Best Practices of Sharing and Publishing (Maritime) Heritage Linked Data: the case of the Eastern Mediterranean

Report produced by Dr Crystal el Safadi and Dr Mohammed Ragab as part of *the 'Establishing the guiding principles for maritime heritage data and databases in the eastern Mediterranean: initial review of feasibility and potential'* project supported by the Honor Frost Foundation.

University of Southampton

Contents

Summary
This report provides a brief summary of a project funded by the Honor Frost Foundation, which aimed at establishing best practices for sharing and publishing maritime heritage data in the eastern Mediterranean (primarily for Lebanon, Cyprus, Syria and Egypt). This first-phase project involved reviewing existing data systems in the region and proposing a preliminary best practice approach. However, further research and discussions are needed to understand the needs, concerns, opportunities, and aspirations of the various stakeholders in the maritime heritage sector, as well as within the broader marine environment. This is an ongoing project, and this report presents the first version.

Background on Maritime Heritage in the eastern Mediterranean
The eastern Mediterranean boasts a rich tapestry of tangible and intangible heritage, reflecting a deep history of engagement with maritime lifeways and early maritime technological advancements (Broodbank, 2013). This region has also experienced significant alterations to its maritime environment due to anthropogenic impacts on urban and coastal areas, as well as natural and climate-induced changes (Safadi et al. 2022). Consequently, the area is abundant in historical and environmental indicators of transformation but faces substantial risks of deterioration and destruction.

Extensive research and surveys have been conducted to uncover, map, document, assess, manage, protect, and conserve this valuable heritage. These initiatives have generated a wealth of digital information at local, national, and regional levels. This data is not only crucial for cultural heritage preservation but also serves as proxy data for studying changing environments, particularly the dynamic coastal and marine ecosystems, evolving landscapes, and climate variations. Moreover, it provides insights into human values and livelihoods.

Ensuring the accessibility, reusability, and discoverability of this data, as well as its sustainability beyond the lifecycle of individual projects, is paramount. Clean data aligned with a common, standardised vocabulary is easier to integrate into Artificial Intelligence (AI) tools. Such tools can significantly enhance heritage management, improving the efficiency and effectiveness of preserving, analysing, and making heritage information accessible for future generations. Existing heritage project databases and national heritage inventories in the region use a variety of data management systems, ranging from bespoke databases to spreadsheet formats. This diversity in managing maritime heritage data leads to heterogeneity, stemming from differences in research needs, technical expertise, and available funding. While these approaches may suffice at the project-specific level for documenting and storing data, they fall short in ensuring consistency of information across multiple initiatives. This inconsistency poses significant challenges from both managerial and research perspectives for three primary reasons:

- Data Consolidation: The integration of data from disparate datasets becomes difficult, hindering research and the effective management of maritime heritage resources.
- Data Compatibility and Accessibility: The absence of standardised systems or guidelines to ensure data compatibility and accessibility impacts future data availability, limiting research potential

and interdisciplinary work, particularly in integrating natural and cultural heritage practices (Breen et al., 2021).

- AI applications: Inconsistent and fragmented data hinder AI tools, which need large, well-organised datasets to work effectively. Without standardised and accessible integrated data, AI's potential to enhance heritage management and research is significantly diminished.

The number of maritime research projects in the region employing independent heritage inventories to store and manage data continues to grow. Notable examples include the Honor Frost Foundation Lebanon Database, El Max project- Egypt, the Maritime Endangered Archaeology (MarEA) and Endangered Archaeology in the Middle East and North Africa (EAMENA) projects (Breen et al., 2022), the Cyprus Coastal Archaeological Project (CCAP), Wrecks at Risk, the Tyre Landscape project and the Beirut Port Project (BPP). With the prospect of future research projects and initiatives generating even more data inventories, it is crucial to ensure thorough planning and foresight in maintaining the sustainability of this data, as well as compatibility between data formats and platforms across different projects. In addition to the independent data management schemes employed by various research projects, there is also directed support towards enhancing national heritage inventories in the region. For instance, the EAMENA and MarEA projects have supported the improvement and establishment of heritage inventory systems in Lebanon. This support underscores the importance of creating a cohesive and sustainable data management infrastructure that can accommodate the growing volume of maritime heritage data.

In recent years, Semantic Web technologies and Linked Data (LD) have emerged as powerful methods for publishing heterogeneous data resources, enhancing their accessibility, interconnectivity, and relevance for both humans and machines (Bizer et al., 2011). These approaches have seen significant advancements in standards, available expertise and tools, and data linking and publication models. Notable examples include large-scale projects like ARIADNE, which has made substantial contributions to data integration and whose e-infrastructure has been repurposed for one of the UK's largest maritime heritage projects, Unpath'd Waters. These initiatives demonstrate that LD offers a sustainable, discoverable, and interoperable means of data management. However, the application of such data structures in the Levant and Egypt remains limited. This is primarily due to the perception that implementing these technologies is technically complex and demanding, coupled with a lack of examples and real-world demonstrations showcasing the benefits of adopting such practices. Consequently, there is a need for more tangible evidence and successful case studies to illustrate the value and feasibility of Semantic Web and LD approaches in these regions.

Heritage and maritime research encompass a diverse and interdisciplinary domain, incorporating a range of disciplines and sciences such as geology, environmental sciences, history, geography, and spatial sciences. Numerous research projects are dedicated to the collection and analysis of maritime heritage data, each pursuing distinct objectives and employing various methods for documentation and data archiving. Additionally, researchers utilise a variety of vocabularies, and in some cases, different languages, to describe and document information related to various maritime heritage artefacts. Given these factors, it is not surprising that the data in this field is often heterogeneous.

Alongside the ongoing initiative of FAIR data management, the prospect of facilitating the reuse of maritime heritage data represents an essential and intriguing path for the research communities to explore. Nonetheless, a set of guidelines that showcase how to reuse existing maritime heritage data resources is still missing. Therefore, finding best practices for publishing and if possible integrating variety of heterogenous data resources about maritime heritage holds significant promise in fostering success within the research community.

In this report, we showcase how to use the principles of Linked Data for linking and publishing various data sources from maritime data management projects/systems (Tommasinin et al., 2020). In particular, we walk through the life cycle of publishing and maintaining maritime Linked Data using the standard technological stack of the Semantic Web (Tommasinin et al., 2020). To demonstrate the feasibility of the proposed approach, we select samples of maritime projects data and show how to represent and publish their data using Linked Data principles and existing semantic web tools. Finally, we discuss the challenges of applying Linked Data principles when publishing data from the maritime heritage data field.

Objectives

O.1: Allow various maritime heritage projects to share their public data with best practices that others can use.

O.2: Help enabling better discovery, accessibility, and reusability of the maritime Heritage data.

O.3: Allow various maritime heritage projects to interoperate with each other.

Background on Linked Data

Linked Data principles are a set of best practices for publishing and connecting structured data on the web. This approach promotes the creation of a global data graph where information from different sources can be linked together.

- Uniform Resource Identifiers (URIs): Each resource should have a unique and resolvable URI. This enables data consumers to access detailed information about the resource.
- HTTP and Linked Data URIs: Resources should be accessible via HTTP, and their URIs should be dereferenceable to retrieve machine-readable data (typically RDF).
- RDF and SPARQL: RDF is used to structure data, while SPARQL is a query language for querying RDF data. Together, they allow for data integration and retrieval across diverse datasets.
- Linking Data: Data publishers should link their resources to related resources using RDF triples. This interconnection forms a global network of data, enabling traversal from one resource to another.

