An Exploratory Approach to Archaeological Knowledge Production

Costantino Thanos · Carlo Meghini · Valentina Bartalesi · Gianpaolo Coro

Abstract The current scientific context is characterized by intensive digitization of the research outcomes and by the creation of data infrastructures for the systematic publication of datasets and data services. Several relationships can exist among these outcomes. Some of them are explicit, e.g. the relationships of spatial or temporal similarity; whereas others are hidden, e.g. the relationship of causality. By materializing these hidden relationships through a linking mechanism, several patterns can be established. These knowledge patterns may lead to the discovery of information previously unknown. A new approach to knowledge production can emerge by following these patterns. This new approach is exploratory because by following these patterns, a researcher can get new insights about a research problem. In the paper, we report our effort to depict this new exploratory approach using Linked Data and Semantic Web technologies (RDF, OWL). As a use case, we apply our approach to the archaeological domain.

Keywords Knowledge bases · Knowledge discovery · Ontology · Linked Data · Semantic Web · Archaeological knowledge

1 Introduction

The current scientific context is characterized by an intensive digitization of the research outcomes (datasets, tools, and services). The result of this intensive activity is the creation of data islands (isolated datasets/collections) that can include textual and visual documents. Data islands are continuously created and accessed online. However, the researchers need to aggregate these data islands in order to discover new insights. Instrumental in aggregating the data islands is the publication of the research outcomes. Publishing the research outcomes, that is, making them openly accessible is achieved through the development of domain specific catalogues/registries (Research Data Infrastructures). The publishing of the research outcome is, furthermore, supported by a policy platform of the European Union that prescribes the openly sharing of all research outcomes as early as is practical in the discovery process. This policy establishes the new paradigm of science, the open science.

In order to implement an efficient aggregation of research data we propose a three-step procedure:

- The first step towards the goal of aggregating/integrating the so created research data, is the publication of these data on domain specific data registries. The publication of research data through data registries is instrumental in making research data FAIR (findable, accessible, interoperable, reusable)[1].
- The second step is the discovery of existing relationships between the data islands. Some of these relationships are explicit, such as for instance the relationships determined by the sameness or the similarity of topical, spatial or temporal coverage between datasets; while some others are hidden/implicit, such as relationships of causality that hold between, e.g., a report about migration and an excavation dataset. The discovery of data relationships must be supported by specific ontologies.

Costantino Thanos, Carlo Meghini, Valentina Bartalesi, Gianpaolo Coro
Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" – CNR Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
E-mail: {costantino.thanos,carlo.meghini,valentina.bartalesi,gianpaolo.coro}@isti.cnr.it

---

[1] https://www.go-fair.org/fair-principles/

– The third step is the actual expression (or materialization) of these relationships through a linking mechanism that allows the connection of several data islands, establishing thus data patterns.

The execution of these three steps lead to the creation of a connected research data space. By following these data patterns new information, previously unknown, can be discovered, that is, they constitute knowledge patterns. A new approach to knowledge production can emerge by following these patterns. The nature of this new approach to knowledge production is exploratory [2,3]. In essence, by following these patterns, a researcher can discover new insights about a research problem.

A technology that is instrumental in the implementation of this new approach, is Linked Open Data (LOD) [4,5]. This technology allows identifying archaeological resources via HTTP IRIs (such as for instance `http://example.com/mydataset`) making them web resources. These IRIs can then be used to express knowledge about these resources via one of the recommended Semantic Web languages (e.g. RFD[2], OWL[3]), and to create the links between them that constitute the patterns/graphs at the basis of the above described exploratory approach. For instance, the RDF statement, also known as RDF triple:

```
http://example.com/mydataset
http://www.cidoc-crm.org/cidoc-crm/P1297\_is\_about
http://vocab.getty.edu/aat/300054328
```

expresses the fact that my dataset (identified by IRI http://example.com/mydataset, the subject of the triple) is about (the property of the triple) archaeology, which in the Getty Art and Archaeology Thesaurus[4] is identified by the IRI `http://vocab.getty.edu/aat/300054328` and is the object of the triple. The statement can also be seen as a link between the subject and the object; the link is labelled by the property. In this sense, the statement is said to link those two resources.

A set of triples like the one above is called an RDF graph. An RDF graph is a most common example of Linked Data dataset. A list of properties is called a vocabulary. An ontology is a set of logical axioms that define the terms of a vocabulary according to a conceptualization [6].

RDF graphs have the potential of being the knowledge patterns needed to realize the exploratory approach

to the creation of knowledge, resulting from the interconnection of data islands. Moreover, given the pervasiveness of the web, and its global nature, LD technologies have also the potential of supporting interdisciplinary research, crossing the barriers inevitably created by a RI.

