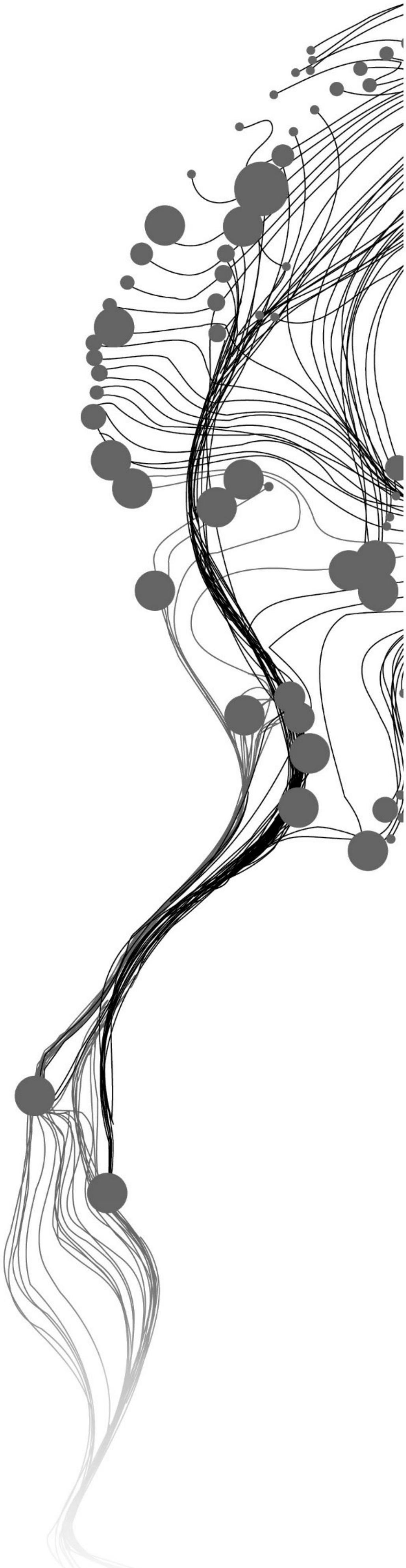
Improving the usability of Commercial Archaeology data through semantic integration.

ALEXANDRE ROBERTO PEIXE
August 2022

SUPERVISORS:
DRS. B.J. KÖBBEN
DR.IR. R.L.G. LEMMENS



Improving the usability of commercial archaeology data through semantic integration

ALEXANDRE ROBERTO PEIXE
Enschede, The Netherlands, August 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Drs. B.J. KÖBBEN
Dr. Ir R.L.G. LEMMENS

THESIS ASSESSMENT BOARD:
Dr F. O. Ostermann (Chair)
Dr M. C. Chipofya (External Examiner, ITC-PGM)

ABSTRACT

A source of archaeological data not explored in its full potential is commercial archaeology, which is mandatory by law before construction or renovation works. In the Netherlands, the Council for Culture acknowledges that 90% of the archaeological projects are done by private archaeology companies. This hints at a considerable amount of data being misplaced or completely ignored, as the nature of the commercial market does not provide enough time and resources to perform all analysis needed. The general objective of this work is to improve the usability of commercial archaeology spatial datasets that are available through public repositories. This research intends to address it by testing a method to enhance their explorability for future researchers. The study involves different knowledge areas and encompasses digital archaeology, the Dutch commercial archaeology norms, the FAIR data principles and the implementation of ontologies and knowledge graphs. The work uses a case study of an archaeologist studying palaeolithic sites by integrating commercial archaeology data to evaluate their presence within the Enschede municipality. To achieve this it was divided into six steps: (1) a source data assessment to evaluate the current datasets; (2) a processing data step to extract and integrate the finds recorded; (3) a data standardization step, as each company has its own set of standards; (4) a user assessment to evaluate how the data have been currently used; (5) the definition of a proof-of-concept geoportal functional requirements and as the last step (6) adapting an ontology for converting the integrated data into knowledge graphs. The integrated dataset contains 4.235 finds extracted from 32 different datasets. A direct search for Palaeolithic artefacts in the public repository finds three reports matching the researcher's interest. Searching the integrated file in the geoportal it rises to eight matches, and the knowledge graphs obtain the same result. However, in the Neolithic, knowledge graphs retrieve 10 matching reports instead of eight obtained by searching the integrated file. Semantic integration also makes the implementation and maintenance of the whole system easier as its classes can be interrelated without having to modify the basic data itself. The process reinforces the importance of planning the data for reusability, as it was not possible to extract all the information available partly due to a lack of specific knowledge of Dutch typologies and periodizations, but also due to the lack of standardization of concepts and codifications. Professionals and authorities pointed out the lack of depth in the artefact analysis that makes difficult its reuse. However, the discussion about the data falling short in information is only possible if the data properties are available for comparison. The use of knowledge graphs allowed the data to be queried in more useful ways, even though the ontologies used can always be improved to enhance interoperability, but this asks for a team effort to ensure their effectiveness. In any case, the results have shown how the knowledge graph format can allow queries to reach more results and enrich the exploration of the dataset content.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. Many thanks to my supervisors, Barend and Rob for their patience and support, not to mention invaluable feedback and suggestions that helped a lot.

Thanks to the ITC for awarding me the ITC Excellence Scholarship Programme which allowed me the financial means to complete this course. And finally, thanks to my girlfriend, and family who endured this long process with me, always offering support and love.

Also thanks to Stanislav Ronzhin who helped with the important insights on the knowledge graphs.

I would also like to acknowledge the contribution of the team of the archaeology department in the Nijmegen municipality and from the Archaeology Station at DANS.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background and justification

The amount of data available for archaeology has been increasing, as it also happens in many other scientific fields. It does not affect all of its branches, places, and periods in the same way, but it is perceived as an issue by several authors since the 1970s. The topic has been brought more and more into the discussion with the rising popularity of data analysis as a subject (Clarke, 1973; Richards, 1997; Kletter & De-Groot, 2001; Allison, 2008; McCoy & Ladefoged, 2009; McCoy, 2017).

For the sake of simplicity, the majority of archaeologists prefer to obtain their data from the field and then integrate them with interpretations and conclusions from previous works. It seems easier than tackling the challenge of integrating legacy data, as they have been called. In Digital Archaeology (the branch of archaeological science devoted to the application of information technology to archaeological data), legacy data is seen as data that was created following practices and methods which are no longer compatible with contemporary standards, and in most cases are not in any digital format (Allison, 2008).

However, a huge source of data usually not explored in its full potential is the so-called rescue, preventive, contract or commercial archaeology, which is mandatory by law in many countries before all sorts of construction or renovation works. This branch of archaeology is defined by Demoule (2012), as the one that is practised by private companies and is dedicated to performing its services for developers surveying and excavating future construction sites for profit. The data produced is usually deposited in a university, research or governmental institution repository. But that's usually the end of the line for it, as it is not always simple to integrate it into the analysis.

To put it in perspective, in 2019, a study by the Chartered Institute for Archaeologists estimated that almost 75% of the personnel working in Archaeology in the United Kingdom were employed by private companies practising commercial archaeology (Aitchison, Rocks-Macqueen, 2020). Archaeology mainly relies on manual labour, which makes a direct connection between where the bulk of the manpower is applied and its main data source. In the Netherlands, the Council for Culture also acknowledges that 90% of the archaeological projects in the country were carried out by private archaeology companies (Raad voor Cultuur, 2022).

This hints at a considerable amount of data being, at best, misplaced or even completely ignored, as Pearson (2019, online) points out, "archaeologists working in commercial archaeology usually have to be generalists". They don't have time or resources to focus on a specific period or a particular theme, because they are spatially bound to the area they are digging for the rescue project. This means that in the morning they can be digging an Early Modern layer (roughly 15th to 18th centuries) and in the afternoon they are on a Neolithic layer (in Europe corresponding roughly between 7.000 and 1.700 years before present, depending on the region).

The commercial archaeology data are fragmentary in most cases, as it is bound by the area intervened by construction, renovation or infrastructure improvement. The information collected is seldom enough to answer archaeological questions on its own, its main task is to assess what is the archaeological value of a specific area and evaluate the disturbance caused by the works proposed to be done there. The commercial archaeology data are fragmentary in most cases, as it is bound by the area intervened by construction, renovation or infrastructure improvement. The information collected is seldom enough to answer archaeological questions on its own, its main task is to assess what is the archaeological value of a specific area and evaluate the disturbance caused by the works proposed to be done there. If the data never reaches a specialist, a lot of information can be lost. The archaeologists are just digging it from one place and burying it in another. The data at their "second burial" site sometimes becomes harder to be accessed because it can be not only geographically far away from its original location, but also under many layers of bureaucracy. More often than not, the soil is easier to dig than the paperwork required to get the data deposited in certain institutions.

This can lead to strange paradoxes that can involve even the United Nations (UN) Sustainable Development Goals. The UN General Assembly (2015) resolution states in its goals list "11.4 Strengthen efforts to protect and safeguard the world's cultural and natural heritage", and the indicator for it is the "Total per capita expenditure on the preservation, protection and conservation of all cultural and natural heritage, by the source of funding (public, private), type of heritage (cultural, natural) and level of government (national, regional, and local/municipal)." However, in the current scenario for Archaeology, is noticeable how this increase in expenditure may result in no perceivable contribution to achieving the sustainable development goal.

The current approach for the geospatial archaeological data is to incorporate the geospatial information produced as a single point, normally representing the archaeological site in systems like the Ariadne Portal (Ariadne Plus Project, 2021), for instance. This method is followed also by the majority of heritage protection governmental institutions, as a way to record the excavations done in their area of interest or responsibility. The remainder of the information is, at best, in the next layer under it. This layer is seldom online, and when it is, the different files and table formats usually do not allow automatic access by queries and the rediscovery of the data recorded in these files relies solely on chance.

Archaeological excavations have recorded spatial data in three dimensions since the early 20th century. Even before geographic information systems existed, archaeologists had been using points, lines and polygons to represent artefacts, ditches or roads, and buildings respectively. Although those records were kept only on 2D drawings and tables, this information can potentially be recorded with at least its grid (x, y) and layer (z).

However, the extraction of these data is done individually, segmented by research group, region, period or research institution. The concern about this segmentation was raised by several authors from Clarke (1973) to Pearsons (2019). Talking with colleagues from Brazil, Portugal, Italy and the United Kingdom it is striking how pervasive the problem looks, as the legislators usually create regulations focused only on mandatory data collection and safeguard, but they do not show the same care for their posterior usability.

This lack of attention can be noticed in the structure of the data stored, an interested researcher would usually have to guess what keywords were chosen for describing it. This adds an extra challenge for the researcher and also limits the results retrieved from the repository. Increasing the number of keywords to enhance the system has its limits as it still depends on the exact wording to work. One improvement that could help with this problem would be to expand the meaning of the terms commonly used in searches in a way that logically connects them with other relevant or related words.

For this purpose, knowledge graphs have been used to enhance knowledge integration in different fields, including archaeology (Bakker, 1987; Stokman, F.N. & Vries, 1988; Bergamaschi et al, 1999; Zhang, 2002; Binding, 2015, Mountantonakis & Tzitzikas, 2019; Ronzhin et al., 2019). As the name infers, they represent the knowledge in the form of a graph, where entities (that can represent material things or concepts) are modelled as nodes and the edges of the graph present the relationships connecting them (Ronzhin et al., 2019). Many of the above-mentioned studies apply the Resource Description Framework (RDF) as a viable solution for its implementation, as the framework provides a working model to operationalize the data interconnectivity.