The Linked Data initiative follows principles for representing, storing, structuring, and publishing data on the web. By adhering to these principles, you create a web of interconnected and discoverable data that fosters interoperability and allows users to explore and gather knowledge across various domains and sources. This approach is fundamental to the vision of the Semantic Web and Linked Data. The principles can be summarised below as follows:

- (1) Use URIs as Names for Things:

  URIs (Uniform Resource Identifiers) should serve as unique identifiers for every resource or entity you want to describe in the dataset. These URIs act as names for the things you are representing. They should be meaningful and specific, making it clear what the URI identifies. For example:

  URI for a ship: http://example.org/ships/queen-mary

  URI for a person: http://example.org/people/john-doe

  URI for a maritime event: http://example.org/events/sinking-of-titanic

- (2) Use HTTP URIs for Lookup:

  HTTP URIs enable people and machines to access information about the resources they identify. When someone looks up a URI in a web browser or programmatically, they should receive a response containing information about that resource.

  For instance: Accessing http://example.org/ships/queen-mary should return data about the ship Queen Mary in a machine-readable format (typically RDF).

- (3) Provide Useful Information Using RDF and SPARQL:

  When someone accesses a URI, the response should include RDF data that describes the resource. RDF (Resource Description Framework) is a format for representing structured data in triples (subject-predicate-object). SPARQL is a query language for querying RDF data. You should provide RDF data that is rich in context and relationships, making it useful for both humans and machines to understand.

  Example RDF data for the Queen Mary ship:

```
@prefix ex: <http://example.org/ships/> .

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

ex:queen-mary

   rdf:type ex:Ship ;

   ex:name "Queen Mary" ;

   ex:builder "John Brown & Company" .
```

The above RDF graph can be visualised as follows:

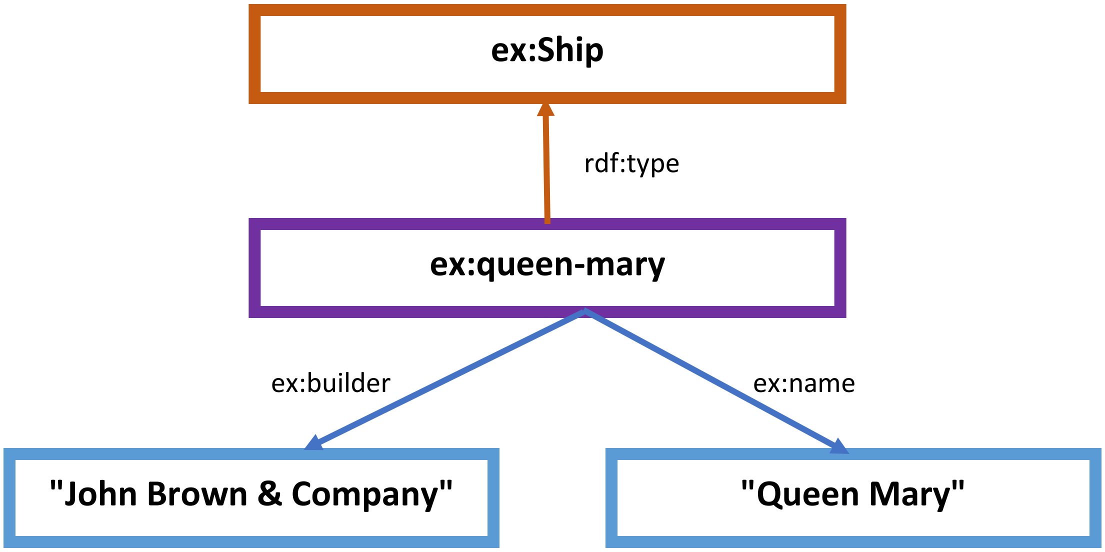


Figure 1- Visualisation of the RDF Graph

- Include Links to Other Datasets (URIs):

  The dataset should include links (URIs) to related resources in other datasets. These links enable data consumers to discover more information and navigate between different datasets seamlessly. For example, you might link to datasets containing information about shipbuilders, historical events, or geographical locations.

```
ex:queen-mary ex:builder <http://example.org/people/john-brown> ;
                         ex:event <http://example.org/events/sinking-of-titanic> .
```

Methodology

We approach the objectives of this project adopting the Linked Data Paradigm for publishing and linking/integrating diverse and dispersed datasets. Publishing heterogeneous data resources using Semantic and Linked Data principles enhances data discoverability, accessibility, and interoperability, making it easier for users to combine and analyze diverse datasets for valuable insights.

The Linked Data paradigm is built on two simple ideas:

- (1) RDF provides an expressive data model for representing structured information.
- (2) RDF links are set between entities in different data sources.

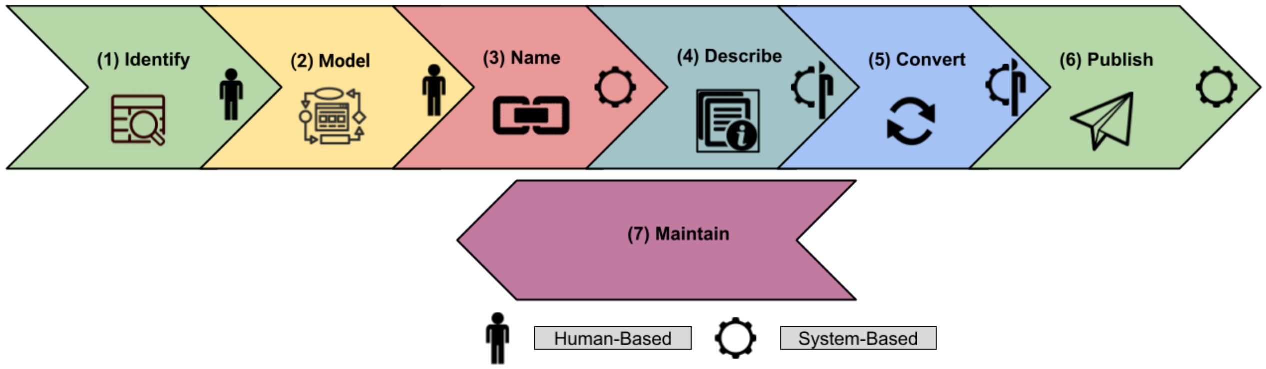


Figure 2- Linked Data life-cycle

Publishing maritime Heritage Linked Data involves specific considerations due to the unique nature of the domain. The Semantic Web can be seen as a new layer of metadata being built inside the Web. Creating and maintaining a Linked Data project involves several steps, from identifying and selecting datasets to publishing RDF data. Figure above shows a high-level overview of those steps.

Herein, we present the publication lifecycle of maritime Linked Data. Figure 2 shows the sequence of those steps that can be summarized as follows:

1. **Identify and Select Dataset(s)**
   - Determine and understand the scope and objectives of the maritime heritage project.
   - Identify and select the datasets that are involved for the data sharing and publishing practice, even at a later date.
   - Ensure that the selected datasets are available in digital format and can be accessed for conversion into Linked Data.
2. **Model Linked Data**
   - Define the data entities and their relationships.
   - Design an RDF ontology that represents the data model.
3. **Name Things Using Persistent HTTP URIs**
   - Assign unique, persistent HTTP URIs (Uniform Resource Identifiers) to each data entity and relation in the maritime heritage dataset.
   - Ensure that these URIs resolve meaningful information about the identified maritime heritage resources when accessed through a web browser.
4. **Describe Things with Previously Used Vocabularies**
   - Use existing standard ontologies relevant to maritime heritage to describe the dataset entities and relationships. This ensures compatibility and ease of data integration.
   - Extend existing vocabularies or create new ones if necessary.
5. **Convert the Data into Linked Data (RDF)**
   - Transform the data from its original format (e.g., CSV, JSON, XML) into RDF triples, because as mentioned each maritime data management project could maintain their data in different formats.
   - Use tools and libraries like Apache Jena, RDFLib, or custom scripts to perform the conversion.

6. **Publish RDF Data**
   - Make the RDF data accessible through various means, such as:
     - Hosting RDF files on a web server.
     - Providing a SPARQL endpoint for querying the data.
     - Using linked data platforms like Virtuoso or Fuseki.
     - Utilizing a triple store for efficient storage and retrieval of RDF data.
7. **Maintain Data Availability and In-Service:**
   - Regularly update and synchronise the generated maritime Linked Data with the source datasets to ensure data freshness.
   - Implement versioning and backup strategies to protect against data loss.
   - Monitor the availability of generated maritime Linked Data service and ensure it remains accessible.