In the archaeological field, the necessity of aggregating data islands, in the context of "Oceans of Data", was at the center of the topics addressed in the 44th "Computer Applications and Quantitative Methods in Archaeology" Conference (CAA-2016) [7]. In particular, in the CAA context the need for well-defined common conceptual models/domain ontologies has been stressed in order to implement the LOD technology [8].

In the archaeological field, the LOD technology has been largely adopted since at least a decade, as witnessed by the numerous projects and tools available today, first of all the ARIADNE research infrastructure for archaeology[1]. As an example, the LOD technology has been used for representing and linking numismatic concepts [9]. The Implementation of this technology in the numismatic field was made possible thanks to the activities of the Nomisma.org, an open access thesaurus of numismatic concepts that conforms to the principles of LOD. In particular, it establishes stable IRIs for the description of coin types, and for the vocabulary terms used to describe these coins, with a focus on Greek and Roman coins.

However, there are still some limitations that prevent the full exploitation of the potential of the LOD technology. We have identified two main limitations.

– From a conceptual point of view, the number of existing types of relationships between datasets is far more than the relationships defined and axiomatized by existing ontologies. This limits considerably the potential of the exploratory approach.
– From a practical point of view, a considerable number of relationships between datasets are unknown or hidden/implicit. Therefore, it is necessary, first, to discover these relationships to express them in an RDF graph. We still lack a systematic study of the algorithms to carry out this kind of discovery, and this also limits the potential of the linking mechanism.

To further complicate the problem, any data space organized according to the LD paradigm is composed of a number of pre-established patterns/graphs that a researcher has to follow to obtain the information she/he is looking for. In essence, she/he has to traverse a static data space. However, the discovery activity requires a researcher be able to dynamically create links between datasets and in this way document collections based

---

on her/his cognitive state. The researcher's cognitive state is continuously updated as the investigation proceeds and, therefore, new information needs that require the discovery of additional relationships can arise. In essence, a researcher needs some tools that enable her/him to discover relationships among datasets dynamically. The above considerations motivate our effort for a new exploratory approach to knowledge production carried out in a dynamic context where the discovery activity is conducted. As a test case, we apply our approach to the archaeological domain.

The paper is organized as follows: In Section 2, the characteristics of the archaeological data space that mainly influence the proposed exploratory approach are described. In addition, the different types of relationships, both explicit and hidden/implicit, that can exist between archaeological datasets and that allow the creation of archaeological data patterns are described. In Section 3, different modes of exploring the linked archaeological data space are illustrated. In Section 4, the characteristics of the inference engines, that is, the software that discovers relationships between datasets are described. In Section 5, an example of discovering an implicit spatial relationship between two archaeological datasets is described. Finally, in Section 6, some concluding remarks are given.

2 The Archaeological Data Space

The Archaeological Data Space (ADS for short) includes the resources that have been collected by archaeologists during their research activities, whether on the field (e.g., excavations), in the laboratory (e.g., instrumental analyses), in meetings with other scholars (scientific reports, conference proceedings) or alone (master or PhD theses, scientific papers). From a technological point of view, such resources include therefore digital objects of various kinds, ranging from data to images, videos and texts. Furthermore, a resource may also be a collection including several instances of the kinds mentioned above. In fact, an ADS is a very heterogeneous space also from a granularity point of view: it may include data from very short and localized events, such as the sampling of a certain phenomenon or object, to much wider events such as the excavation of a site.

In the rest of this Section, we will describe the features of the archaeological data space, focusing first on datasets and then on collections, and using the general term of "Data Resource" (DR) to refer to an element in this space, whether an individual dataset or a collection.

2.1 Archaeological Data Resources

The archaeological DRs are organized and managed in various ways: some are traditional (relational) database management systems [10], others are geographic information systems [11], repositories, Linked Data datasets, and so on. The main subjects are the excavated units and the artefacts found there. Artefacts discovered in an excavated unit are mainly described in terms of their features (for example, ceramic type, dimension of artefacts, provenance, appearance, stratigraphic position, etc.). Also, the details of the data collection are described. Further, of paramount importance, it is the description of how artefacts are related to each other as well as in relation to artefacts found in other excavation units. In summary, the following types of DRs populate the archaeological data space:

– DRs that describe excavated units;
– DRs that describe fieldwork (fieldwork archives);
– DRs that describe artefacts discovered during an excavation activity and collection details;
– DRs that describe burials discovered during an excavation activity;
– scientific DRs that report the results of an in-depth analysis of the discovered archaeological material conducted by archaeologists;
– DRs that describe archaeological museum holdings.