According to Miller (1998: 17), the RDF "defines a simple, yet powerful model for describing resources" and also "provides the ability for resource description communities to define the semantics". This is achieved by using RDF statements to represent the interconnected data. These statements are structured as a subject-predicate-object, usually called a triple. The subject is the entity being described, the predicate states the relationship and the object is the entity that is related to the subject.

A crucial step for this implementation is to structure an ontology from which the semantic definitions will be drawn. In this sense, ontology represents "a formal description of knowledge as a set of concepts within a domain and their internal relationships, formally specifying their components as instances, classes, attributes, relations, restrictions and rules" (Ontotext, n.d.). All scientific fields use a shared vocabulary to express their concepts, but the same word can be used in other fields with different meanings, or differ greatly in its common use, which can generate misunderstandings. The RDF triple format demands an extra layer of standardization to remove ambiguity and ensure machine readability to make it able to retrieve the information. For this, the role of the ontology is to limit the number of types of relationships and keep the system manageable (Zhang, 2002; Ehrlinger & Wöß, 2016).

Even though theoretically sound, translating the actual datasets into RDF statements can become very challenging. Among the challenges is the dependency on how the data are described and organized. Any aggregation or reclassification of the data is limited by the amount of information already provided. To avoid the ambiguity mentioned earlier can demand using a less detailed classification than the ones provided in the data, but able to encompass all entries. Otherwise, it could become too arbitrary and not be able to assemble the information provided in a meaningful way.

In an attempt to address these issues, we did this research which follows the current procedures needed to perform a search in the current state of these datasets and propose ways on how it could be improved using a different structure.

1.2. Thesis structure

The thesis is structured in six chapters, including this introduction. Chapter 2 presents what we try to accomplish and what specific objectives are being pursued. The third chapter provides a literature review about how archaeologists approach digital means and translate their finds into organized datasets and how this process is guided by the legislation regarding heritage protection in the commercial archaeology environment. It also addresses how the FAIR principles could be useful for such datasets and how the knowledge graphs structure can contribute to finding the data.

The fourth chapter focuses on describing the operationalization of the process, starting with a presentation of the study area, and then evaluating the way the data for this study area can be retrieved from the appropriate public repository. The proper work starts with an overview of the process to be proposed, the data processing involves their extraction from the reports and their merging into a common table. It also includes the contribution of archaeologists that work in the Netherlands and their familiarity with these datasets. Their insights can give a better grasp of the uses and problems found in the current state of these datasets. The presentation of these merged results through a geoportal that gives the option to search the data using their content at a level that was not possible in the original repository and finally how the data can be structured into knowledge graphs to add an extra layer of meaning to the stored data and enable a broader spectrum of searches in their content.

The fifth chapter describes the results obtained from the data processing and what decisions needed to be taken to integrate these datasets. It will also present the tests performed for assessing the usability obtained at the end of the whole process comparing it with the process needed to be taken before. The final chapter will present conclusions and recommendations to discuss what was possible to be accomplished and the shortcomings and problems encountered. Not only about the processing and integration of those datasets, but also relating to the possibilities and issues raised about the commercial archaeology data. It will also include suggestions on possible proceedings of the study and what steps could be taken in the direction of making the original data a little more suitable for this kind of integration.

2. RESEARCH OBJECTIVES

This research intends to address the fragmentation of archaeological data stored in public repositories by testing a method of enhancing their explorability as a way to improve their usability for future researchers. Explorability, as understood here, is the way the user can explore and sort the contents of the dataset without the need to individually open the dataset itself.

It assumes the reuse of the data is linked to not only the ability to find it and access it but also to how explorable they are. Allowing better searches into their content can help researchers to link and integrate them in a way that all stored data can contribute their part to a better understanding of our archaeological past.

The questions that will guide the work for a proper evaluation of the procedures to be taken are:

1. How is the data commonly produced by Commercial Archaeology stored?
2. What kind of archaeological information can be found in these datasets?
3. Are there standards for naming or codification used for creating these datasets?
4. How is the geospatial information recorded in them?
5. Are these datasets being used by archaeologists in the Netherlands?
6. What is required to integrate those datasets in a single geospatial visualization?
7. Can the knowledge graphs offer a way to retrieve information that would be harder to be obtained otherwise?

The general objective is to improve the usability of commercial archaeology spatial datasets that are already available through public repositories. This is meant to be accomplished by enhancing the ways the data can be explored and it aims to relieve the future researcher from having to open several datasets and file reports one by one to evaluate which are relevant.

To achieve the general objective the work will be structured in a set of smaller steps to accomplish the main task. It starts with the assessment of the existing data to acknowledge how the expected fragmentation appears in a real-world setting.

Departing from that, the information from these datasets will be extracted to create a single standardized source with the capability for geospatial visualization of its content. This study does not intend to replace the individual reports, but to test a way to improve their usability for future researchers. The improvement is meant to be achieved by allowing these researchers to select the datasets containing the information they need without browsing hundreds of reports unnecessarily.

In the next step archaeologists in the Netherlands will be contacted to assess the current dataset usage and to have an idea about the strategies they use for it. Also, they were questioned about possible applications for the data and tools that could enhance their analysis.

Using the answers as a starting point, the requirements for presenting the information in a geoportal were derived. It will also include how the current data can be queried, and how the answer will point back to the original datasets in the repository.

In the end, it will the data will be converted into knowledge graphs using a chosen, or adapted, ontology to be explored the possibilities of querying that are not possible, or easily done otherwise.

These specific objectives can be summarized as follows:

I. Assess the current state and capabilities of the commercial archaeology datasets in the Dutch deposit institution (DANS);

II. Devise a way to integrate the different finds datasets and make the geospatial information included in them findable, accessible and able to be visualized;

III. Contact archaeologists working in the Netherlands to evaluate their usage and opinion about the commercial archaeology datasets.

IV. Make the integrated data accessible through a proof of concept geoportal that allows searches based on their content;

V. Explore knowledge graphs as an option to enhance searches in the dataset in ways not possible, or too difficult otherwise.

3. LITERATURE REVIEW

The study involves different knowledge areas in an intersection between humanities and technological sciences. The analysis of the content recorded into commercial archaeology datasets will be backed up by digital archaeology concepts to understand how the finds are recorded using digital means and why. But to understand how and why specific information was selected to be recorded the way they are and why this method was chosen among others it will be also important to delve into how commercial archaeology is organized in the Netherlands. Some of the principles are defined by international agreements and standards, but the adaptation to the local practices plays an important role in the making of the heritage protection environment and affects how the excavations are conducted and how the records are produced and stored either digitally or otherwise.

The measuring of usability increment can feel subjective if based solely on direct comparisons, as it can be difficult to objectively describe one's experience. The FAIR principles are a set of guidelines that intend to enhance the reusability of scientific data by improving the infrastructure of the repositories. The acronym FAIR stands for Findable, Accessible, Interoperable and Reusable and they seem to fit well in this case, as they can be used as a mark to compare the evolution of usability based on their descriptive steps. The proposal to explore knowledge graphs also aims to contribute to the enhancement of FAIR principles, as it is expected that increasing the chance of a resource being found, will also result in facilitating the task of those who want to reuse it.

3.1. Digital archaeology

The term Digital Archaeology can have many definitions and it is used in different fields to represent different concepts. This particular study will understand it as Daly & Evans (2005: 3) propose the study of the five main areas computers have impacted Archaeology. However, here only two of those topics will be explored:

- Recording and representing archaeological data;
- Increasing the potential to share information with both peers and the public.

However, the rise of digital methods in Archaeology raised concerns in some authors about what they call data fetishism. Their complaints focus on data-driven analysis becoming so detached from the actual findings, that is difficult to relate it to the past behaviours they are presumed to represent. (Huggett, 2004, 2015, 2020; Gattiglia, 2015).

This can often occur when the software packages and algorithms used are based on implied assumptions that are not well described. It can become problematic for a posterior researcher, who will not be aware d of those effects, and use these interpretations as if they were raw data. These conclusions, if taken to face value, can result in later applications linking findings to behaviours that should be not related.

As a way of mitigating this, one of the main reminders presented is that Archaeology does not have something that can be called raw data. Archaeological data are "theory-laden, process-laden, and purpose-laden, created by different people, under different conditions, for different purposes, at different times", as Huggett (2022:103) points out.

The data result from a process of choice and analysis since it is collected from the ground. Not all finds have their exact location pinpointed or are catalogued individually, or even go through complete cleansing and classifying processes.

The whole process of data creation, analysis, archiving and sharing has been discussed as archaeologists grow concerned with a rush for the adaptation of theoretical approaches to fit into digitization and not the opposite (Stenborg, 2018; Garstki, 2022; Harrison, 2022).

Describing and recording digitally the data in a proper way is more difficult than it looks at first glance and it has been fuelling many academic studies (Backhouse, 2005; Bradley, 2005; Finat et al, 2010; Caraher, 2015, 2019; Huggett, 2015, 2022; Reilly, 2015; Burg, 2017; McCoy, 2017; Rabinowitz, 2022).

Some authors (Caraher, 2015, 2019; Huggett, 2015, 2022; Harrison, 2022) consider that the main problem with the current digital approach is to convey more importance to the data collection itself than to the information it can convey. They advocate a different approach, what they call Slow Archaeology, using a three S model (Slow, Small and Sure) in opposition to the big data three Vs model (Velocity, Volume and Variety). They argue the attempts to apply the latter to Archaeology, especially focusing on velocity, have not created the expected advances and put too much attention only on the first step (data creation), but falter to produce a sustainable framework for treating archaeological data. They consider the three S model as more suitable to fulfil the needs of the research and of the community it takes place.

Due to the non-renewability of the archaeological data, professionals discuss the necessity of collecting more of it. Huggett (2022: 101) reasons that "increasing the amount (or size) of data does not necessarily increase the information that may be derived from it". The concern about it comes from the technological advances that later could extract more information from the data being underused now. There are plenty of examples of this happening in the past. For decades burnt coal from prehistorical sites was discarded because they were of no use until the 1950s. The development of radiocarbon methods put them as the main proxies for determining the age of the whole site. And even then, many projects cannot afford to use it as it is relatively expensive depending on the country and the project's available funding, so still to this day, there are many where the current technology capabilities cannot be explored in their full potential.

The merits of Slow Archaeology are not the focus of the current study, but they highlight the archaeologists' perception that archaeological data specificities are not always compatible with data approaches currently applied in other scientific fields. The "Slow" approach suits well for academic research, as it attempts to link data creation to research needs and objectives, through a research design that seeks to address the whole data cycle from creation to curation and sharing.

However the data which will be used in this study not only suffer to a greater or lesser degree from the problems pointed out by digital archaeology, but they are also constrained by the rules and regulations that guide the practice of commercial archaeology in the Netherlands.

3.2. Commercial archaeology in the Netherlands

The approach suggested by the slow archaeology bumps into the lack of freedom for research design that commercial archaeologists endure. Nonetheless, the scope of single commercial project datasets is usually too small to make sense alone, so it would also benefit from a widely integrated approach as slow archaeology proposes. Commercial archaeology's position in the friction point between heritage management, construction and infrastructure industries, more often than not takes away the control over most of the variables during the fieldwork from archaeologists. They cannot expand the research area, they cannot extend the excavation period, and they do not have time or funding to go deeper into any of the finds understanding, or how they connect with the regional bigger picture. Since 2016, with the implementation of the Heritage Act, the Dutch legislation does not require an excavation permit for archaeological research, it has been replaced by a regulated certification for parties that carry out this work. There is no full description of requirements for the activity in the Act, it just states that the certified part "shall ensure that when an excavation is carried out the actions performed and the archaeological finds discovered are documented, the finds are preserved, and a report is drawn up summarising the results" (Heritage Act, 2015: 17).