Putting the Lifecycle into Practice (Use Cases):

Here, we present representative examples of real-world maritime heritage datasets from various projects published as Linked Data. The goal is to provide guidelines that motivate the samples of RDF maritime data models.

(1) MarEA database

The MarEA (Maritime Endangered Archaeology) project aims to address the significant threats facing the coastal and underwater cultural heritage in the Middle East and North Africa (MENA) region (Breen et al., 2021). The project recognises the critical importance of preserving and documenting the coastal and underwater cultural heritage in the MENA region. By adopting established methodologies, the project creates a comprehensive database (using Arches DB) that addresses the dynamic nature of these sites and the various threats they face, ultimately contributing to their protection and conservation.

The MarEA project feeds into the EAMENA project database and this database is based on the Arches software and incorporates a Linked Data structure. However, for demonstration purposes, a subset of MarEA data from a .csv file was used to demonstrate how this data, outside of the Arches database, could be transformed. Publishing Maritime Linked Data from the MarEA project requires implementing the standard Linked Data publication life cycle mentioned above (Antoniou et al., 2004) and depicted in Figure 2.

1. **Identify and Select Dataset(s)**

   For this project, we work on a sample of CSV data, exported from the Arches DB. This sample mainly describes the attributes of assessment activities of various maritime heritage resources. This data resource describes details of the resource alongside listing all kinds of conducted resource assessments, e.g., conditional, environmental, and archeological.

On this sample, we decide on the required attributes to keep or discard. As shown in Figure 2, this step is human-based and requires not only clear understanding of the dataset, but also the goal of the required mapping to the target RDF representation.

2. Model Linked Data

This step highly depends on the previous step, as it entails modeling the concepts and entities within the database and deciding the relationships among those concepts. For example, for the database in our hands, we can elicit this global modeling based on the understanding of the dataset internal structure. We have an "Assessment activity" has been done by an "Assessor/Archeologist" for a "Resource", on several levels of assessment "Environmental", "Archeological", and "Conditional". Each of these generic classes/entities has multiple proprieties or attributes and could have one or more sub-classes. For example, the Archeological assessment can entail "Chronological", "Material" sub-classes. Figure 3 shows a generic modelling of the main classes/entities in dataset and their relationships (visualized using the webvowl ontology visualization tool).

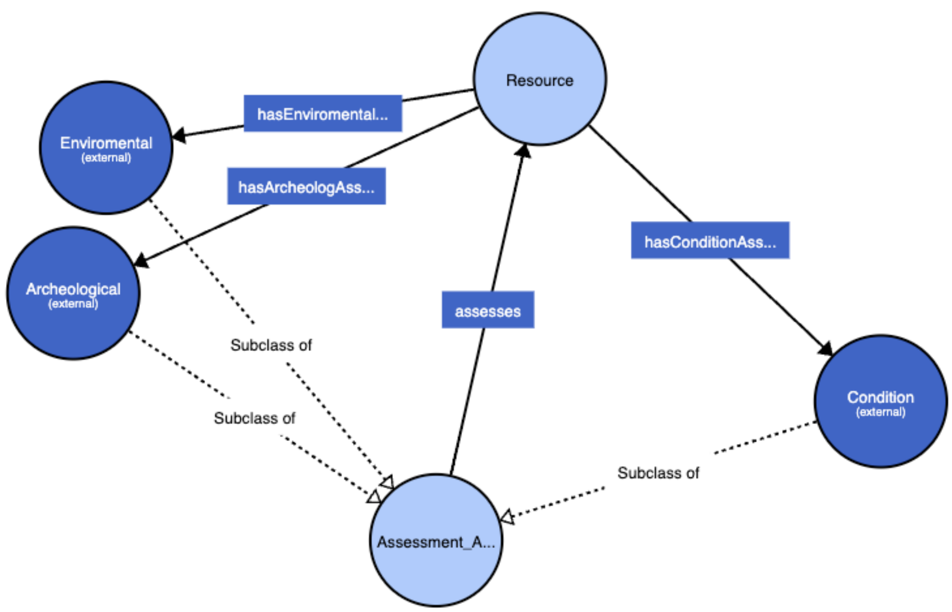


Figure 3- Generic modelling of the classes/entities using webvowl.

Below, we can see part of the full data model for structuring the relationships between concepts and entities of the project's data.

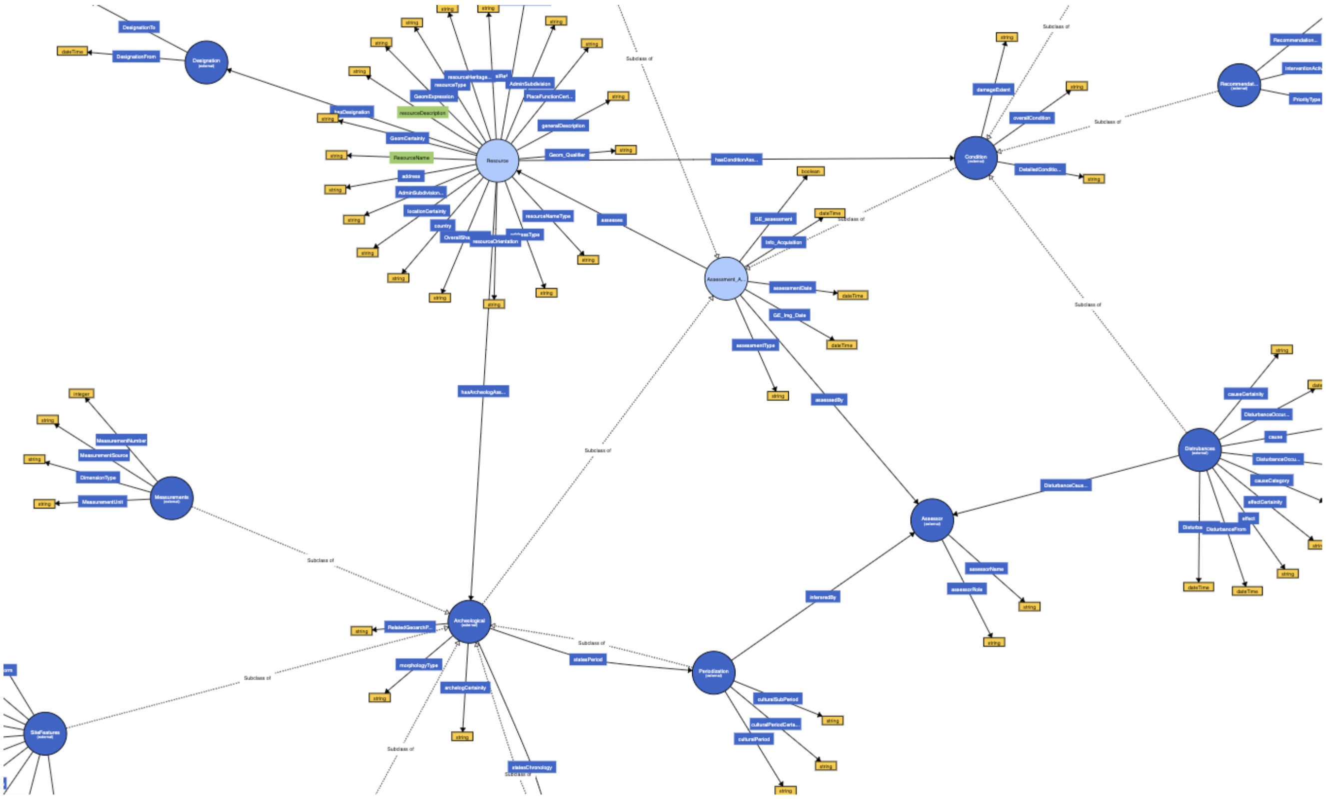


Figure 4- Full structure of the data visualised.