Such a classification schema is supported by the ontology defined by the ARIADNE archaeological community [1] for its e-infrastructure[5]. Archaeological DRs are endowed with metadata. A dictionary/ontology documents the meaning of the variables that are included in the metadata scheme. A widely accepted ontology is CMRarchaeo[6] that enables the encoding of metadata about the archaeological excavation process. The goal is to provide means to document the excavation activity.

2.2 Archaeological Collections

Archaeological collections are groupings of archaeological resources related to some archaeological activity (such as excavations) or to some research activity in the context of which the members of the collection have been created [12]. These collections often include textual documents or images providing visual descriptions of excavated units, artefacts discovered, and other findings. In addition, a considerable number of published papers and reports is available that contain the results of an analysis activity of the found artefacts.

---

[5] http://legacy.ariadne-infrastructure.eu/
[6] http://www.cidoc-crm.org/crmarchaeo/home-3

The ARIADNE project, during its first phase of activity, has aggregated descriptions of over 1,5 million datasets, over 50 thousand textual documents, and over 40 thousand collections. This number is going to be increased by the just started follow-up project.

2.3 DR Identity

Each DR has distinguished characteristics that contribute in defining its identity. By DR identity it is intended a number of characteristics that make a DR definable and recognizable allowing, thus, to distinguish it from other DRs but also to establish relationships between different DRs. Identity must be an intrinsic characteristic of the DR [13]. Several characteristics concur to establish the identity of a DR; for the purpose of identifying relationships between DRs, we consider the following four characteristics of DRs:

Class. The archaeological community has identified five DR classes:

1. DRs that describe excavated units;
2. DRs that described artefacts discovered during an excavation activity;
3. DRs that describe burials discovered during an excavation activity;
4. DRs that report the results of an in-depth analysis of the discovered archaeological material conducted by archaeologists;
5. DRs that describe archaeological museum holdings. In essence, a DR class indicates that the data contained in a DR concerns the same subject or has a common theme, that is, there is a semantic relatedness between them.

Spatiality. The spatiality of a DR has two aspects: topological spatiality and geometrical spatiality. The topological spatiality indicates the present location of the DR on the earth surface. The geometric spatiality describes the geometry of the objects (for example, artefacts) described in a DR like their shape, extent, etc. The topological spatiality of a DR should be associated with its geometric spatiality. In order to describe the topological spatiality of a DR the ISO/TC211 standard can be used. It deals with the modelling of geographical information. In particular, the ISO Standard 19107 provides concepts for describing the spatial characteristics of geographical information. The geometrical spatiality, instead, can be described by a domain-specific ontology.

Temporality. A DR provides knowledge about some phenomenon or set of phenomena, that has occurred sometime in the past. The period of time during which the phenomenon of the DR has occurred is generally termed the temporal coverage of the DR. The temporal coverage of a DR is an interval of time, which may be of varying width, depending on the involved phenomenon. It occupies a position in a temporal reference system A problem, typical in archaeology, is the fact that, often, a temporal coverage denoted with the same name, for instance, "Bronze Age", in different geographical regions may refer to different periods in absolute terms. This problem is typically addressed by creating common reference resources that map such names to an absolute time scale. One of these resources is the perio.do gazetteer[7]. We can therefore assume, as already said, that the temporal coverage of an archaeological DR is available as a period on an absolute scale. However, it should be noted that the temporal coverage of a DR is often vague and, therefore, a fuzzy approach for representing time has been suggested. Worthwhile to mention is the proposal of defining several time categories concerning archaeological temporality. Six-time categories have been proposed, i.e., excavation time, database time, stratigraphic time, site phase time, and absolute time. These categories can be profitably used in order to establish temporal relationships between DRs. In conclusion, the identity of an archaeological DR is defined by its class, its spatiality both topological and geometrical and its temporality.

Metadata Schemes for Archaeological DRs. A metadata scheme "is a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality of values" (ISO 23081). The metadata scheme of a DR must, formally, define those elements that concur to establish the DR identity [14]. For each DR class, the metadata scheme will contain elements that characterize this particular DR class, for example, spatiality, temporality, et al. Having classified DRs into five classes implies that, also, the associated metadata will have different features related to each class [15].