The act defines an archaeological find as "a remain, object, or another trace of human presence in the past originating from an archaeological monument". (Heritage Act, 2015: 6). Archaeological monuments are sites considered "part of cultural heritage due to the remains, objects, or other traces present there of human presence in the past, including said remains, objects, and traces" (Heritage Act, 2015: 6).

The act prohibits, without a proper certificate, any kind of excavation. For this purpose, excavation is loosely defined as "actions involving the detection, investigation, or acquisition of cultural heritage, or parts thereof, which results in a disturbance of the soil or disruption or total or partial displacement or removal of an archaeological monument or underwater cultural heritage" (Heritage Act, 2015: 17). In this study, the focus will be only on the procedures leading to the actual excavation of findings on land, as there are completely different protocols for underwater sites, and it will also not comprise their posterior conservation and storage.

A more thorough description can be found on the Foundation Infrastructure Quality Assurance Soil Management (*Stichting Infrastructuur Kwaliteitsborging Bodembeheer* – SIKB) website and brochures (https://www.sikb.nl/archeologie/). The SIKB is a network organization aimed to integrate government and industry efforts to develop guidelines for not only archaeology but also soil, water, soil protection and data standards. It publishes the requirements and standards for archaeological activity through Assessment Directive (*Beoordelingsrichtlijn* - BRL) and the protocols called Quality Standard for Dutch Archaeology (*Kwaliteitsnorm Nederlandse Archeologie* - KNA). The BRL guidelines are meant to regulate the way certification bodies test whether an organization meets the technical requirements for obtaining or maintaining a certificate. The KNA protocols indicate the requirements set for archaeological work, from inventory research to the preservation of archaeological material.

The archaeological investigation aims to identify and evaluate possible archaeological remains and it takes place in steps. Each one of them contributes to deciding about possible or necessary follow-up actions. It starts with relatively simple research methods in its initial phase progressing to more complex and expensive activities when it is required. It is not always necessary to conclude all phases, as an earlier result may show that there is no reason for further investigation/ However in some cases, the results of the mandatory archaeological survey can demand an adjustment of the plans if there are sites worth preserving in the plan area, or if it is needed a full archaeological excavation before the project construction. (SIKB, 2007).

Any person or organization who carries out projects in which the soil is disturbed must take into account the possibility of existing archaeological finds in the area. Due to this, an archaeological preliminary investigation is mandatory. The preliminary investigation consists of two parts: the Office Investigation and the Field Survey, each with associated standard reports. These are also written proof of compliance with the licensing authority's requirement and must give enough information for the follow-up decision about the site, conservation in situ, release or excavation (SIKB, 2007).

The Office Investigation (*Bureauonderzoek*) seeks to evaluate the archaeological expectation in the specified area, supported by information from existing sources. It results in a report describing the specified archaeological expectation and giving a suggestion to the authorities about the decision on whether or not to have a follow-up investigation is needed to be carried out. (SIKB – Protocol 4002, 2018)

The Field Survey (*Inventariserend Veldonderzoek* - IVO) seeks to supplement and test the archaeological expectation obtained from the Office Investigation report. It is conducted by performing observations in the field to record the presence or absence, the nature, the extent, the date, the integrity, the conservation and the quality of the archaeological site. It can be performed by trial trenches and test pits (IVO-P) or by mapping, drilling or geophysical research (IVO-O). Its final report will contain the description of the procedures taken as well as advice through which the authorities will take their policy decision (SIKB – Protocol 4003, 2018).

Based on the results obtained from the previous reports it will be decided on what action will be taken. The specified area can be decided to be: Preserved in situ, which means the archaeological monument is recognized to be of such importance for the local history that will be preserved in its present location; Released, which means the process is finished and the construction, renovation or demolition works can proceed; or Excavation, that means the area can contain relevant archaeological vestiges and must be systematically excavated before the construction works may proceed and disturb the soil. Only the last procedure is covered in this work, as it is the only one that can produce more archaeological findings.

The Excavation (*Opgraven*) seeks to process and secure the archaeological material from sites, and record their information, which may contribute to the knowledge about the past. It results in a report submitted to the authorities and notified to Archis (Archaeological Information System for the Netherlands), and also deal with the destination of the material recovered and documents produced (SIKB – Protocol 4004, 2018).

These archaeological investigation phases are organized in what is called Archaeological Monument Care Cycle (*Archeologische Monumentenzorg Cyclus – AMZ-cyclus*). It is structured as a multi-stepped process by the Cultural Heritage Agency (*Rijksdienst voor het Cultureel Erfgoed*). It covers all procedures from the initial assessment to the final protection and conservation of the site and its recovered material (Fig.1A). The work of commercial archaeology companies through the archaeological investigation steps originate the reports (Fig.1B) used in this study. They are described in a more detailed manner in the KNA protocols. (SIKB – BRL 4000, 2018; Cultural Heritage Agency, 2022).

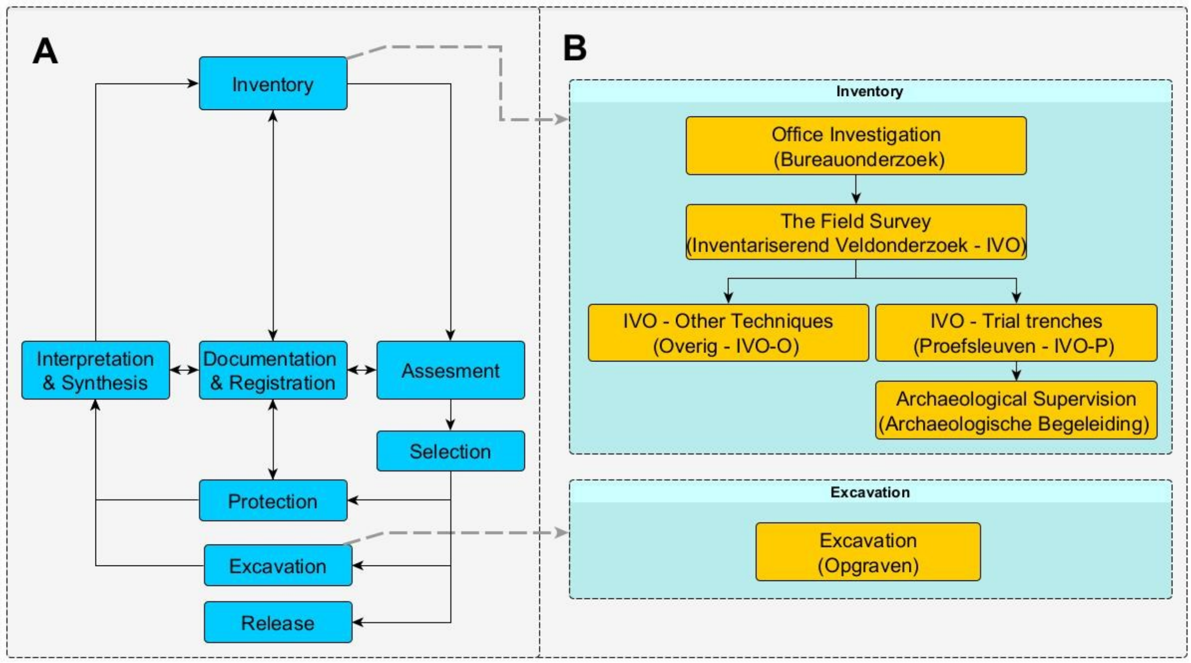


*Figure 1 — **A** - AMZ-Cyclus schema; **B** - Detailed view of steps that generate the reports used in the study – Source: https://www.cultureelerfgoed.nl/*

The reports produced during the inventory phase also include recommendations for the authorities on what should be the specific site follow-up actions. The assessment and selection of what will be done are up to the local authorities. If the site is considered irreplaceable, it cannot be destroyed, so the planned construction works have to be changed to ensure its protection. If the site is considered relevant, further excavation is demanded to extract as much information as possible from it, and it also generates an extra report about the findings. Otherwise, if the inventory step is considered sufficient according to the evidence found, the site is released for the continuation of the construction works (SIKB – BRL 4000, 2018; Cultural Heritage Agency, 2022).

The commercial archaeological reports and datasets are submitted to the municipality where the site is located and the digital copies are deposited into the Data Archiving and Network Services (DANS). DANS is the Dutch national centre and repository for research data and it currently makes them available through an online archiving system for depositing and recovering data. It contains the collections of the former Netherlands Historical Data Archive (NHDA) and the Steinmetz Archives and accommodates the "E-depot for Dutch Archaeology" that contains the datasets that are the focus of this research (E-Depot for Dutch Archaeology, 2021).

Projects presented individually are usually small in scale, as they are required to follow the path of the works, which can be as narrow as a street or limited to a single building. These constraints explain the mosaic nature of the resulting datasets. They serve their legal purpose of evaluating the archaeological value of the area, but they usually do not carry enough information for a full-fledged analysis of the site. For better use of the information contained in them, they need to be integrated into larger sets that allow expanding their scope and contribute to the broader analysis. This requirement only can be fulfilled if the data is findable, accessible and interoperable, which can allow it to be reused for the necessary integration.

3.3. FAIR data

The FAIR principles intend to establish an infrastructure that encourages the reuse of data, initially, the main effort was directed to create and disseminate proper metadata that would support it, as they are a crucial factor for its dissemination and it can be seen also in the description of the principles, as metadata are planned to survive even to the actual data disappearance (Wilkinson et al. 2016).

Since then, the principles have been applied and adapted to several organizations. Initiatives such as Go-Fair (Go-Fair, n.d.) have expanded their description and coverage and combined them with new technologies to reach even more data and organizations. For cultural heritage, and archaeology per consequence, a similar initiative called Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (PARTHENOS) also developed a set of guidelines (Table 1) for helping heritage organizations to make their data compliant with FAIR principles.

FAIR Principles	PARTHENOS Guidelines
	1. Invest in people and infrastructure;
Findable	2. Use persistent identifiers;
	3. Cite research data;
	4. Use persistent author identifiers;
	5. Choose an appropriate metadata schema;

Accessible	6. Choose a trustworthy repository;
	7. Clearly state accessibility;
	8. Use a data embargo when needed;
	9. Use standardised exchange protocols;
Interoperable	10. Establish well-documented machine-actionable APIs;
	11. Use open well-defined vocabularies;
	12. Document metadata models;
	13. Prescribe and use interoperable data standards;
	14. Establish processes to enhance data quality;
	15. Prescribe and use future-proof file formats;
Reusable	16. Document data systematically;
	17. Follow naming conventions;
	18. Use common file formats;
	19. Maintain data integrity;
	20. License for reuse;

Table 1 - Guidelines compiled from PARTHENOS et al. (2018).

As it is closely related to the subject in this study, and also has members of DANS taking part in this specific initiative, the PARTHENOS guidelines present a good match for evaluating characteristics that enhance the reusability of the data produced.

The current study focuses on usability, so the last five guidelines are the main topic for consideration and comparison, but as its implementation passes through many of the other guidelines, they are also touched in one way or another.

The PARTHENOS guidelines (PG) under the Reusable principle start with PG-16 "Document data systematically", which seeks to clarify the content of a dataset or repository, so the user will know what can be found, or not, in such dataset or repository. A common way to do it is with metadata, putting as much information as possible, but in a vast array of interests and subjects, it can be tricky to find descriptions that satisfy all the users, present and future, so it is always a work in progress (PARTHENOS et al., 2018).