3. Name Things Using Persistent HTTP URIs

This step entails selecting and using persistent HTTP URIs for both the entities and relationships for the target RDF model. In this case, we use the following base URI http://martime.soton.ac.uk/ontology/marea (please note this domain is only an example and is not active) for identifying and naming resources and concepts in the ontology. For example, the "Assement_Activity" concept/resource can be identified (accessed) using the URI as follows: http://martime.soton.ac.uk/ontology/marea/# Assement_Activity.

Also, this step entails using existing vocabularies (if existing) to select uniform and common naming of things (I.e., classes or relationships). In the field of maritime databases, we can rely on the existing vocabulary of CIDOC-CRM ontology (https://www.cidoc-crm.org/) (Bruseker et al., 2017). This ontology is widely used in the applications of information integration in the field of cultural heritage. It achieves this by providing definitions and a formal structure (I.e., Ontology) for describing the implicit and explicit concepts and relationships used in cultural heritage documentation and of general interest for the querying and exploration of such data.

Figure 5 below shows a concise depiction of the basic concepts of the CIDOC-CRM based ontology (Karagiannis et al.)The "Exx" codenames denote classes (entities), as defined in the CRM, and likewise "Pxx" denote object properties. The arrows represent subsumption relations and faded arrows represent object properties; one may notice some of the original classes of the CRM and some new classes that we have added (the new classes are denoted as "Exx.y", where "xx" denotes the codename of the CIDOC-CRM superclass).

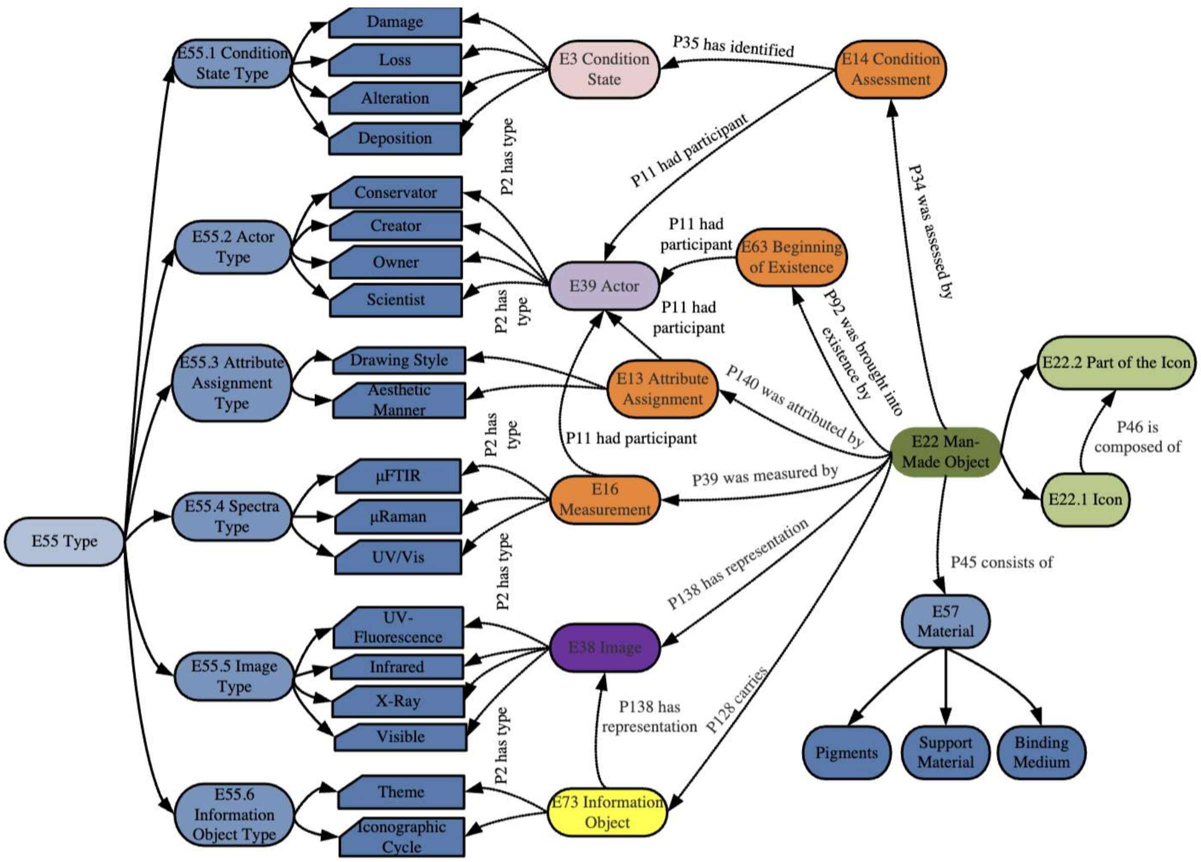


Figure 5- concise depiction of the basic concepts of the CIDOC-CRM ontology

Hence, we endeavor to maximise our utilisation of the formal concepts employed within well-established and widely-adopted ontologies, such as CIDOC-CRM. This is done in pursuit of both formal precision and seamless future data integration with other extant datasets.

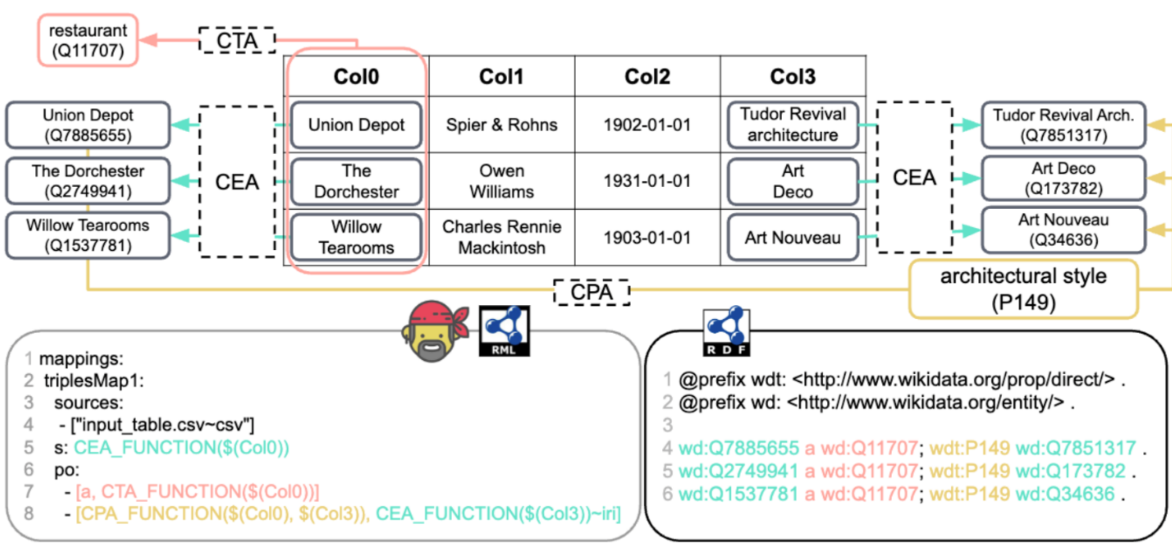
As an illustrative example, within our modeling of the MAREA DB, we have a class known as "Condition_Assessment." To maintain consistency and alignment with the established model naming (E14: Condition Assessment), we employ the same nomenclature to denote the identical concept (depicted in the accompanying figure). In instances where concept definitions are not readily available in existing vocabularies or ontologies, we undertake the task of defining them ourselves, introducing new vocabulary terms as needed.

4. **Convert the Data into Linked Data (RDF)**

This step entails converting the data from its original format (e.g., CSV, JSON, XML, ..etc.) into RDF triples. The Semantic Web community provides several mappings that describe how existing data can be represented using the RDF data model.