2.4 Relationships between Archaeological Data Resources

Several relationships between DRs exist in the archaeological data space; for example, relationships between excavation units, between artefacts, between artefacts and their surroundings, between artefacts/excavated units and documents, both textual and visual, that describe
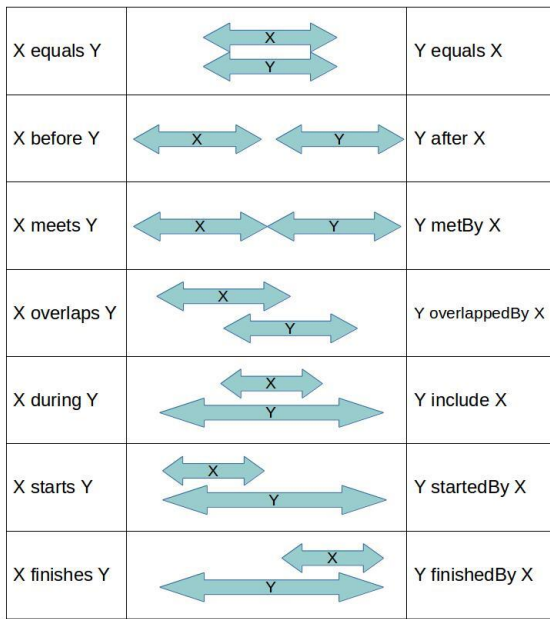
---

[7] https://perio.do/en/

X equals Y		Y equals X
X before Y		Y after X
X meets Y		Y metBy X
X overlaps Y		Y overlappedBy X
X during Y		Y include X
X starts Y		Y startedBy X
X finishes Y		Y finishedBy X

Fig. 1 A graphical representation of Allen's temporal relations

them. For our study, we have identified the following relevant relationships: temporal relationships, spatial relationships, spatio-temporal relationships, semantic relationships and anachronistic relationships.

Temporal relationships
Two DRs are temporally related based on an ordering relation between their respective coverage. A useful set of temporal relations between time intervals is given by Allen's relations. These relations form a jointly exhaustive set of pairwise disjoint relations covering all possible ways in which two-time intervals can relate. An illustration is provided in Figure 1.

The thirteen basic relationships given in the above Figure can also be combined in disjunctions that capture intuitive temporal notions. For instance, the coverage of a DR includes that of another DR if the time period of the former DR equals, or is included in, or starts or finishes the temporal period of the latter. Using these notions it is possible to link the DRs in a data-space based on their temporal relationships and exploit those links for addressing research questions, as it will be argued below.

Spatial relationships
A spatial relationship between two DRs is the relationship between the topological spatiality of the two DRs. It indicates a topological relation, for example, a distance relation/ a directional relation/etc. among them.

In essence, a topological relationship describes a relationship between DRs in space.

Spatio-temporal relationships
A spatio-temporal relationship is a relationship between the topological spatiality of a DR and its temporal coverage. These relationships are of paramount importance as they can be used for gaining new insights about archaeological DRs of interest. They allow, also, the outline of what information was available at a fixed time in history [16,17].

Semantic relationships
A semantic relationship between two DRs is the association that exists between the meanings of variables contained in the metadata schema associated with these DRs. Several semantic relationships have been identified [18]; of particular importance for the archaeological field is the Inclusion relationship, which describes situations where one entity comprises or contains other entities. Three different types of inclusion have been identified: class, meronymic, and spatial.

1. Class inclusion: is the standard subtype/supertype relationship often expressed as is-a , (A is-a B, where A is referred as the specific entity type of B). Other examples include relationships of classification, generalization, and specialization.
2. Meronymic inclusion is the relationship between something and its parts. Examples include the relationships: component-object, member-collection, phase-activity, and place-area.
3. Spatial inclusion is the relationship between an object and another object that surrounds it without being part of the surrounding object.

Therefore, a relationship between the geometrical spatiality of two DRs can be characterized as a kind of semantic relationship, for example, spatial inclusion.

Anachronistic relationships
An anachronistic relationship is a special type of temporal relationship. It is the relationship between a DR that exists and another DR that does not exist anymore. The anachronistic relationships are of paramount importance in the archaeological domain.

The automatic identification of temporal/spatial/semantic relationships between archaeological DRs is of paramount importance for the successful implementation of an exploratory approach to knowledge production. Therefore, the exploratory approach to knowledge production will be successful if inference engines will be developed that efficiently and effectively discover relationships between DRs distributed worldwide. Obviously, different

types of logic, depending on the type of the sought relationship, should be employed for implementing such engines: deductive, inductive, modal, causal, temporal, topological, etc. We envision, in the near term, the creation of libraries/catalogues of specialized inference engines for the archaeological domain. Such catalogues will enable the creation of specialized data patterns. Finally, we distinguish the relationships between DRs in two categories: explicit and implicit.

Explicit relationship

An explicit relationship between two DRs exists when it is represented by common variables in their respective metadata schemes. For example, in a relational database, an explicit relationship between two relations/tables exists when one table has a foreign key that references the primary key of the other table. Explicit relationships are intentionally created by the designers of the database schemes. Explicit relationships are discovered by a query processor; for example, in a relational database a query processor, based on the relational calculus, can identify explicit relationships between relations/tables.