That also leads to the next guideline, PG-17 "Follow naming conventions", which recommends electing a "precise and consistent naming convention" (PARTHENOS et al., 2018, 10). It is stated to be followed for data file names, but it can also be applied for the dataset content as a whole. When the data content follows a standardized nomenclature, it enhances the possibility of reuse as it is more likely to be integrated with other databases.

PG-18 "Use common file formats" advises the use of common and standard file formats currently in use by the scientific community in the research discipline chosen. In its notes, it refers to guideline 19 (Prescribe and use future-proof file formats) to give preference to open-source or at least open-specification formats over more closed proprietary ones that can become black-boxes after their use becomes less popular due to technological evolution (PARTHENOS et al., 2018).

Focusing more on the data content itself, PG-19 "Maintain data integrity" prescribes that the changes in the datasets by posterior revisions should be traceable to assure that the research data that was once retrieved can still be correlated to the current version accessed (PARTHENOS et al., 2018).

And as a closure guideline, PG-20 "License for reuse" seeks to remind the data producers to allow the data to be reused through the assignment of compatible intellectual property licensing scheme (PARTHENOS et al., 2018), which, if forgotten, can create insecurity for future researchers interested in the study of the datasets produced.

Even though the focus seems to be connected mainly to the reuse of the datasets, finding them is a crucial step in this process. The smoothness of the journey to the datasets can amplify their odds of reuse, and testing the use of knowledge graphs as a way to explore the data can open new ways to do it.

3.4. Use of knowledge graphs

Researchers have been exploring the possibilities of knowledge graphs for a long time with different approaches and goals, like Bakker (1987), Stokman and de Vries (1988), van de Riet and Meerman (1992) and Zhang (2002). In 2012, Google presented its project on knowledge graphs as a tool to improve its search results, using the slogan "things not strings" and attracted a wider interest in knowledge graphs research. Bergman (2019) listed 27 different definitions of knowledge graphs collected from academic studies since 1974, of which 19 were published after the Google project presentation.

The interest derives from the ability of knowledge graphs to use conceptual meanings to solve queries and identify a broader array of answers to users' questions. This is accomplished by expanding machine capabilities of understanding text sentences and semi-structured tables using its node/edge architecture. It enables applications to go beyond simple keyword searches to seek and retrieve information (Yan, J. et al, 2018; Kejriwal, 2019; Ronzhin et al., 2019)

Many studies give structural descriptions and/or applications of knowledge graphs, but much fewer propose a synthetic definition. Zhang (2002: 19) describes knowledge graphs as a "method of knowledge representation" that "belongs to the category of semantic networks". The method's representation is composed of concepts and relationships, where the concepts are represented by nodes in the graph and their relationships by the edges connecting them. (Zhang, 2002). The meaning is expressed by the graph structure, so the statements need to be unambiguous and reduce the number of relation types to a limited set. Ehrlinger and Wöß (2016, 18), for instance, propose that a "knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge". Both present an interesting framework to be taken into account during the evaluation of the possible contribution of knowledge graphs to the usability of archaeological datasets.

However, the integration of such datasets incurs the risk of similar terms being used in different contexts with different meanings. Brandsen et al (2021) found out the two more common negative feedback received on their searching mechanism for papers in the Dutch archaeological domain were "concepts" and "facets". The ten professionals they interviewed mentioned them six and five times, respectively. In this context, concepts were the named entities (artefacts or structures) being sought and facets were the filtering terms for specific metadata values (archaeological periods or document types).

Their results highlight a more pervasive issue of needing to find the correct term to include in the search that matches the limited amount of keywords chosen to characterize the paper or dataset when it was stored. One of the examples they presented is for finding "all beakers from graves in the late Neolithic" (Brandsen et al, 2021:9/21). To accomplish it the query would necessarily include the terms "beaker"(defined as a small drinking vessel without a handle or spout that can be made of glass or metal), "late Neolithic" and "grave". Brandsen included also "prehistory" and "Neolithic" to further filter the results, showing how difficult it can be to pinpoint the periodization, even on a national level. The search could be enhanced even more by including the terms "burial" and "tomb" to increase its reach. The use of keywords become an issue in finding the desired content or being flooded by undesirable one. For instance, if in the same search, they tried to avoid the periodization uncertainties by omitting it, the results would probably be flooded with entries from the Bell Beaker culture (archaeological culture found in the European Bronze Age)

The use of the same word "beaker" in two completely different settings is just one of many examples of how hard can be to integrate heterogeneous data sources and avoid such semantic conflicts, Gagnon (2007) proposes an ontological approach as a solution to circumvent the limitations in interoperability. For Guarino (1998:4) an ontology is constituted by a "specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words". And Uzdanaviciute & Butleris, (2011:36) reminds us that in this specific context it "refers to a machine-readable representation of knowledge, particularly for automated inference".

The importance of having an ontology presenting a set of well-defined concepts to structure the knowledge in a way it can be found and understood not only by its readers but also by the search engine used to retrieve the results. The question then becomes which ontology is more suitable for use. Due to the limited scope of the data in the datasets used in this study, the complexity falls mainly into choosing a chronology that can match the classification already applied in the datasets.

DANS suggested the Getty Vocabularies (https://www.getty.edu/research/tools/vocabularies/aat/index.html) as a "neutral" ontology agreed by European heritage institutions to be used in the exchange of semantic information between countries. According to them, it is considered "neutral" because it comes from the Getty Museum in the US, which would not favour any European-based institution ontologies (personal communication, October 18th, 2021). This vocabulary is already in use in the ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking) Portal (https://portal.ariadne-infrastructure.eu/). The portal is the geospatial branch of the ARIADNEplus research project that aims to put together in a unified database the European archaeological repositories.

Meroño-Peñuela et al (2015) presented an overview of semantic technologies applied to historical datasets and managed to list 48 different ontologies and vocabularies. In the next year, Calvanese et al (2016) in their work propose a framework for improving the scholars' access to data about the Roman Empire's food production and trading system already added ten more vocabularies suitable for being used. This highlights the difficulties of making an exhaustive search on this subject.

Two other options were chosen to be explored as options in the heritage domain, the IDAI Chronontology (https://chronontology.dainst.org/ and the CIDOC Conceptual Reference Model (https://www.cidoc-crm.org/)

The IDAI Chronontology was chosen because it presented an interesting concept that maybe could be useful. Their data model assumes periods as "spacetime volumes (STV), i.e. they have a geographical extent in addition to their temporal extent". (IDAI - online). The depth of the concept is above the ambition and the needs of the current study, as it is limited to the Enschede municipality. However, compatibility with it could contribute to future studies, as even inside the borders of the Netherlands there are variations between archaeological periods and cultures between different regions as can be seen in Fig 2. For instance, it can be seen how the Early Neolithic starts slightly earlier in the south of the country. A feature that allows binding geospatial boundaries to the periods could ensure answers to a wider array of searches and fit them to the local periodization and type of the material being sought.

On the other hand, CIDOC was chosen as it has been used in different cultural heritage studies for mapping information to a digital equivalent representation (Binding et al., 2008, 2015; Eide et al., 2008; Hedges et al., 2017; Sinclair et al., 2006; Varnienė-Janssen & Kuprienė, 2021). According to its conceptual reference model it aims "to enable information exchange and integration between heterogeneous sources of cultural heritage information" (Doerr et al 2020: I) and to meet the needs of "all types of material collected and displayed by museums and related institutions" (Doerr et al 2020: II).

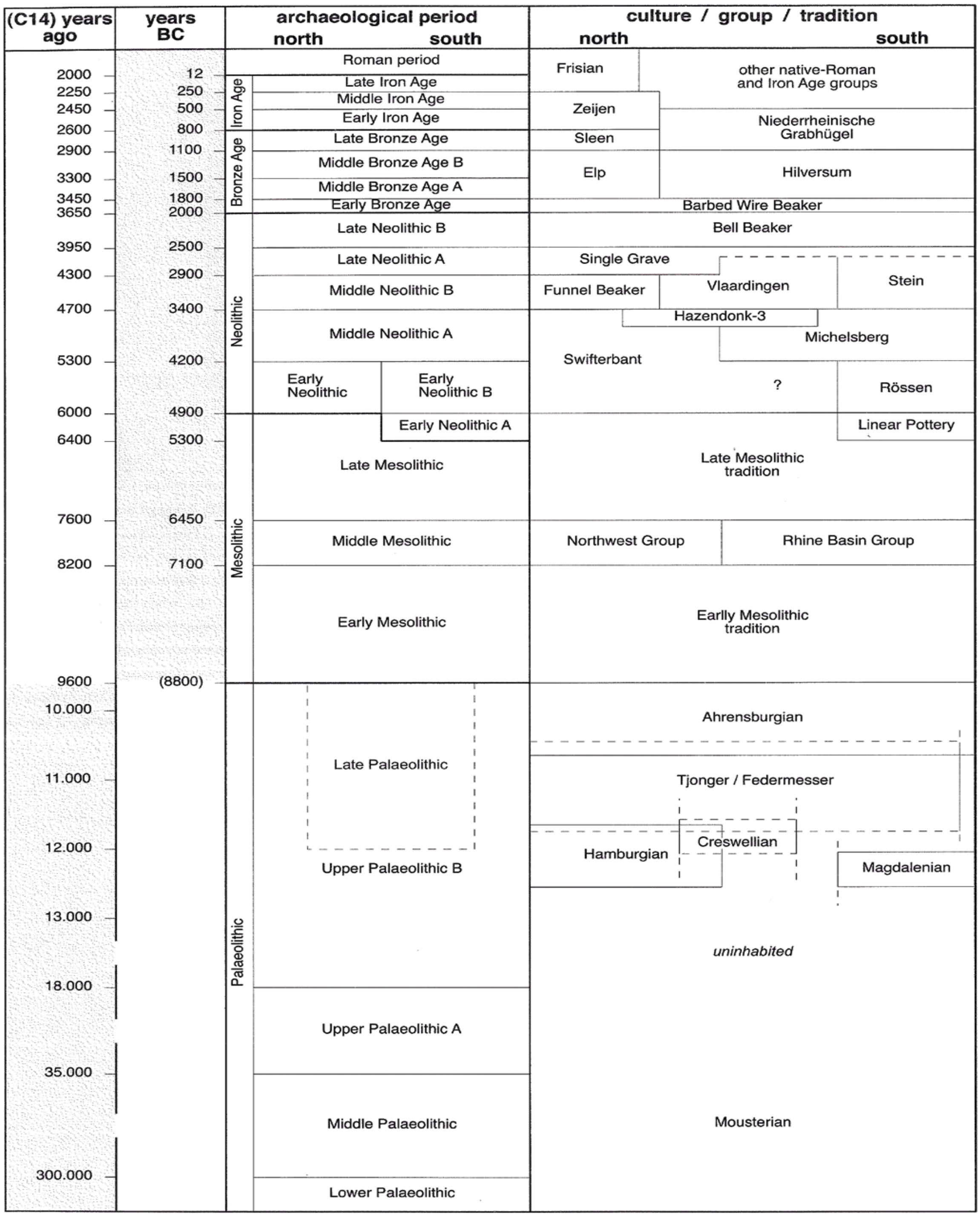


Figure 2 - Chronological and cultural variations between the northern and southern parts of the Netherlands Source: Lauwerier et al, 2017:26

4. MATERIALS AND METHODS

4.1. Study area

The area chosen to be studied is the municipality of Enschede, as it has enough interventions to expect a sufficient number of findings, but on the other hand not so many that could become overwhelming.

Deposited in the DANS database, were found 224 datasets, ranging from 2002 to 2021. The graph distributing them by year (Fig.3) shows that most of the reports stored were written after the establishment of the DANS, in 2005.