The most popular language for mapping heterogenous data into RDF is the RML (https://rml.io/specs/rml/). RML (Dimou et al., 2015) is a language for expressing customised mappings from heterogeneous data structures and serializations to the RDF data model (currently defined for sources in a structured format, e.g. databases and CSV, TSV, XML, and JSON data sources) to RDF datasets. RML is based on and extends [R2RML (https://rml.io/specs/rml/#bib-r2rml)]. R2RML [6] is defined to express customised mappings only from relational databases to RDF datasets. An RML mapping contains one or more Triple Maps which on their own turn contain a Subject Map to generate the subjects of the RDF triples, and zero or more Predicate Object Maps with pairs of Predicate and Object Maps to generate the predicates and the objects respectively for each incoming data record.

Herein, we see a simple example that shows how to convert CSV input data into RDF data in RML mappings [9]. The example shows how the language specified that all instances inside the first column are the subjects of the triples (s:CEA_FUNCTION($Col0)). Then, the code specifies the type of each of those subjects as of type "restaurant" using the function (a, CTA_FUNCTION($Col0)), and that its architectural structure corresponds to the column 3 values using the function (CPA_FUNCTION($(Col0),($Col3))). The RDF output of the mappings is shown on the right as a form of RDF triples.



For simplicity, the developers of Semantic Web community aimed to make the life of RDF non-practitioners easier by providing tools that hide the complexity of RDF RML languages and mapping rules, and automatically convert heterogenous data resources into RDF datasets. One of the most popular and widely used tools for such tasks is OpenRefine (https://openrefine.org/). OpenRefine [10] is a powerful free, open-source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

To empower OpenRefine with RDF transformation capabilities, it is required to install an extension (https://github.com/AtesComp/rdf-transform) that transforms data into RDF formats. Based on the RDF extension. Figure 6 shown below shows the steps of installing OpenRefine and its RDF extension. It is important to mention that OpenRefine enables reuse of existing ontologies and vocabularies for naming conventions of the dataset instances.

Figure 6- Steps to download OpenRefine.

The steps of using the OpenRefine tool empowered with the RDF extension are as follows (see Figure 7below):

1. Import the CSV input files of the dataset subject to conversion.
2. Create a project inside OpenRefine with a meaningful name.
3. Click “RDF” and then “Edit RDF Skeleton” and verify the RDF Skeleton has been created.

4. By clicking the “Edit RDF Skeleton” button, you can check/modify the current mapping between your data and the RDF skeleton.
5. Upload the existing ontologies or your newly created vocabularies for describing the dataset entities for the target RDF dataset.
6. Finally, click “Export” then “RDF in Turtle” to get an RDF description of your solution.

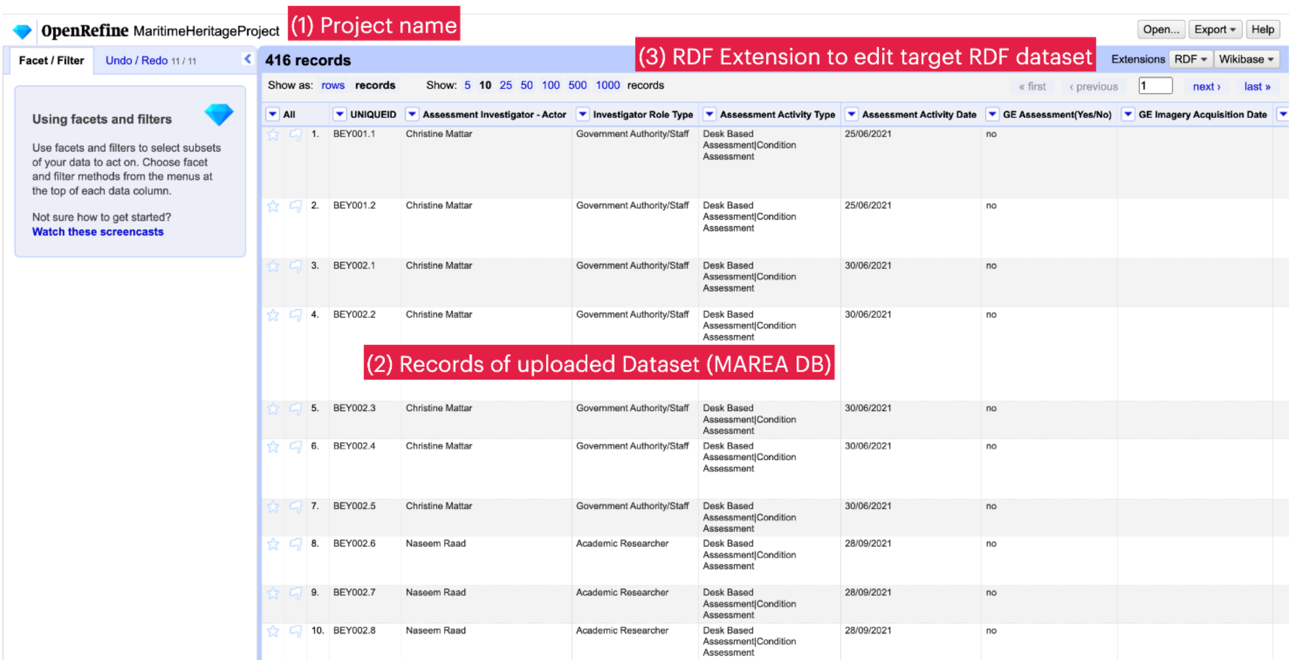


Figure 7- OpenRefine steps

Figure 8 below shows how to upload the CIDOC-CRM ontology to describe and name entities in the MAREA database. Notably, the CIDOC-CRM ontology can be downloaded as a “.ttl” file from the official project website (https://www.cidoc-crm.org/).