Implicit relationships

An implicit relationship between two DRs exists when there are no common variables in their respective metadata schemes, but there exists a relationship (for example semantic, temporal, spatial) between variables of the corresponding metadata schemes. In essence, an implicit relationship is a hidden relationship that can be revealed to a researcher by an inference engine. The inference engine is based on a logic that depends on the type of sought relationship. Once revealed, a relationship can become explicit by connecting the two DRs involved in this relationship by a linking mechanism.

Discovering implicit relationships and making them explicit by a linking mechanism enables the creation of an interconnected archaeological data space. More implicit relationships are revealed and made explicit more tightly interconnected will result the archaeological data space. An interconnected archaeological data space enables the creation of data patterns. By a data pattern, we intend a directed graph whose nodes are DRs and whose arcs represent relationships between DRs. A data pattern may be cyclic or acyclic, depending on the relationships represented by the arcs. These data patterns contain implicit and often previously unknown information, i.e., knowledge. In essence, they constitute knowledge patterns [19]. A researcher, by following these patterns, can gain new insights. It could be possible to create data patterns that are characterized

by the type of relationship represented by the links between the DRs involved in the patterns. Data patterns are implemented by a citation mechanism. This means DRs should be endowed with an identifier assigned by an archaeological authority or with an IRI.

3 Exploring the Archaeological Data Space

Seeking data in the archaeological data space can be carried out in two modes: navigational querying or navigational browsing [20].

- In the navigational querying mode, data seeking occurs in an intentional way, that is, the user has a specific target in mind that is described via a linguistic expression, known as query; the query is submitted to the system that manages a DR of the data space; by processing the query, the system produces a subset of the queried DR containing all and only the data/documents of the DR that satisfy the given description. Successively, the user can refine her/his query, based on the information contained in the subset so far obtained. This refined query can be issued against the same DR or any other DR of the data space obtaining, thus, another subset that is more closed to her/his information needs. This mechanism can be iterated until the user succeeds to obtain the exact information she/he is looking for.
- In the navigational browsing mode, the user does not have a definite target in mind, and the data seeking occurs in an extensional way. The user navigates in a data space where the hidden relationships are made explicit by linking the DRs involved in a relationship. The user follows different data patterns within such a linked data space, in the hope that she/he might find an interesting DR.

In essence, the distinction between these two modes of data seeking is determined by the cognitive state of the user. The navigational querying mode is appropriate when the user knows exactly what data is looking for and where to search for the desired data. The navigational browsing mode is appropriate for an exploratory approach to knowledge creation as such approach facilitates an investigation activity to be started without a strong preconception.

Data Publication

Instrumental in the implementation of an exploratory approach to knowledge creation is data publication. By Data Publication, we mean a process that allows researchers to discover, understand, and make assertions about the trustworthiness and fitness for purpose of the

DRs in a data space. The ultimate aim of Data Publication is to make scientific data available for reuse both within the original disciplines and the wider community. Among the main functions that the data publication process performs, we distinguish the following two that are of paramount importance for the creation of data patterns: data registration and data semantic enhancement. The purpose of registration is to make a DR citable as a unique piece of work, while the purpose of semantic enrichment is to make it understandable. Once accepted for deposit, a DR should be assigned a "Digital Object Identifier" (DOI) for registration. A DOI [21] is a unique name (not a location) within the archaeological data space and provides a system for persistent and actionable identification of data. In addition, the DR should be assigned appropriate metadata. Instrumental in the publication of DRs is the development of domain-specific catalogs/registries where these DRs are published. In the context of the ARIADNE project, a catalog, AO-Cat, is under development. In this catalog, all the archaeological resources, events, as well as concepts will be described at a conceptual level supporting, thus, discovery and research activities. In addition, an authority in the archaeology domain will assign the DOIs to the DRs created by the archaeologists.

4 Inference Engines

As already said, the automatic discovery of relationships between archaeological DRs is of paramount importance for the successful implementation of an exploratory approach to knowledge production. Therefore, the development of software able to infer relationships between variables contained in the metadata schemes of different DRs in order to establish interconnections between DRs must be hastened. An inference engine should calculate a measure of dependence between variables, contained in metadata schemes, in pairs of DRs. Most of the data relationships can be modeled as functions, but not all are well modeled by a function. The modeling of data relationships is a domain-specific task and it must be supported by domain-specific vocabularies. We envision the development, in the near future, of inference engines specific for each type of relationship. This kind of software will enable the creation of "specialized" data patterns. Inference engines must be adequately described in order to enable potential users to find them. The inference engines should be described at three distinct levels [22]: the computational, the algorithmic and the implementation level. At the computational level, the logic of the abstract computational model is described. In essence, at this level, the

