The ubiquitous need for integrating data for e-science and the data transformation framework of FORTH-ICS-ISL

Yannis Marketakis and Yannis Tzitzikas
{marketak|tzitzik}@ics.forth.gr
Institute of Computer Science, Foundation for Research and Technology (FORTH-ICS), GREECE, and
Computer Science Department, University of Crete, GREECE

Introduction

The role of data in research and e-science:

- Driving insights and discovery
- Enabling collaboration across disciplines
- Supporting complex computational methods
- Facilitating reproducibility and transparency
- Accelerating research outputs and innovation

Major Challenge:

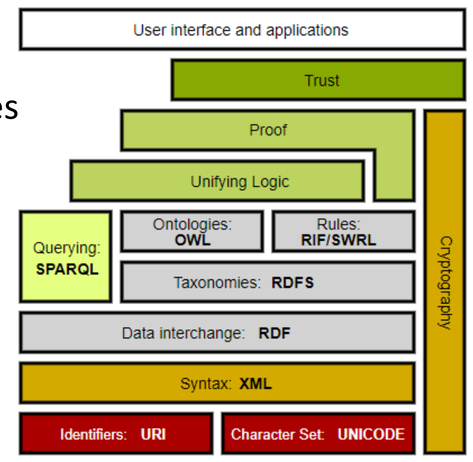
- Integration of heterogeneous data, as varied formats and structures hinder seamless data analysis and sharing across disciplines

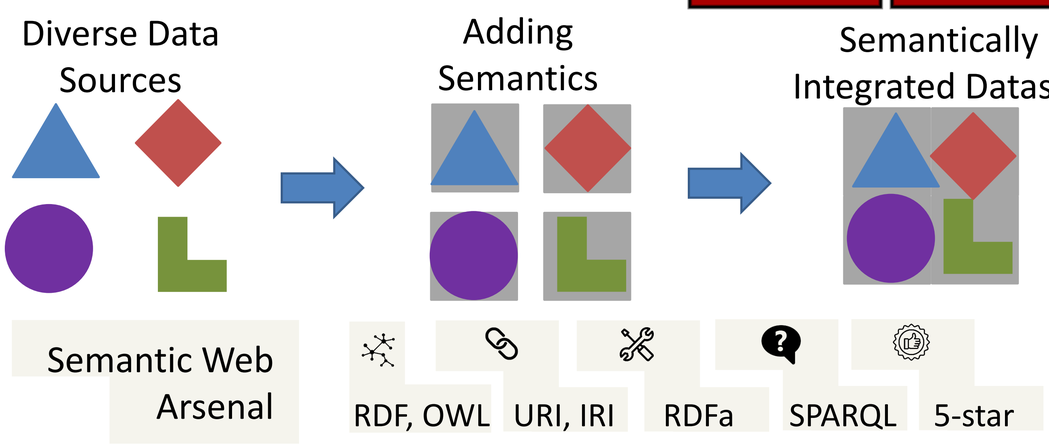
Challenges

- Diverse data types: data exist in numerous forms such as structured (e.g., relational databases), semi-structured (e.g., JSON, XML), and unstructured (e.g., text, images, videos).
- Different datasets often use distinct schemas or data models. Even when representing similar information, the structure and terminology may differ
- Semantic ambiguity is quite often in heterogeneous datasets

The Role of Semantic Web

The Semantic Web provides a structured and interoperable framework that enhances machine readability and understanding of data across different sources. Key features: (a) Standardized Data Representation, (b) Ontology-based Data Integration, (c) Interoperability, (d) Data Linking and Federated Queries, (e) Data Enrichment

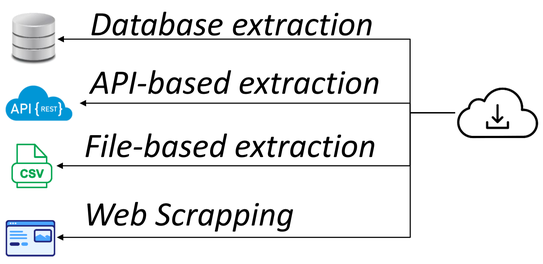




The Data Transformation Framework

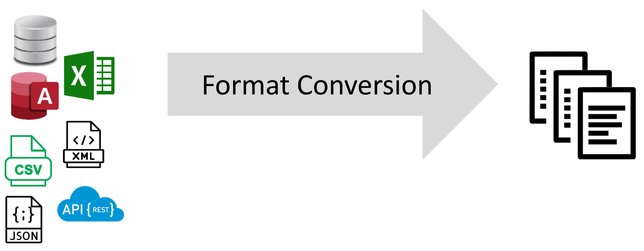
1.Data Collection

Collects entire datasets or some parts of them from their original sources



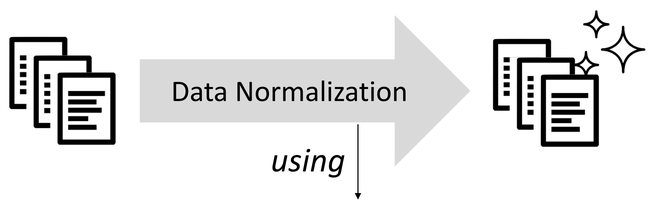
2.Data Format Conversion

Convert collected data into a common format to simplify their further processing



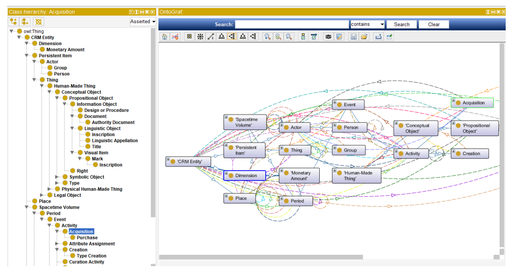
3.Data Curation / Normalization

Normalize data to remove ambiguity, standardize values, and facilitate their exploitation and transformation