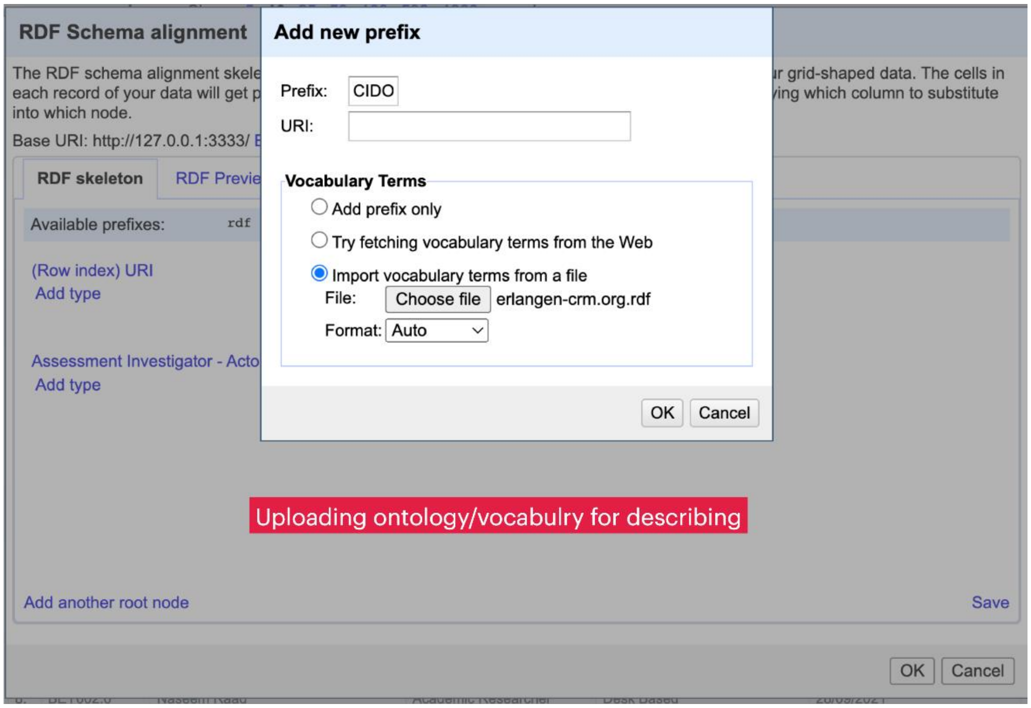
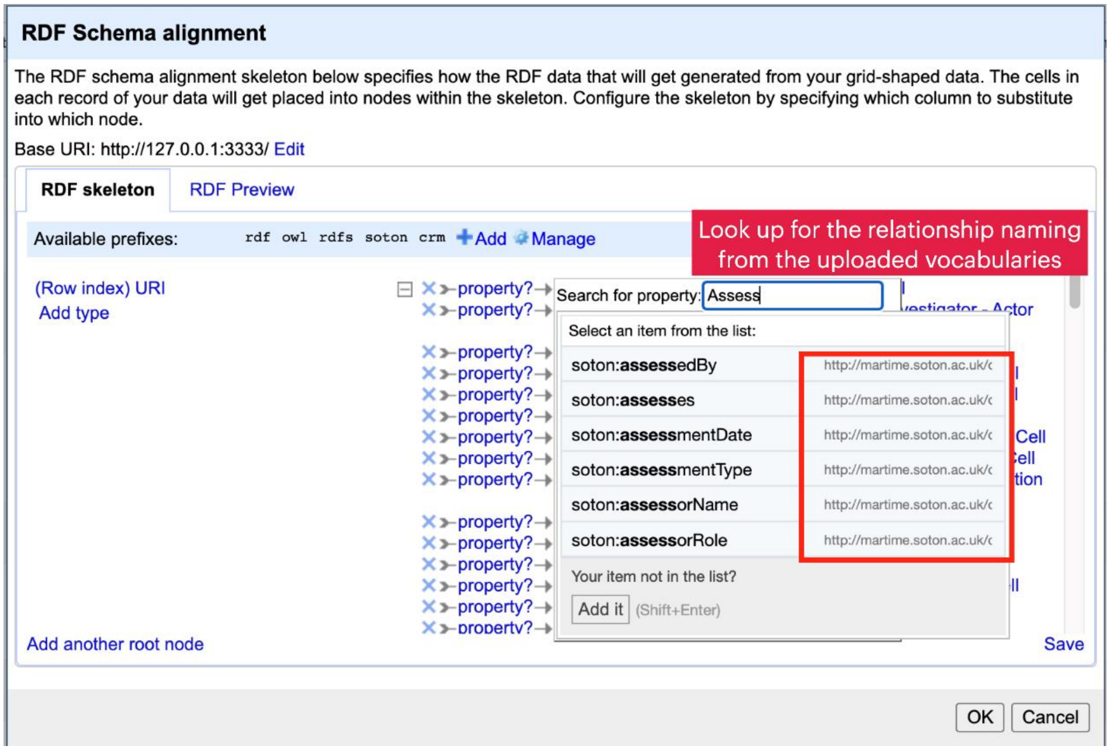
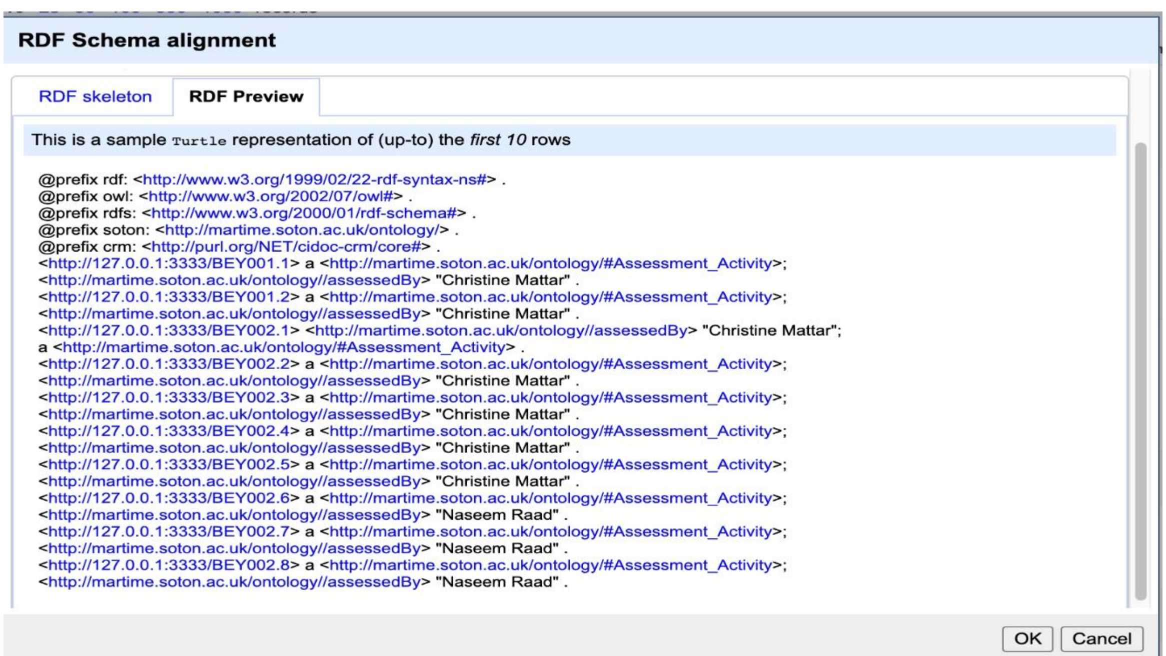


Figure 8- How to upload existing ontologies.

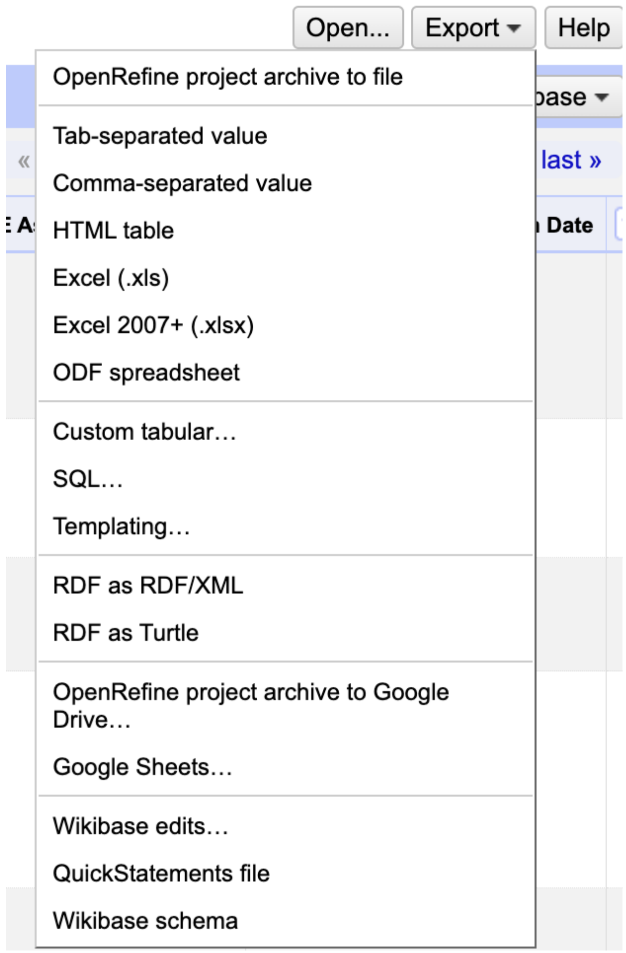
Next step is to pass by the data columns in the input CSV dataset and assign classes and properties (relationships) naming from the uploaded vocabularies/ontologies (e.g., CIDOC-CRM ontology).



Finally, we can preview the output RDF dataset (in the form of triples) from the "RDF Preview" button. Figure below shows an example of the MAREA DB in the form of RDF. It describes each of the rows as of the type "Assessment Activity", and that each unique assessment is "Assessed By" an Archeologist (shown by his/her name).



From the export button as shown below, we can download the target output RDF datset.



(2) Database Example 2

In this report, we provide a second example of mapping a resource of maritime data into RDF utilising the same standard LOD life cycle.