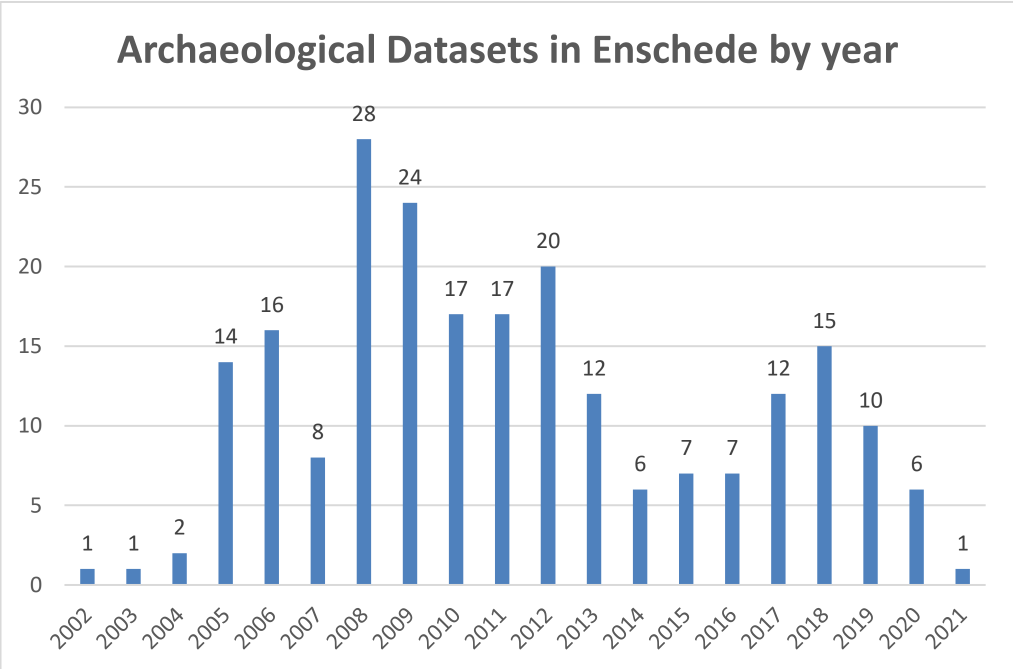


Figure 3 - Archaeological Datasets found in DANS in Enschede by Year Data Source: E-Depot on Dutch Archaeology, 2021.

4.2. Overall methodology

The main objective is to improve the usability of these datasets, in other words, to make available in one place as much information from them to future researchers so that they will not lose unnecessary time looking into reports that they do not need.

As a way to exemplify it, we will assume a hypothetical case of an archaeologist that desires to study palaeolithic sites in the Netherlands and will try to integrate commercial archaeology data to evaluate the presence of palaeolithic remains within the Enschede municipality.

In the current situation, there are not many options to look for palaeolithic sites in Enschede directly, neither in DANS nor in Ariadne. As a way to carry out the task, the work was divided into six steps: (1) Source Data Assessment; (2) Data Processing; (3) Standardizing Findings Data Description; (4) User Assessment; (5) Geoportal Functional Requirements and (6) Ontology and Knowledge Graphs.

In the first step (1), the source data will be assessed to evaluate what is their actual state and how much it complies with the FAIR principles proposal. Each one of the projects will need to be accessed individually to gather the data available and to evaluate how the data can contribute to achieving the final goal. The understanding of how the data is organized inside the report projects on the DANS website is a mandatory preliminary step for our research on the palaeolithic sites, as it can help to sort which reports are relevant for our study; In the next step (2), the data will be extracted from the reports and the goal is to create an integrated artefacts table from the data extracted from the datasets with as much information as possible to be made available later. The format chosen for the integrated table file is the CSV format. It is a semi-structured lightweight format that can be read by the majority of the GIS packages and is easily convertible to other formats if needed which complies very well with PG-18;

To create this table is needed to extract the individual tables containing the archaeological artefacts recovered, their classification and location from the reports/datasets available in the repository. In many cases, this spatial data is not included directly in the table, but it can be found in the report on maps or other images included in it.

Based on these extracted tables, the information contained is compared to evaluate what data is common to all reports, or what kind of information can be retrieved from most of them in a way that it can be used for a posterior analysis or treatment.

In the third step (3), the findings table will be reviewed as it is already expected that, regardless of the extraction and structuring, it can still be difficult to standardize the classification of the obtained records. The integration of so many different sources through an extended period can encompass several different approaches and methodology shifts that can make it hard to find a common denominator.

Due to the limited time and scope of this study, the fields with more information will be reviewed to evaluate how much it will be possible to standardize them and if it is worth the effort.

The fourth step (4) will consist of a user assessment, as the theoretical palaeolithic study is an example to show how research can be carried out in these datasets, it is a practical problem that posed a challenge for other professionals too. Collecting impressions and strategies used in their work can help to give ideas of procedures and also to evaluate what is the type of information that can more relevant for future users. This is intended to be achieved by interviewing professional archaeologists working in the Netherlands that could aggregate these data into their works.

The fifth step (5) is to present the structure and the functional requirements for the proof-of-concept geoportal are presented based on the data obtained by the data processing and standardization and following the observations that were the result of the interview.

The final step (6) is dedicated to exploring how knowledge graphs can be constructed from the dataset that was generated, discussing how the definition of an ontology and formatting of the information obtained could benefit the palaeolithic study and also future researchers in finding the information they need.

4.3. Step 1: Source data assessment

The research aims to propose a way to improve the usability of the Commercial Archaeology spatial datasets already available in the repository. An assessment of the current state of the data is needed to evaluate what are the steps needed to integrate and make them available.

The direct access through the DANS website (https://dans.knaw.nl/en/) does not allow a spatial search, only textual (Fig. 4). The alternative, suggested by them, is to access their database using the ARIADNE Portal which allows searches made by location. However, the number of results shown on the screen (Fig. 5) is not linked to the actual number of finds. It is displayed considering a bounding box created by the visualization on the screen, so it sums up all reports inside it. As it does not give the boundaries of Enschede municipality, it can mistakenly exclude reports if the wrong zoom level is used when the button "Display as a research result", located in the top left, is clicked. It seems not to be a pure spatial search as it includes also reports that mention the place Enschede, not only its spatial position.

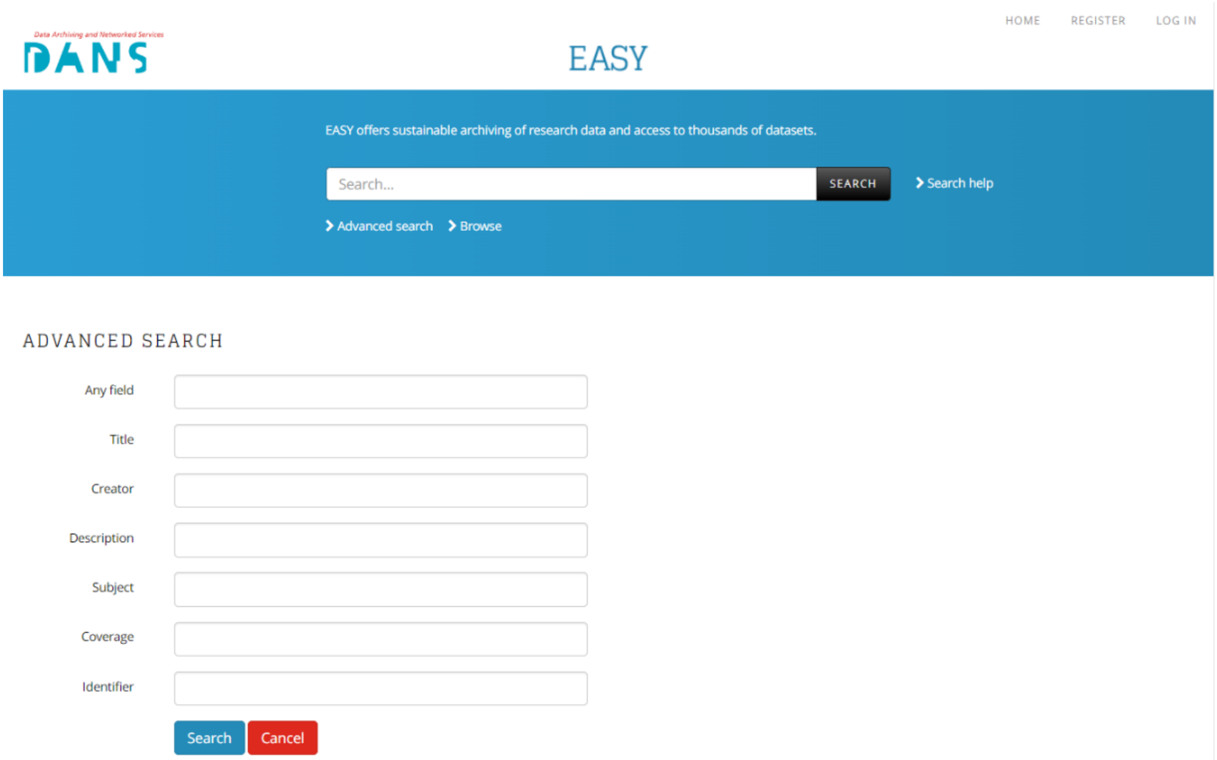


Figure 4 - Search screen on the DANS website

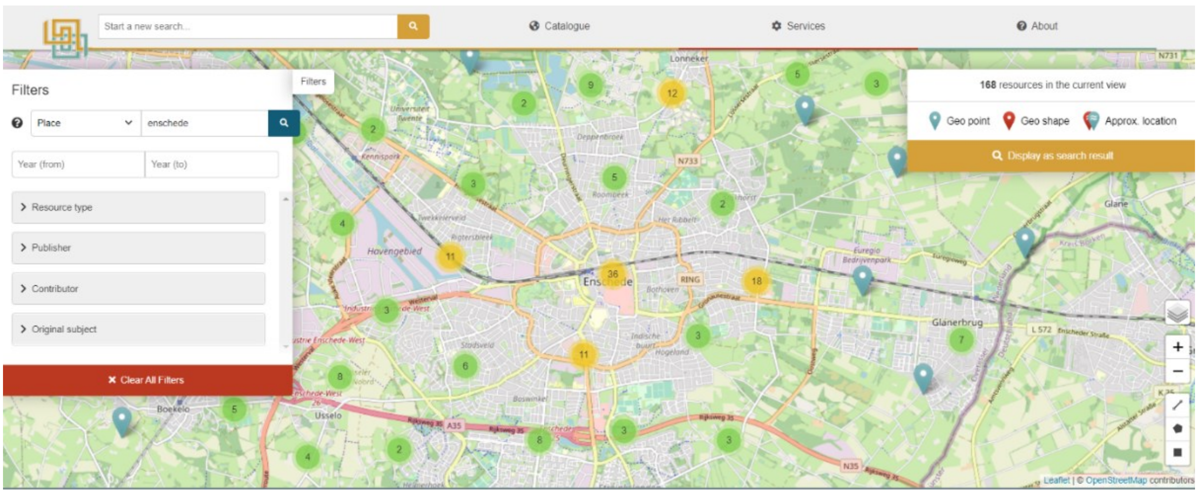


Figure 5 - Ariadne Portal spatial search

When opened in ARIADNE, each report shows the metadata related to it and there the persistent identifiers, according to the FAIR principles, can be found for each report. The identifier redirects the user to the report landing page at the DANS server (Fig. 6). There, three tabs are available, an Overview tab with the citation details for its use and a brief description provided by the company that deposited the data. A description tab presents the metadata available for the report, including the coordinates for the location of the point used in Ariadne. And a data tab gives access to the data itself, most of the time it contains only the report itself for download in the Portable Document File (PDF) format, as shown in Fig. 6.