goal of the computation is described as the identification of a certain type of relationship between variables contained in the schemata of two DRs. As said in Section 2, several types of relationships can exist between these variables. The computational model, in essence, implements an appropriate logic that must guide the discovery of a particular type of relationship sought by a user. Examples of logics, that can be adopted, include conventional, modal, causal, temporal, topological etc. At the algorithmic level, the input and output values to the inference engine are described. The input values are the values of two variables, contained in schemes of two DRs, that the engine has to infer whether exist a relationship among them. The output values are the existence or not of a certain relationship between the input variables. At the implementation level, the inference engine is a software with a discoverable and invocable interface. All these three levels of description are included in the metadata of the engine. As for the DRs, also the inference engines must be published in order to make them discoverable. This means that an archaeology-specific catalogue has to be developed. This catalogue should include, at least, for each engine:

- a description that is contained in the metadata;
- an identifier IEI (InferenceEngineIdentifier);
- the type of the inference engine;
- how to request the inference engine;
- how the inference engine delivery is fulfilled.

A graphical representation of the entire proposed workflow that implements an exploratory approach to archaeological knowledge is reported in Figure 2.

4.1 An example of discovering an existing implicit data relationship between two data archives

Let's consider two archaeological archives containing information about coins [23, 24]. The Cambridge Fitzwilliam Museum archive (FWM)[8] that contains information about metals and coins discovered during excavations or coming from various acquisitions or donations for a total of 1670 numismatic records. The metadata schema of this archive includes the following attributes (variables):

- coin maker,
- production location,
- mint,
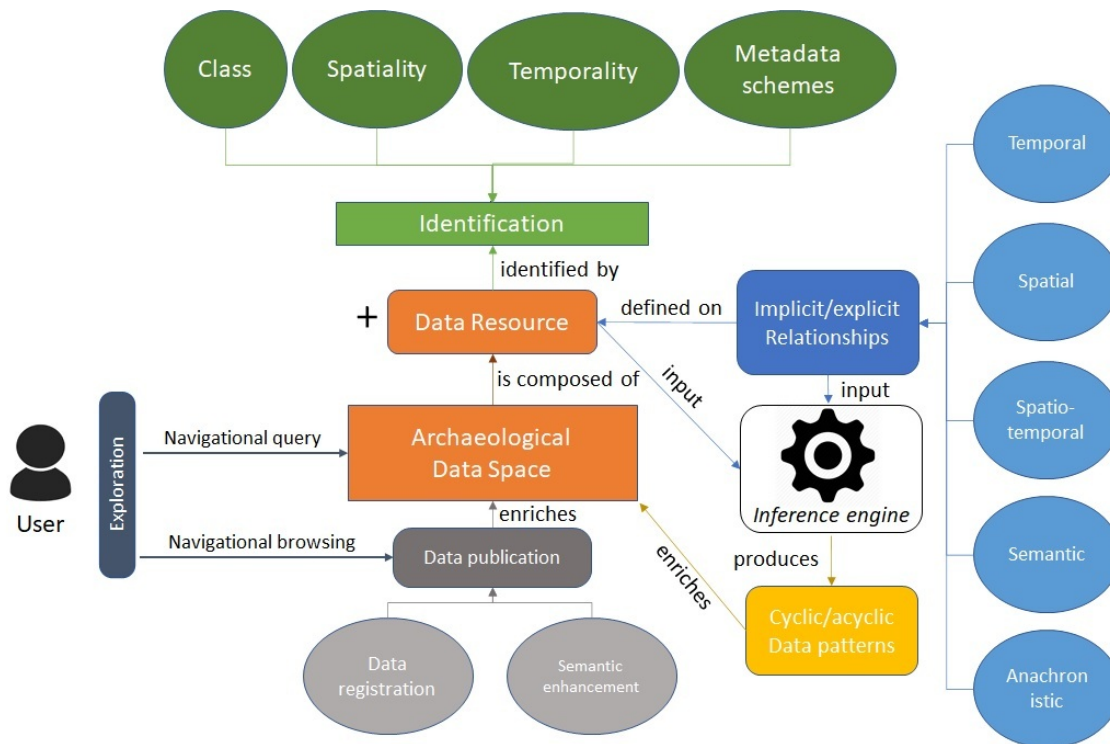- coin type,
- category,
- coin name,
- inscription,

---

[8] https://www.fitzmuseum.cam.ac.uk/about-us/collections/coins-and-medals

Fig. 2 A graphical representation of the proposed workflow that implements an exploratory approach to archaeological knowledge

- dimensions,
- production technique, and
- references to images.