This dataset from an existing project encompasses details regarding diverse shipwreck resources located at multiple coastal sites. It houses information about the characteristics of these shipwrecks, their geographical positions, and the data acquired through archaeological assessments conducted by institutions or exhibitions. These assessments encompass findings within the shipwrecks, their condition assessments, and an examination of the different threats they encounter. Ultimately, this information aids in safeguarding and preserving these historical sites.

Herein, we show how we implement the standard Linked Data publication life cycle mentioned above (Antoniou et al., 2004) and depicted in Figure 2 for publishing Maritime Linked Data from this project's database. The steps are as follows:

1. Identify and Select Dataset(s)

For this project, we worked on a sample of CSV data, recorded by the project team. This sample mainly describes the description details of assessment activities of various maritime heritage shipwrecks resources. This data resource describes details of these resources alongside listing all kinds of conducted assessments, e.g., conditional, chorological, and geographical.

2. Model Linked Data

Like what we did with the MAREA project, we elicit the main concepts/entities in the project's database and get a generic relationship of those concepts. Again, this step highly depends on the previous step, understanding the internal structures of the selected database. For example, in this database, we can elicit the concept "Assessment" conducted by archeologist ("Assessor") in a "Management" organisation ("Institute", or "Exhibition"). Each "Shipwreck" resource in the database may have various types of assessments (e.g., "Chronology", "Geography", and "Condition"). The assessment also entails information about the "Site" of the shipwreck resources. The shipwreck resource is also described by details of its "Cargo-Details", "Equipment", "Tools", "Gallery-Wares", "Personal Belongings", etc. Figure 9 shows a generic modelling of the main classes/entities in dataset and their relationships.

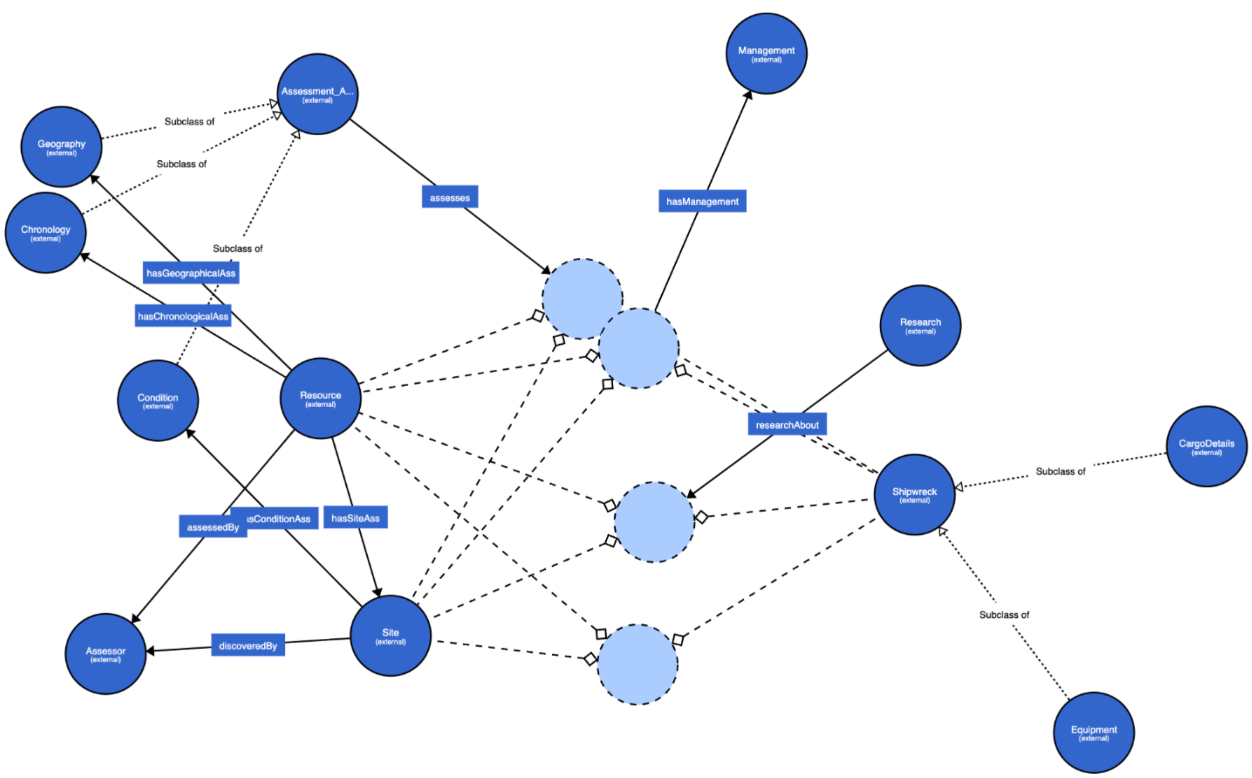
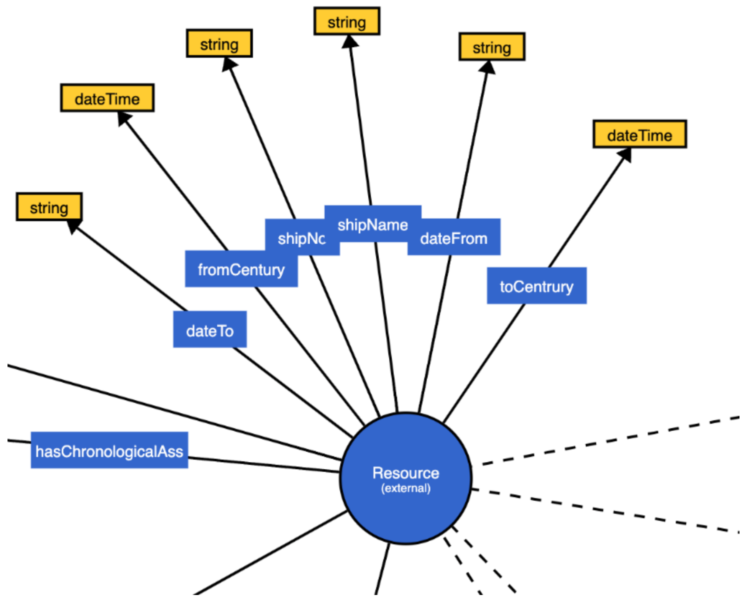
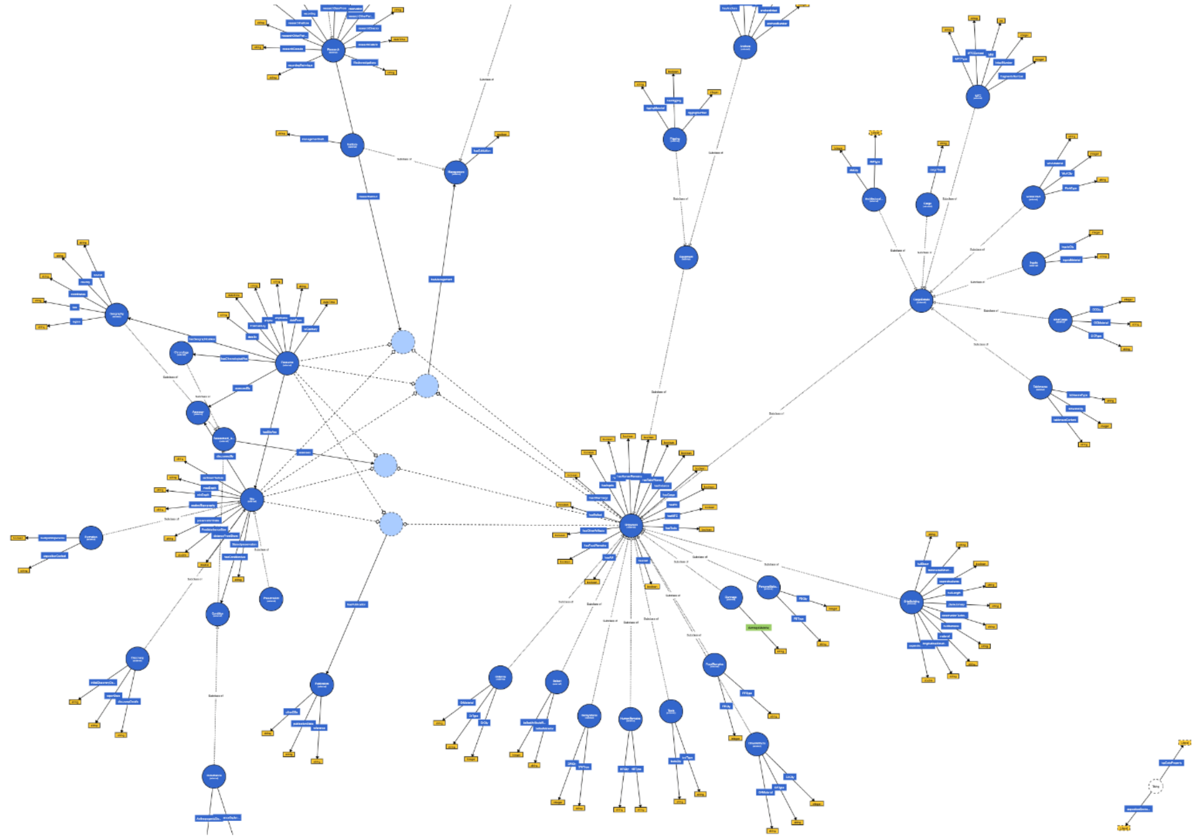