Evaluating the DANS handling of these datasets, it is noticeable they follow the PARTHENOS guidelines as much as possible. However, as they are not the producers of the data, so they only are capable to ensure the application of the principles as far as the report landing page.

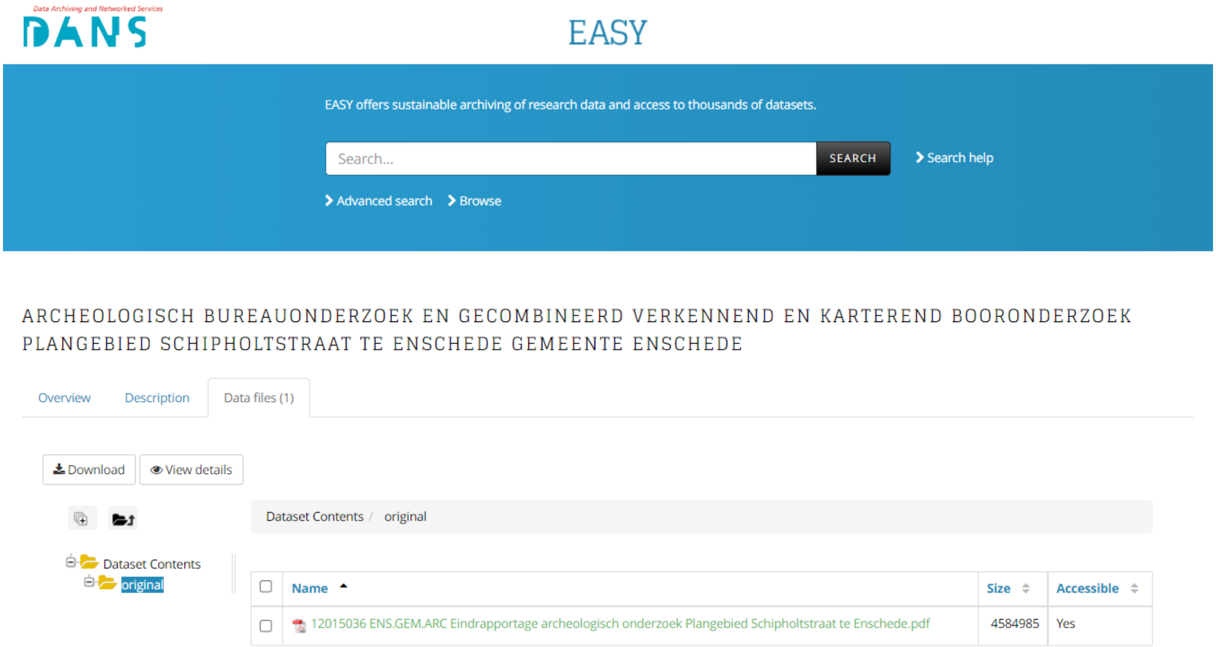


Figure 6 - Example of the data files tab on the report landing page at DANS.

The metadata has a comprehensive array of elements that allow its classification, including a description field that could help to fill whatever gaps could still exist. Despite this, the data filled in the field end up not contributing as much as they could due to its not standardised content. The text is provided by the companies, and usually gives a brief introductory view to the report, but it seldom gives any hint on the results and when does, it usually talks about the legal recommendations prescribed for the area excavated but not about what were the archaeological findings in the site or periods identified. In some cases, the field has just the name of the report contained in the accompanying PDF.

If considering the usability PARTHENOS guidelines individually for the content: (PG-16), when there is more than just the report, the relationships between the files included are not well documented; (PG-17) the file name conventions are kept inside the same company, but they are not applicable for the others; (PG-18), the file formats used other than the reports, usually follow the guidelines, but there are exceptions; (PG-20) the licensing applied was almost entirely appropriate for the reuse the data with some cases asking for creating a user on the DANS website, and very few asking for individual company permission to use the data.

Even though this first evaluation checks most of the guidelines boxes, none of them solves the question of how a researcher interested in palaeolithic sites in the Enschede area could find the commercial archaeology datasets on the DANS public repository without opening them one by one.

4.4. Step 2: Data processing

The first steps deal with assessing the current state of the datasets, and verifying how different institutions stored them (file types, classes and attributes used). The data, as they are deposited, can vary in the ways of recording the information, even though the reports share a similar structure due to legal requirements. They depend on the company or institution's common practices and procedures for their archaeological interventions.

Twenty-four different companies worked in Enschede according to the datasets recovered (Fig.7). Five of them are responsible for a little more than 50% of the investigations. It is not surprising as the timespan being considered is almost twenty years. However, it adds an extra layer of complexity to the changes in practices and technologies throughout the period as it also needs to equalize different approaches, nomenclatures and sets of specifications.

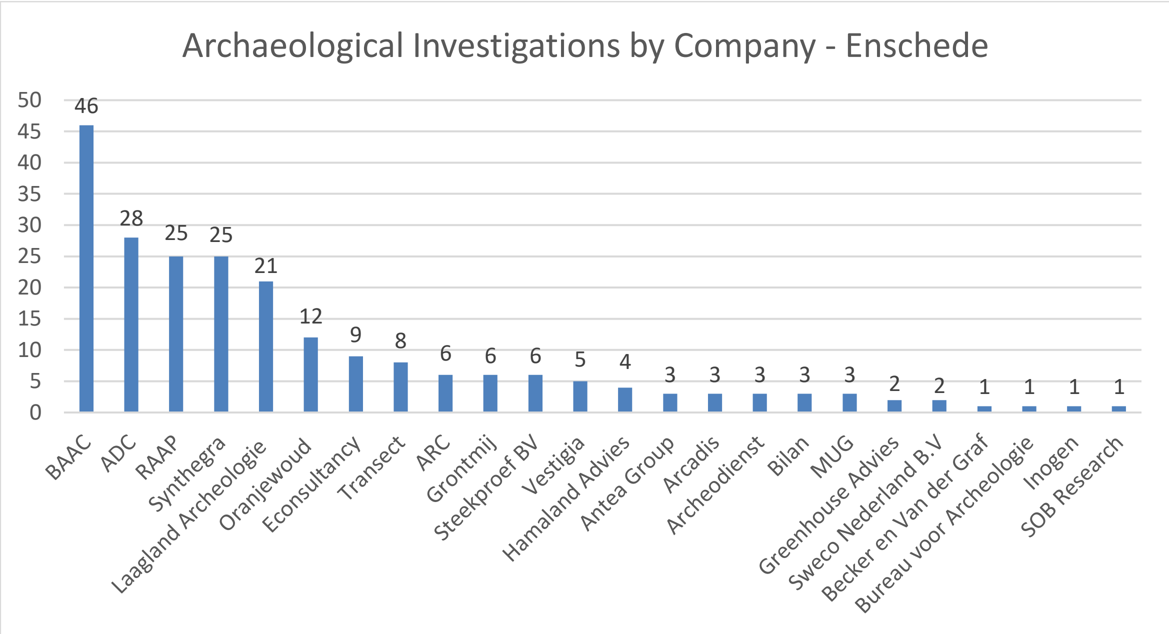


Figure 7 - Archaeology Investigations by Company Data Source: E-Depot on Dutch Archaeology, 2021

The datasets also contain different kinds of information, due to their position in different stages of the AMZ cycle (Fig. 8). Office Investigation reports usually do not have any new archaeological finding information because they do not demand any visit to the actual site to be written, and other types of the report have different objectives and procedures. There are also types of reports not explicitly provided for in the protocols, as they have to adapt to the specific environment of the area where the work is being carried out. For instance, the Archaeological Supervision reports are made in cases when it is not feasible to carry out a trial trench investigation before the implementation works, therefore a preliminary investigation cannot be carried out. As a solution, the earthworks necessary are done under archaeological supervision to record any possible archaeological material to be found. A large number of reports combine more than one type, like the Office Investigation succeeded by a Field Survey that is the most common combination. To those, it can also be added the 27 instances combining both and a third investigation type (IVO-O) that can include mapping, drilling or geophysical techniques.

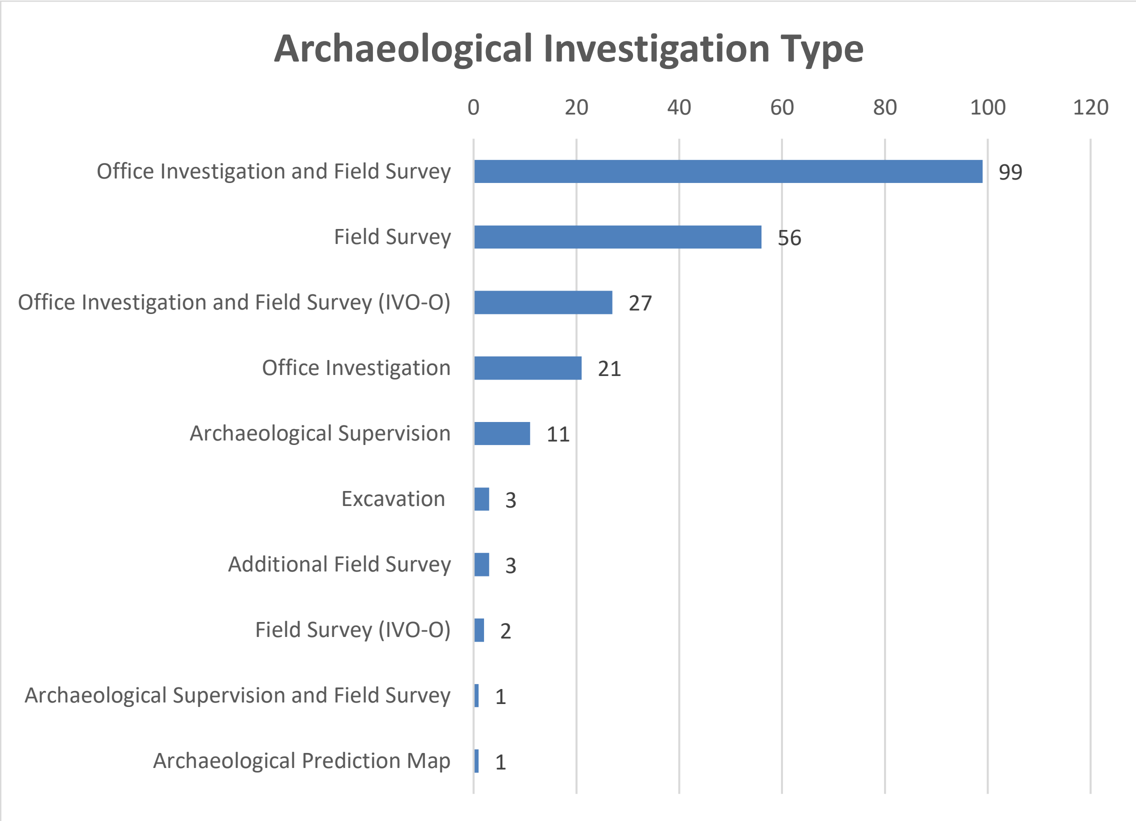


Figure 8 - Archaeological Investigation Type Data Source: E-Depot on Dutch Archaeology, 2021

The majority of the datasets were found in the surveying step of the AMZ-cycle, with a very small amount containing actual excavation or other instances. From those reports, 42 (19%) recorded archaeological finds. The recorded finds distribution (Fig. 9), follows roughly the reports' distribution, but their relationship depends on many other variables and it is not a 1:1 correlation.

