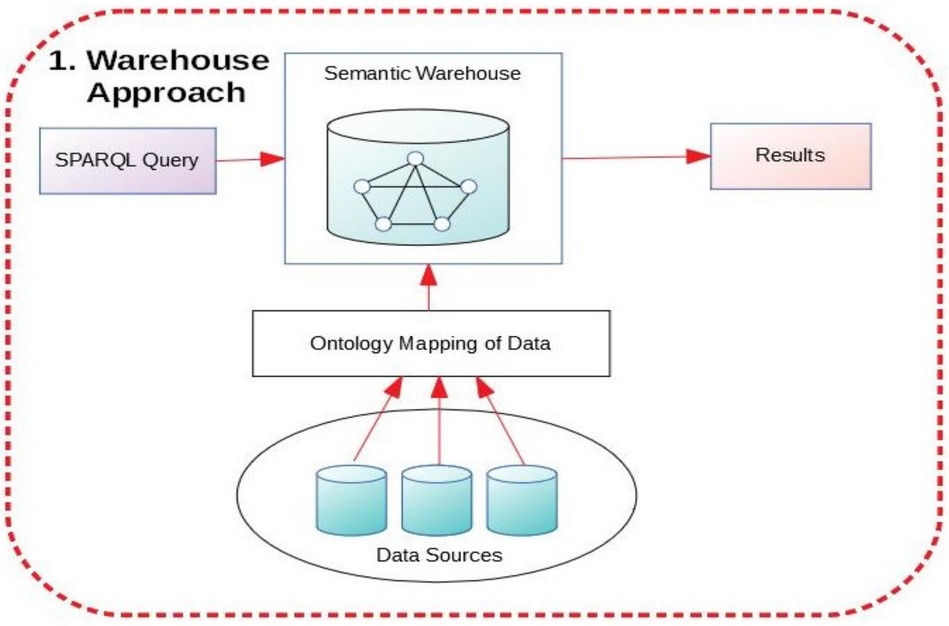
Presenter contact details: **a.kennedy28@nuigalway.ie**

This project looks at utilising semantic data to store and query fisheries data. The approach outlined is a Warehouse Approach, where an Extract, Transform and Load (ETL) is performed on the data. The data is extracted from the MI databases, transformed into RDF triples and loaded into a triple store. A simple ontology is also created and loaded. The triple store is then queried, directly, using SPARQL, and results are obtained.

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.

Fisheries data typically becomes available in batches rather than being streamed. The batches can range in size from an individual fishing trip, up to all vessels in a calendar year. These databases are typically siloed, so querying for information that are stored across many databases can be a significant challenge.



FishID	Weight	Age	Name	Sex
1	303	2	Cod	M
2	298	2	Cod	F
3	333	3	Mackrel	F

Data From Relational Database

Subject	Predicate	Object
1	Weight	303
1	Age	2
1	Name	Cod
1	Sex	M
2	Weight	298
2	Age	2
2	Name	Cod
2	Sex	F
3	Weight	333
3	Age	3
3	Name	Mackrel
3	Sex	F