Figure 9- Generic modelling of the classes/entities of the project database

Each of those concepts has several attributes (known as properties). For example, the ”Resource” entity has the following attributes (“shipName“, ”shipNumber”, ”fromcentrury”, ”tocentury”,etc) as shown bellow.



Below, we can see part of the full data model for structuring the relationships between concepts and entities of the this project’s data. It is worth mentioning that Steps 3-5 follow the same structure as for the MAREA_DB use case.



What next and How end-users can use that:

The following step after transforming the input dataset into an RDF asset is “Publishing RDF datasets” (Step 6 in the life cycle). This process typically includes several key steps. First, the RDF data is made accessible through different means, including hosting RDF files on a web server to allow for direct download and access. Additionally, a SPARQL endpoint may be provided, enabling users to query and retrieve specific RDF data using the SPARQL query language. Moreover, linked data platforms like Virtuoso or Apache Jena Fuseki can be employed to enhance data discoverability and interoperability. To ensure efficient storage and retrieval of RDF data, a triple store may be used, facilitating the management and optimization of RDF datasets. These various methods of making RDF data available contribute to the broader goal of sharing and linking structured data on the web for enhanced data integration and knowledge discovery.

To this end, in our project that is related to maritime information and knowledge sharing, end users can make the most of the published RDF data through the following means:

1. Direct Downloads: If Maritime Linked Data provides RDF data files on a web server, end users, including researchers, maritime professionals, and developers, can download these files. They can use RDF-compatible tools to extract and analyse specific data they require, such as shipwreck information, coastal/maritime site conditions, etc.

2. SPARQL Queries: Maritime Linked Data could offer a SPARQL endpoint, enabling users to execute SPARQL queries for retrieving maritime data. This approach allows users to find details about wrecks, harbours or any other maritime-related information by crafting customised queries.
3. Linked Data Platforms: Utilising linked data platforms like Virtuoso or Fuseki within Maritime Linked Data allows end users to explore interconnected maritime datasets in a user-friendly way. They can navigate through the linked data, uncovering relationships and insights that help in maritime research, logistics planning, or maritime sustainability analysis.
4. Triple Store Access: With the use of a triple store, end users can access the Maritime Linked Data through an optimised interface. This enables them to perform efficient queries, explore the data, and discover valuable connections and patterns within the maritime domain.

In the context of Maritime Linked Data, these methods empower end users to access, query, and utilise RDF data to enhance their understanding and decision-making in various maritime-related fields. This structured data serves as a valuable resource for making informed choices and gaining insights into the maritime domain.

Conclusion

This report highlights the transformative potential of Linked Data for the maritime heritage research community. By embracing these principles and developing guidelines for data publication, we can unlock the true value of our maritime heritage data, fostering collaboration and innovation in this interdisciplinary field. As researchers and institutions continue to embrace these practices, we can look forward to a future where maritime heritage data is more interconnected, accessible, and valuable than ever before.

References

Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data: The story so far." Semantic services, interoperability and web applications: emerging concepts. IGI global, 2011. 205-227.

Tommasini, Riccardo, et al. "A first step towards a streaming linked data life-cycle." The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19. Springer International Publishing, 2020.

Antoniou, Grigoris, and Frank Van Harmelen. A semantic web primer. MIT press, 2004.

Best Practices for Publishing Linked Data https://www.w3.org/TR/ld-bp/, last accessed 4 Aug 2023.

Dimou, Anastasia, et al. "RML: A generic language for integrated RDF mappings of heterogeneous data." *Ldow* 1184 (2014).

Rodriguez-Muro, Mariano, and Martin Rezk. "Efficient SPARQL-to-SQL with R2RML mappings." Journal of Web Semantics 33 (2015): 141-169.

Bruseker, George, Nicola Carboni, and Anaïs Guillem. "Cultural heritage data management: The role of formal ontology and CIDOC CRM." *Heritage and archaeology in the digital age: acquisition, curation, and dissemination of spatial cultural heritage data* (2017): 93-131.

Karagiannis, G., et al. "Towards «Cultural» Intelligence: Applying Signal Processing and Semantic Web Technologies in the analysis of Byzantine Iconography."

Chaves-Fraga, David, and Anastasia Dimou. "Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges." *KGCW@ ESWC* (2022).

Verborgh, Ruben, and Max De Wilde. Using openrefine. Packt Publishing Ltd, 2013.

Broodbank, C., 2013. *The making of the Middle Sea : a history of the Mediterranean from the beginning to the emergence of the classical world.* Oxford University Press, Oxford

El Safadi, Crystal, Ray, Nick, Ortiz Vazquez, Rodrigo, Nikolaus, Julia, Blue, Lucy, Westley, Kieran, Andreou, Georgia and Breen, Colin (2022) Maritime cultural heritage and urbanisation in the Middle East and North Africa*. The Historic Environment.* (doi:10.1080/17567505.2022.2075070).

Bruseker, George, Nicola Carboni, and Anaïs Guillem (2017). Cultural heritage data management: The role of formal ontology and CIDOC CRM." *Heritage and archaeology in the digital age: acquisition, curation, and dissemination of spatial cultural heritage data*: 93-131.

Tommasini, Riccardo, et al. (2020). A first step towards a streaming linked data life-cycle. *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November* 2–6, 2020, Proceedings, Part II 19. Springer International Publishing, 2020.

Bizer, Christian, Tom Heath, and Tim Berners-Lee (2011). Linked data: The story so far. *Semantic services, interoperability and web applications: emerging concepts*. IGI global, 2011. 205-227.

Breen, Colin, El Safadi, Crystal, Huigens, Harmen, Tews, Sophie, Westley, Kieran, Andreou, Georgia, Ortiz Vazquez, Rodrigo, Nikolaus, Julia and Blue, Lucy (2021) Integrating cultural and natural heritage approaches to Marine Protected Areas in the MENA region. *Marine Policy*, 132 (104676), [104676]. (doi:10.1016/j.marpol.2021.104676).

Breen, Colin, Blue, Lucy, Andreou, Georgia, El Safadi, Crystal, Huigens, Harmen, Nikolaus, Julia, Ortiz-Vazquez, Rodrigo, Ray, Nick, Smith, Ash, Tews, Sophie and Westley, Kieran (2022) Documenting, protecting and managing endangered maritime cultural heritage in the Middle East and North Africa (MENA) region. *Journal of Maritime Archaeology*, 17 (3), 341-352. (doi:10.1007/s11457-022-09338-z).