A dataset of 630 records coming from the Soprintendenza Archaeologica di Roma (SAR) containing information about the physical features of the coins. The metadata schema of this archive includes the following attributes (variables):

- physical features of the coins;
- the region in which a specific coin was minted;
- chronology information (the age/century/period during which coin minting took place) ;
- obverse/reverse inscriptions of iconography;
- current location of the specific exemplar.

Between two variables of the metadata schemata of the above two archives, i.e., production location (FWM) and region (SAR) there exists a spatial inclusion relationship, that is the relationship between two objects such that one of the two surrounds the other without being part of it.

Let's suppose that the above two archives are implemented as two relational tables and a researcher wants to issue the following query against these two tables:

"Find the physical feature of the coin whose name is X"

A query processor based on the relational calculus is not able to answer this query although the sought information can be derived by appropriately combining the information in the two tables with the addition of an inference step. However, a query processor based on an appropriate spatial logic and supported by a geographic ontology should be able to answer correctly the query. The steps of the query execution, that this query processor has to undertake, are the following (Figure 3):

1. The query processor takes the name of the coin (X) as the input.
2. The query processor extracts the name of the location where the coin was produced (Y) from the FWM archive.
3. From the ontology, it infers that the location (Y) is part of the region (Z). This is the inference step.
4. From the SAR archive, it extracts the physical features (W) of the coins minted in region Z.
5. Finally, the query processor infers that the physical features of coin X are W as the production location Y of coin X belongs to the region Z.

In essence, a so enriched query processor is able to discover an implicit data relationship and answer a query that depends on this relationship.
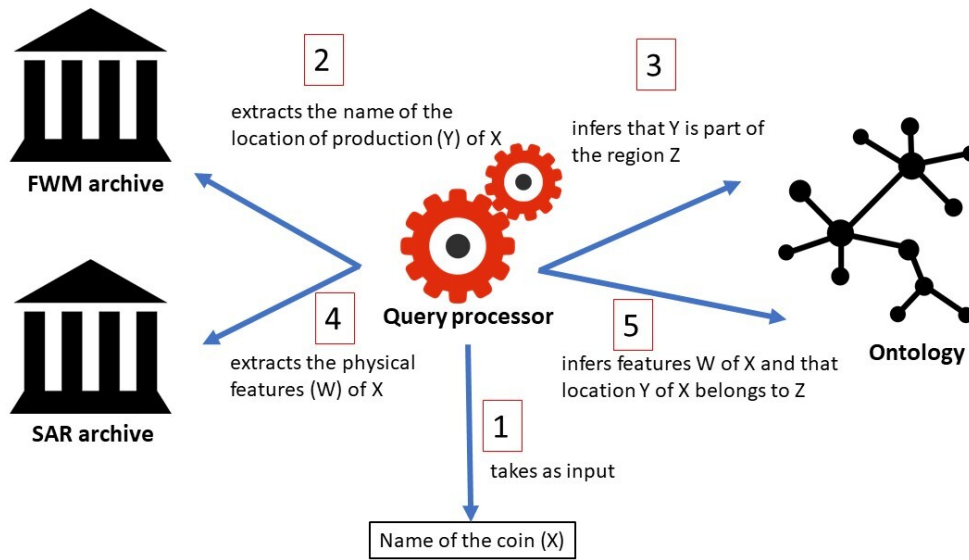
Fig. 3 The steps of the query execution that the query processor undertakes

5 Concluding Remarks

In this paper, we have outlined a new approach to archaeological knowledge creation. The current scientific archaeological context is characterized by intensive digitization of the research outcomes. Several relationships can exist among these outcomes. Some of them are explicit, whereas others are implicit or hidden. By materializing these implicit or hidden relationships through a linking mechanism, several patterns can be established. These knowledge patterns may lead to the discovery of information that was previously unknown. A new approach to knowledge production can emerge by following these patterns. In the paper, we have reported our effort to depict this new approach using Linked Data and Semantic Web technologies (RDF, OWL). Instrumental in the implementation of this approach is the ability of the researchers to create data patterns in the archaeological data space in a dynamic way by exploiting existing both explicit and implicit relationships among the data resources. Realizing this approach implies the implementation of an infrastructure and the development of appropriate tools able to support the researchers in this activity. The infrastructure should provide: (i) linking services to allow the dynamic creation of linked data patterns; (ii) intermediary services in order to make the holdings of the archaeological data resources discoverable, accessible, understandable and reusable; and (iii) navigational services to allow researchers to navigate the linked data patterns. Concerning the tools for the automatic discovery of data relationships, we emphasize the urgent need for the development of specialized inference engines. We envision that in the near future these pre-conditions will be fully implemented thus enabling an exploratory approach to knowledge creation.