Data Converted to Triples

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#" xml:base="localhost://">
   <rdf:Description
rdf:about="http://example.com/fish#1">
   <ns0:Name
xmlns:ns0="http://example.com/fish#"
rdf:resource="http://example.com/fish#Cod"/
>
  </rdf:Description>
...
...
...
...
</rdf:RDF>
```

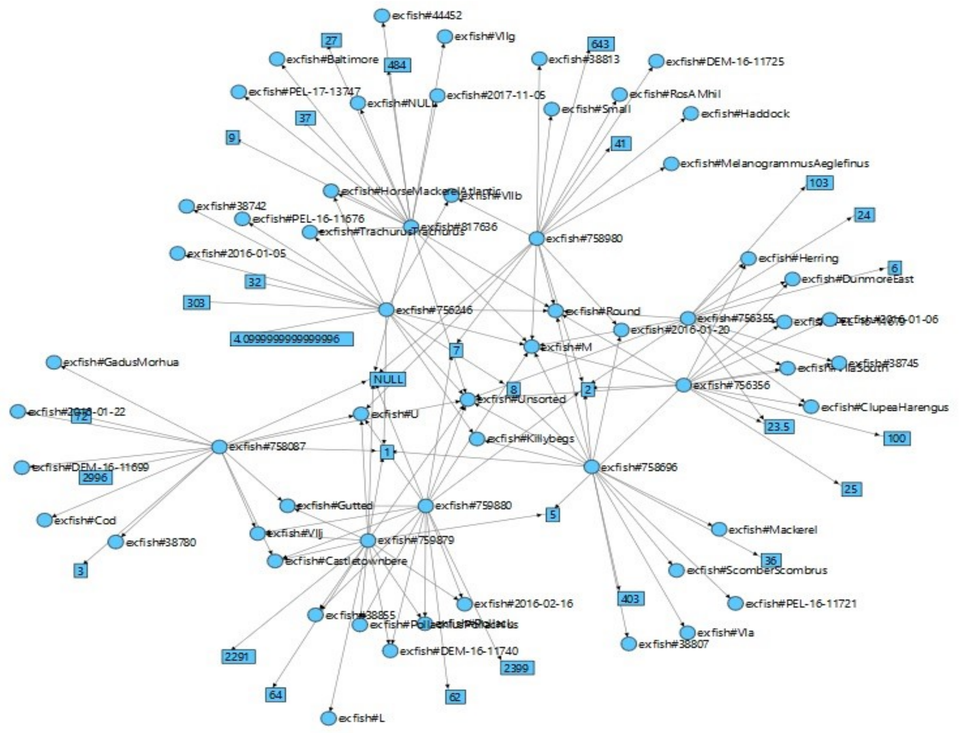
Triples Serialised to RDF File

Conversion of data to RDF triples

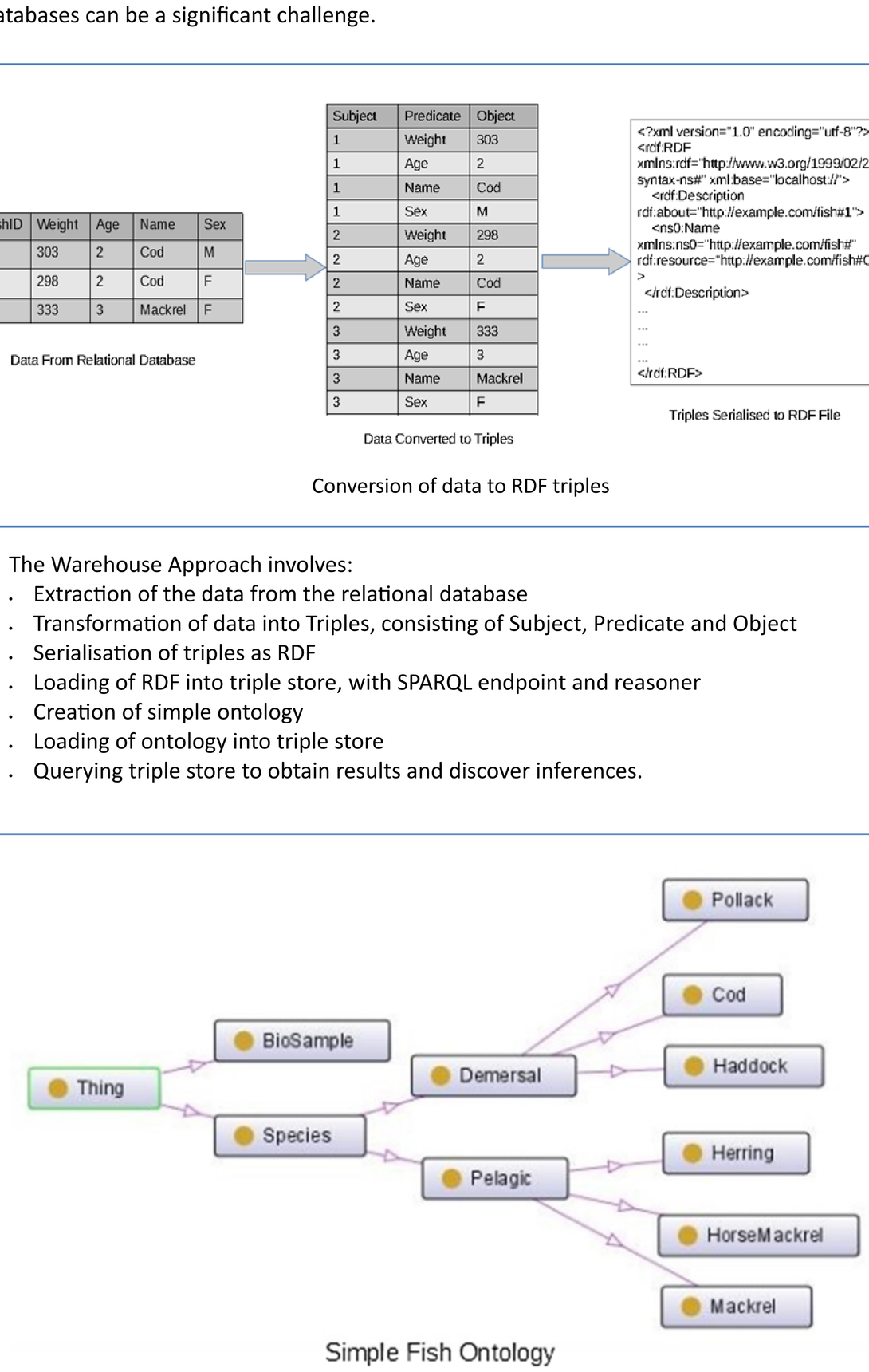
The Warehouse Approach involves:

- Extraction of the data from the relational database
- Transformation of data into Triples, consisting of Subject, Predicate and Object
- Serialisation of triples as RDF
- Loading of RDF into triple store, with SPARQL endpoint and reasoner
- Creation of simple ontology
- Loading of ontology into triple store
- Querying triple store to obtain results and discover inferences.

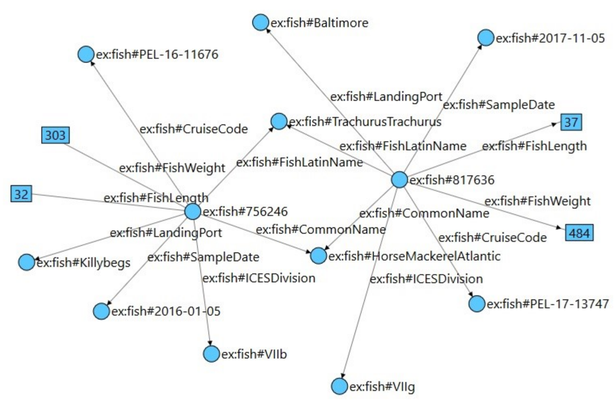
The data in triple format is stored in a Triplestore, which is optimised for storage and retrieval of triples.
The sets of triples can be visualised as a graph, where the nodes are formed by the Subjects and Objects and the edges are the Predicates.



Graph Representation of RDF Triples



Simple Fish Ontology



Graph of Two Samples with Predicates

A simple ontology was created, in OWL, with a small class hierarchy. When applied to the data, the ontology can infer if a fish sample species is a Demersal or Pelagic Fish.
SPARQL is used to query the data. An example query is to return all the samples that are Demersal Fish.

The work carried out so far has concentrated on gaining experience and understanding in semantic modelling of fisheries data. The rest of the PhD will be building on that in the following ways

- Complete Warehouse Approach
- Use a complete data set
- Define a full ontology
- Implement a Mediator Approach, where the SPARQL is converted to SQL and relational database is queried directly
- Compare and contrast Warehouse and Mediator Approaches
- Engage with Fisheries Scientists to build a system that can answer real questions more effectively than existing approaches.

References

Antoniou, G., & Van Harmelen, F. (2004). A semantic web primer. MIT press.

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

Gruber T. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5: 199-220.

Klyne, Graham, and Jeremy J Carroll. (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax. 10 Feb. 2004, www.w3.org/TR/rdf-concepts/

Musen, M.A. (2015) The Protégé project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. DOI: 10.1145/2557001.25757003. (http://protege.stanford.edu)

Tzitzikas Y. et al. (2013) Integrating Heterogeneous and Distributed Information about Marine Species through a Top Level Ontology. In: Garoufallou E., Greenberg J. (eds) Metadata and Semantics Research. MTSR 2013. Communications in

Acknowledgments

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