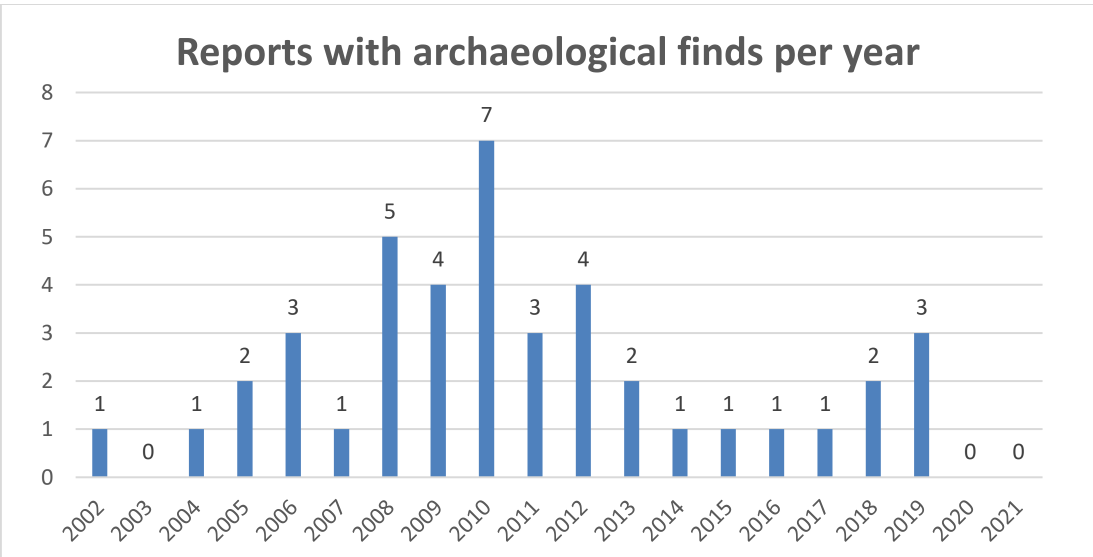


Figure 9 - Report with Archaeological finds in Enschede per year Data Source: E-Depot on Dutch Archaeology, 2021

The selected 42 datasets include all data available about the intervention recorded in different formats (Fig.10). Except for one, all include at least one PDF file containing the written report about the investigation. Others also include Comma Separated Value (CSV) files with the tables containing data about the area and the findings recorded. A few of them include files containing the actual geospatial information about the investigation results, notably the more recent ones (Fig.11), in different formats such as Shapefiles (SHP), Quantum GIS Projects (QPJ), MapInfo Interchange File (MIF) or Geography Markup Language (GML).

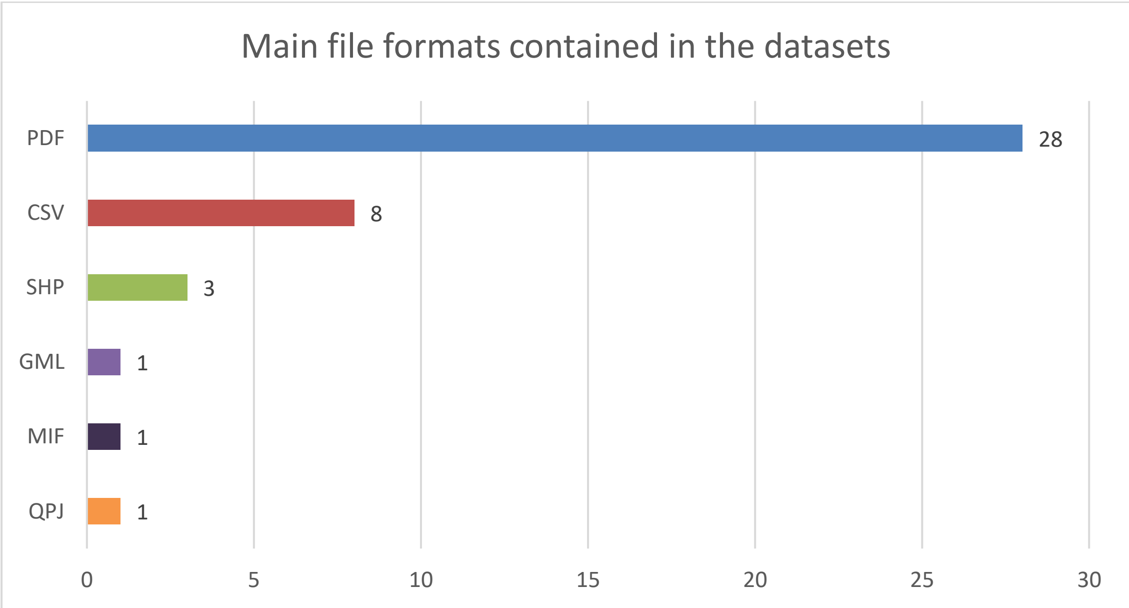


Figure 10 – Main file formats contained in the datasets Data Source: E-Depot on Dutch Archaeology, 2021

The dataset containing the MIF file was the only one that did not contain a written report, which made the whole interpretation rely solely on the geospatial data available. However, it contained enough metadata included that allowed us to make sense of the data.

In more than 65% of instances, it is recorded in a single file that usually contains the full report and all tables associated with it. This procedure is very well suited for archiving, as it keeps all data together and easily accessible as text. However, it does not allow a search of the internal content, except by the use of keywords previously chosen.

The metadata available includes the geographic coordinates of a single point for the whole intervention (project). These coordinates allow the geo-portals to pinpoint the dataset in a specific place, but it is not possible to distribute the finds spatially or into the possible archaeological classifications.

The compliance of the data licensing for reuse was majoritarian (93%), and the three remaining datasets, belonging to two different companies, that asked for specific approval have been obtained through contacting the companies by e-mail asking for permission to use the data in DANS.

For the exemplary research on palaeolithic sites, the processing of the files could be done only to those that include findings from this period. However, for the proof-of-concept geoportal, the full array of findings will be included. In both cases, the same types of files will need to be dealt with according to their characteristics to extract the information and proceed with the work.

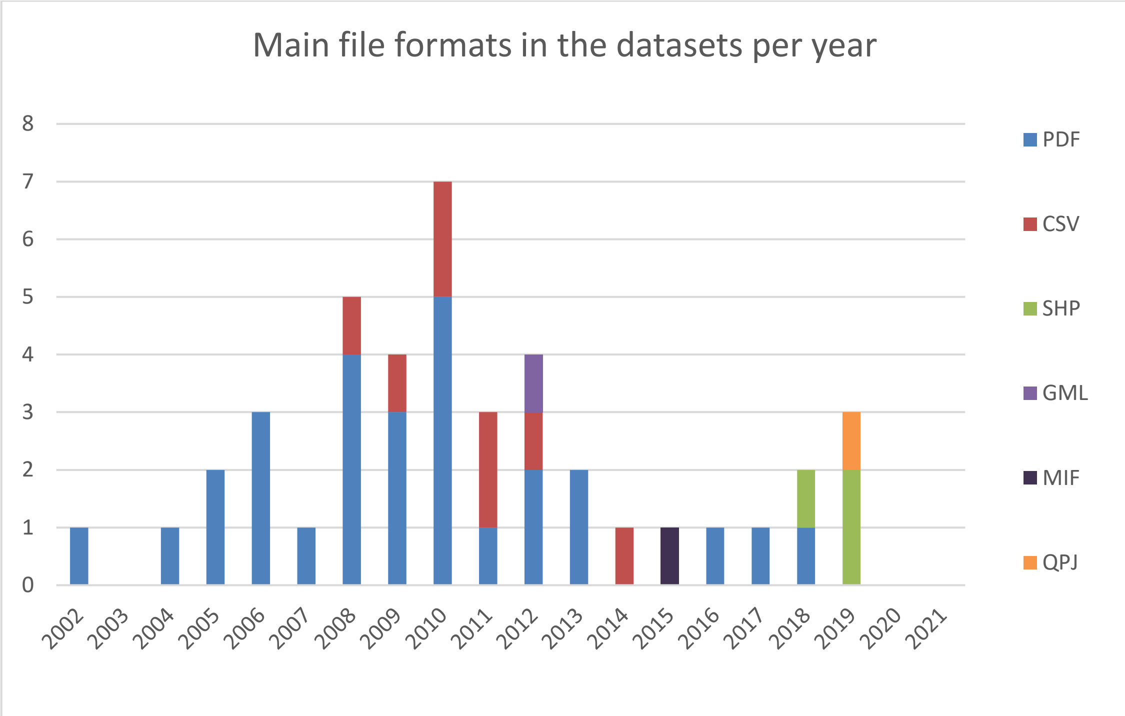


Figure 11 – Main file formats used per year Data Source: E-Depot on Dutch Archaeology, 2021

4.4.1. Text-based files (PDF and CSV)

Initially, all 36 reports were assessed to have an overview of the complete database available. The data was extracted from the (28) PDF reports using Tableau Desktop. Due to the internal processing used by the software, these tables were created individually page by page. When the original table comprised more than one page they needed to be also manually merged. An automatic solution did not have good results, as on different pages there were different problems with the column order, or sometimes with different columns being wrongly merged. In some cases, depending on the table styles chosen by the report-makers some information was lost and needed to be inserted manually into the CSV resulting file, like the column titles. As a solution, the resulting files were opened in the Microsoft Excel file for easier integration and visualization of the final layout.

Some reports also demanded the extraction of other tables, not only the one with the findings. The geospatial information wasn't associated with the findings, but with the trial areas investigated. Those areas contained the trenches or pits dug in the site. The map sequences are divided by those areas, and inside them, the trenches and pits cannot be found directly.

None of the tables (either PDF or CSV) had any coordinates recorded, and the majority of the reports did not contain maps showing the individual findings. This demanded an extra step on the positioning, as the finds need to be assigned to a central point calculated for the pit/trench/layer associated with it. The map itself was georeferenced using its coordinates when available. A better explanation can be made with an instance of one map (Fig. 12). Each of the findings (marked with a green-filled circle) is assigned to the centre of its respective work pit (WP - *werkenput* in Dutch)