References

1. Meghini, C, Scopigno, R, Richards, J., Wright, H., Geser, G., Cuy, S., ... and Vlachidis, A. (2017). ARIADNE: A research infrastructure for archaeology. Journal on Computing and Cultural Heritage (JOCCH), 10(3), 1-27.
2. Yu C H (2010) Exploratory data analysis in the context of data mining and resampling. International Journal of Psychological Research, 3(1): 9-22,Universidad de San Buenaventura DOI: https://doi.org/10.21500/20112084.819
3. Idreos S (2013) Big Data Exploration. Big Data Computing, 3 ISBN: 9780429101366
4. Bizer C (2013) Interlinking Scientific Data on a Global Scale, Data Science Journal, 12: GRDI6–GRDI12 DOI: http://doi.org/10.2481/dsj.GRDI-002
5. Auer S, Bryl V, Tramp S (eds.) (2014) Linked Open Data–Creating Knowledge Out of Interlinked Data: Results of the LOD2 Project, LNCS, vol. 8661, Springer DOI: 10.1007/978-3-319-09846-3
6. Guarino N (ed.), Formal ontology in information systems: Proceedings of the first international conference, FOIS'98, June 6-8, Trento, Italy (1998)
7. Matsumoto M, Uleberg E (eds.) (2016) CAA2016: Oceans of Data: Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology
8. Ore, C E S (2018) Oceans of Data: Creating a Safe Haven for Information.

9. Gruber E, Matsumoto M, and Uleberg E (2018) Linked Open Data for Numismatic Library, Archive and Museum Integration. In: CAA2016: Oceans of Data: Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology (pp.55). Archaeopress Publishing Ltd.

10. Kadar M, Data modeling and relational database design in archaeology. Acta Universitatis Apulensis, 3: 73-80, (2002)

11. Smith J R, Database design, archaeological classification and geographic information systems: A case study from southeast Queensland. Unpublished thesis (PhD), University of Queensland (2000)

12. Miller T M (2012) Specify for Archaeology: A Proposed Data Model for Archaeological Collection Database Management, Unpublished thesis (PhD), University of Kansas

13. Wynholds L(2011) Linking to scientific data: Identity problems of unruly and poorly bounded digital objects, International Journal of Digital Curation DOI: https://doi.org/10.1080/09558543.1991.12031189

14. Farnel S., Shiri A., Metadata for research data: current practices and trends. In: International Conference on Dublin Core and Metadata Applications, DC-2018, Porto, Portugal, pp. 74-82 (2014)

15. Willis C, Greenberg J, White H, Analysis and synthesis of metadata goals for scientific data, Journal of the American Society for Information Science and Technology, 63(8): 1505-1520 (2012)

16. Belussi A, Migliorini S (2017) A spatio-temporal framework for managing archeological data. Annals of Mathematics and Artificial Intelligence, 80(3-4): 175-218 DOI: https://doi.org/10.1007/s10472-017-9535-0

17. Hallot P, Billen R (2016) Enhancing spatio-temporal identity: States of existence and presence. International Journal of Geo-Information, 5(5). Multidisciplinary Digital Publishing Institute DOI: https://doi.org/10.3390/ijgi5050062

18. Storey V C (1993) Understanding semantic relationships. The VLDB Journal, 2(4): 455-488, DOI: https://doi.org/10.1007/BF01263048

19. Gullo F (2015) From patterns in data to knowledge discovery: what data mining can do. Physics Procedia, 62:18-22 DOI: https://doi.org/10.1016/j.phpro.2015.02.005

20. Waterworth J A, Chignell M H (1991) A model for information exploration. Hypermedia, 3(1): 35-58 DOI: https://doi.org/10.1080/09558543.1991.12031189

21. Paskin N (2005) Digital object identifiers for scientific data. Data science journal, 4: 12-20 DOI: 10.2481/dsj.4.12

22. Thanos C, Klan F, Kritikos K, Candela L (2017) White Paper on Research Data Service Discoverability. Publications, 5(1) DOI: 10.3390/publications5010001

23. Gerth P (2016) ARIADNE Deliverable D14.2: Pilot Deployment Experiments. ARIADNE Project. Available at http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/d14-2-pilot-deployment-experiments/. Accessed 21 April 2021

24. Felicetti A, Gerth P, Meghini,C, Theodoridou M (2015) Integrating Heterogeneous Coin Datasets in the Context of Archaeological Research. In: EMF-CRM@ TPDL, pp. 13-27