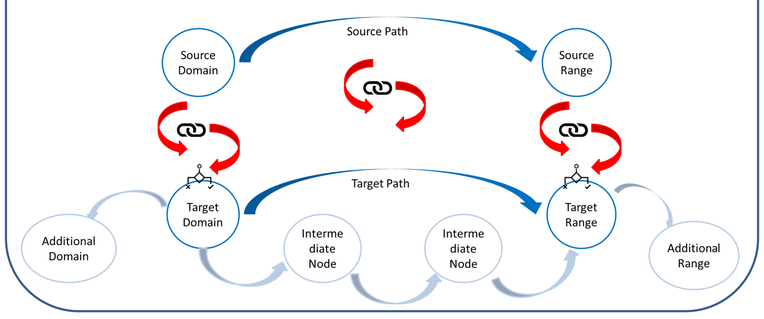
4.Ontology-based Modelling

Definition of a unified model of the domain for integrating data. By reusing or extending existing ontologies



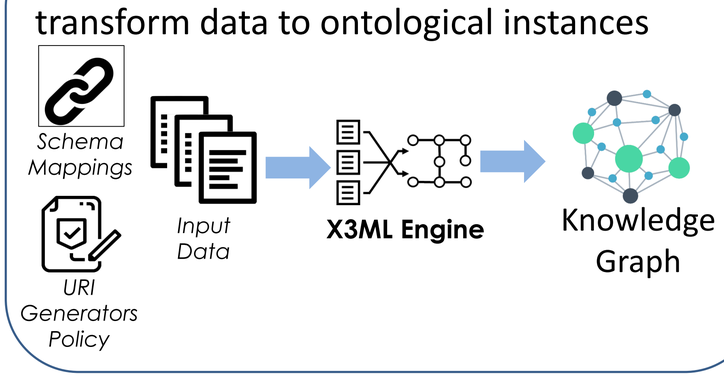
5.Definition of Schema Mappings

How elements from each source are mapped using the adopted ontology(-ies) using **X3ML**



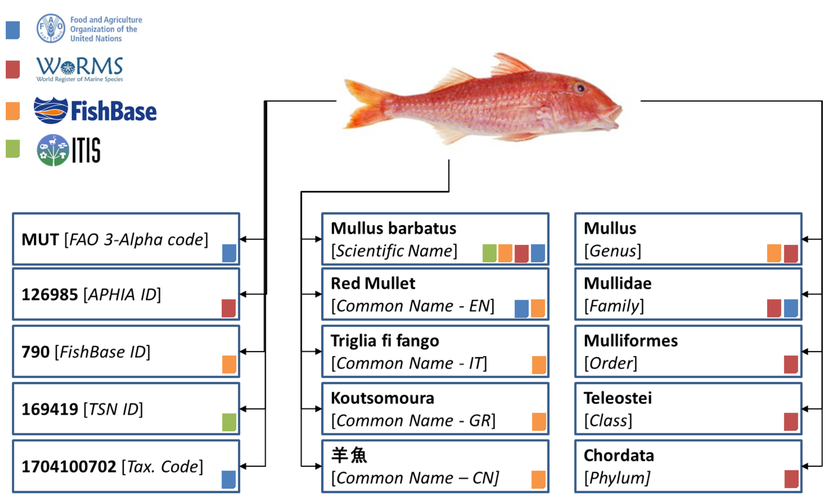
6.Data Transformation

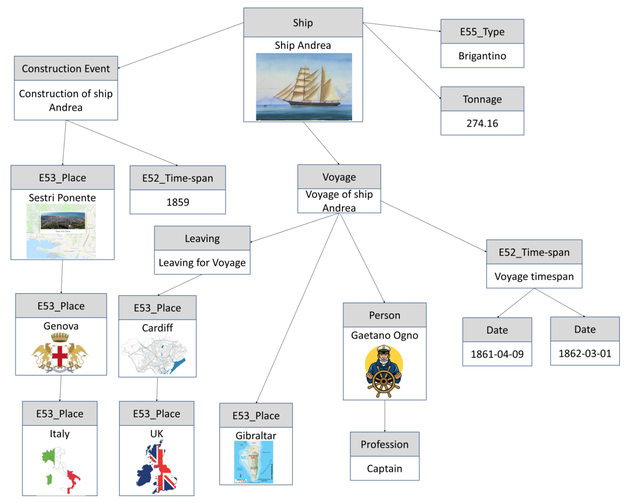
X3ML Engine utilizes schema mappings to transform data to ontological instances



Applications

Applicable to different domains





Successfully applied to various projects:

- BlueBRIDGE
- BlueCloud
- iMarine
- ARIADNE
- ARIADNEplus
- RICONTRANS
- Sealit
- PortADa
- VeriFish
- VRE4EIC

https://github.com/isl/x3ml

14th FORTH RETREAT 2024