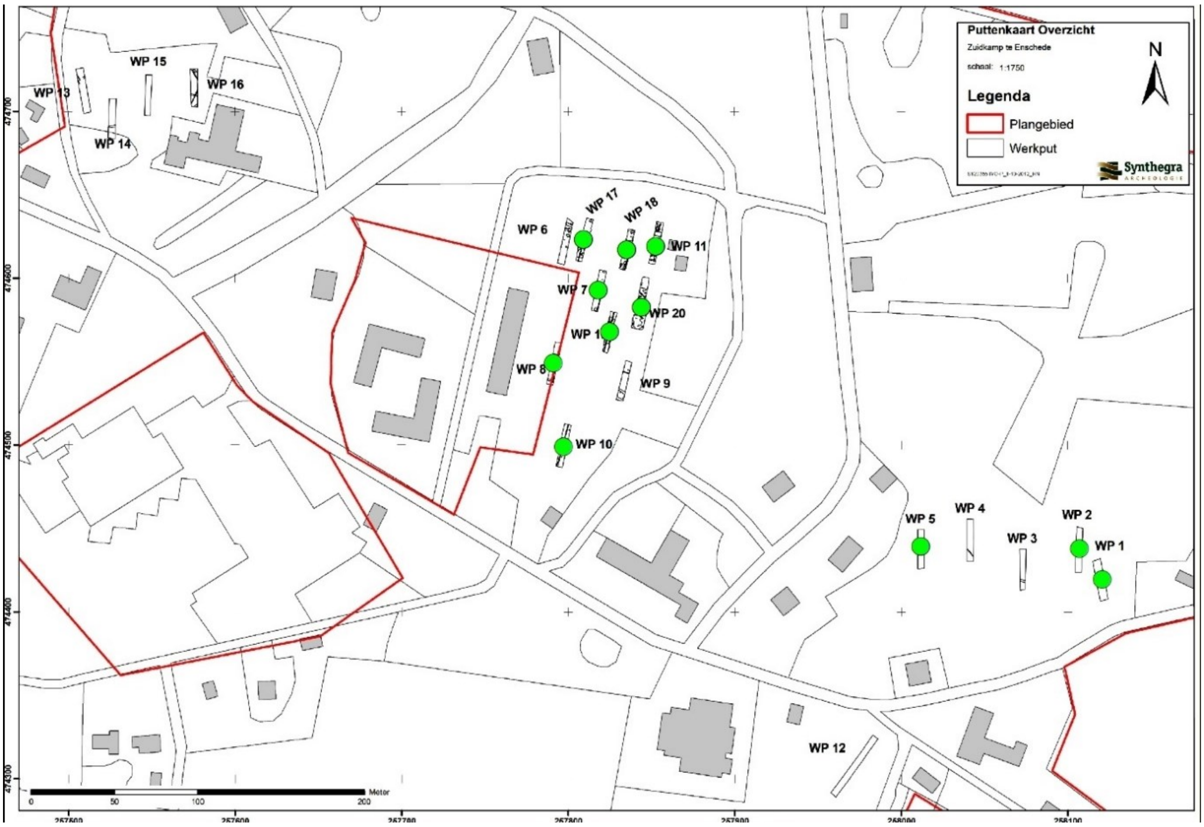


Figure 12 - Instance of map extracted from the reports with its findings associated.

In some cases, the final trench/grid schema is done only on computer-aided design (CAD) software and does not contain any indication of geographic coordinates. Luckily, the reports also contained other geographic indications that allowed the positioning of the finds through a daisy-chain process of georeferencing these CAD drawings into the maps that showed the site location when they were indicated (Fig.13).

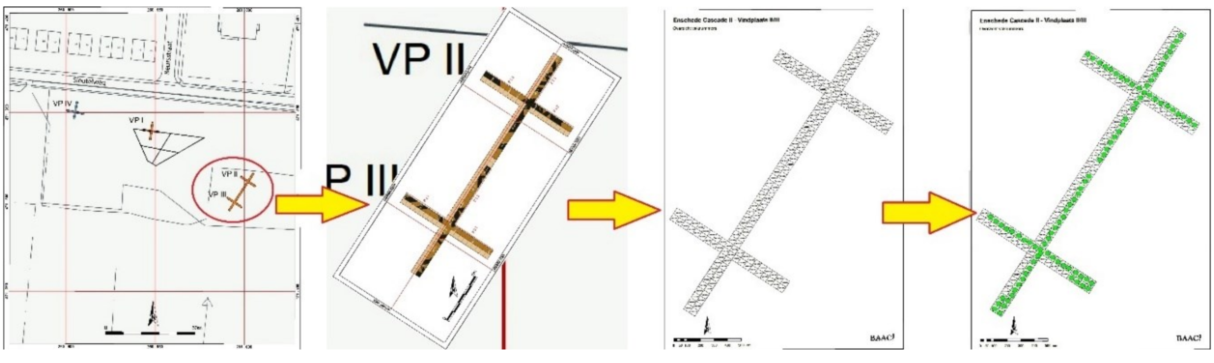


Figure 13 - Instance of daisy-chain process needed to place the findings associated with a CAD drawing.

After the process, the total of PDF reports with finds to be analysed was reduced to 18. Of the discarded reports, one contained no findings within its excavated area. However, the report presented a table of observations contained in the vicinity from non-referenced previous studies. This table was not included in the study as its findings did not mention by whom or how they were obtained. It was not possible to characterize it as a commercial archaeology work. In another case, it was found one report was repeated into two different DANS datasets (53590 and 184989). In another case, the dataset only included the word Enschede in the client address, but the site itself was located in a different municipality, Oost Gelre (131786). And, in the end, seven reports did not present enough geospatial information that allowed to locate the individual findings, which also happened in three of the eight reports that already offered the data as CSV tables, reducing their total to five.

4.4.2. GIS files (SHP, QPJ, MIF and GML)

The data is more easily extracted from the geospatial formats, even though they correspond to little more than 10% of the total number of reports. The main work on its processing was to generate their coordinates into the already existing table and export it as a CSV file. The process itself was done using QGIS in most cases as it was able to open all formats found without any extra import tool.

4.5. Step 3: Standardizing findings data description

At the end of the whole process, the findings of 32 reports were integrated into a CSV file containing the 4.235 points recorded. The artefacts recovered are classified according to many different attributes, and the evaluation of these attributes varied between the different companies throughout the timespan comprised. An excerpt taken from QGIS, shown in Fig. 10, gives an impression of the general appearance of the dataset.

The coordinate system used is the System of National Triangulation (*Rijksdriehoeksmeting* – RD). The RD is the national coordinate system for the Netherlands, and its original axes intersected in Amersfoort, in the centre of the country. The system was changed later, the axes have been shifted to guarantee that all land coordinates in the country have a positive value and the values of x are always smaller than the values for y creating the RD-New, which has been used currently (Kadaster, n.d.).

ReportData2 — Features Total: 4235, Filtered: 4235, Selected: 0

	Artefact	Find	Lat	Long	Depth(Z)	Material	SIKB	MatAlg	MatSpec	Kind	Subkind	Type	TotNr	TotWg	BeginPer	EndPer	Begindate	EndDate	Description	Opmerking	Remarks	EasyDb	Company
1	1	1	258162.9	471407.6	-	KER	KER	STG	SIEGBURG	-	-	-	1	0	MEL	MEL	1300	1400	STG	fijne hoekige ...	fine angular flut...	198193	Laagland
2	2	2	258164.6	471412.3	-	VKL	XXX	-	-	-	-	-	1	0	IND	IND	0	0	-	lbr., zacht, pla...	lbr., soft, plant i...	198193	Laagland
3	3	3	258165.9	471414.1	41.3763	BKR	KBW	DAKPAN	-	golfpan	-	-	1	0	NTA	NTC	1530	1850	roof tile	ruw baksel m...	rough baking w...	198193	Laagland
4	4	3	258165.9	471414.1	41.3763	BKR	KBW	BAKSTEEN	-	handvormsteen	-	-	1	0	MEL	REC	1300	1920	BRICK	d.rood-geel ...	d.red-yellow m...	198193	Laagland
5	5	3	258165.9	471414.1	41.3763	KER	KER	ROOD	SLIBVERS	-	-	-	2	0	NTA	REC	1600	1900	RED	1 fr ongegla...	1 fr unglazed; 1 ...	198193	Laagland
6	6	3	258165.9	471414.1	41.3763	KER	KER	INDUSWIT	-	pearlware	-	iw-kom	1	0	NTB	NTC	1775	1830	INDUSWHITE	blauwig craq...	blueish crackle ...	198193	Laagland
7	7	3	258165.9	471414.1	41.3763	KER	KER	WIT	-	-	-	-	1	0	NTB	REC	1750	1920	WHITE	pot met naar ...	pot with outwar...	198193	Laagland
8	8	3	258165.9	471414.1	41.3763	KER	KER	FAYENCE	-	-	-	f-bor	1	0	NTB	NTC	1650	-1800	FAYENCE	-	-	198193	Laagland
9	9	3	258165.9	471414.1	41.3763	KER	KER	ROOD	-	-	-	-	1	0	NTA	NTC	1600	1825	RED	grote pot/gra...	large pot/grape...	198193	Laagland
10	10	5	258167.1	471417.9	-	KER	KER	INDUSWIT	-	paerlware?	-	-	1	0	NTC	NTC	1800	1830	INDUSWHITE	onderglazuur...	underglaze tran...	198193	Laagland
11	11	5	258167.1	471417.9	-	KER	KER	STGL	-	Münsterland	-	-	1	0	NTB	REC	1730	1860	STGL	konische "afr...	conical 'cream' ...	198193	Laagland
12	12	5	258167.1	471417.9	-	MXX	MXX	SLAK	-	ijzrslak	-	-	1	66	IND	IND	0	0	SNAIL	licht magneti...	slightly magneti...	198193	Laagland
13	13	6	258179.9	471443.9	-	BKR	KBW	DAKPAN	-	machinaal	-	-	1	0	REC	REC	1880	2000	ROOF PAN	grijze platte ...	gray flat pan wi...	198193	Laagland
14	14	10	258184.5	471450.7	-	BKR	KBW	BAKSTEEN	-	machinaal	-	-	1	0	REC	REC	1900	-2020	BRICK	-	-	198193	Laagland
15	15	12	258261.5	471580.5	-	BKR	KBW	BAKSTEEN	-	handvormsteen	-	-	1	0	REC	REC	1850	1950	BRICK	late? handvor...	late? hand mou...	198193	Laagland
16	16	13	258264.5	471583.6	-	KER	KER	KGP	-	-	-	-	1	0	MVR	MVR	900	1350	KGP	mogelijk vari...	possible variant...	198193	Laagland
17	17	13	258264.5	471583.6	-	MXX	MXX	SLAK	-	vloeislak?	-	-	1	27	IND	IND	0	0	SNAIL	magnetischbr...	magnetic brow...	198193	Laagland
18	18	14	258191.4	471460.9	-	KER	KER	ROOD	-	-	-	-	1	0	MEL	NTB	1300	-1750	RED	-	-	198193	Laagland
19	19	16	258193.5	471465.3	-	BKR	KBW	DAKPAN	-	golfpan	-	-	2	0	NTA	REC	1550	-1900	ROOF PAN	-	-	198193	Laagland
20	20	18	258199.8	471479.4	-	KER	KER	KGP	GROFMAG	-	-	-	3	0	MVR	MEL	900	-1300	KGP	-	-	198193	Laagland
21	21	19	258199.1	471477.8	-	KER	KER	KGP	GROFMAG	-	-	-	1	0	MVR	MEL	800	-1300	KGP	-	-	198193	Laagland
22	22	20	258199.4	471478.7	-	KER	KER	KGP	DEKSELGL	-	-	-	13	0	MEV	MEL	1200	1300	KGP	iets verdikte s...	slightly thickene...	198193	Laagland
23	23	20	258199.4	471478.7	-	MXX	MXX	SPIJKER	-	MFE	-	-	1	0	IND	IND	0	0	NAIL	gecorrodeerd...	corroded nail fr...	198193	Laagland
24	24	20	258199.4	471478.7	-	SXX	SXX	BROK	-	-	-	-	3	95	IND	IND	0	0	BROK	STE	STE	198193	Laagland

Figure 14 - Full data table excerpt

Due to the restriction of time and resources, it would be impractical to try to do it to all the fields present in the table, so the fields with more occurrences were selected to be revised and tested for queries in the geoportal. The list and explanation of each field of the data table can be seen in Table 2. The fields more commonly filled (above 50%) were checked for suitability to be queried. For this purpose were chosen the fields Material (100%), MatAlg (80%), BeginPer (67%), and EndPer (67%).

FIELD	TYPE	DESCRIPTION	PRESENT IN (%)
Material	String	Codified description of the main component material of the archaeological find, for instance, ceramic, iron or stone. The key for the codes can be found in the Materiaalcategorie session of SIKB, 2017.	100
BeginPer	String	First period when the artefact was possibly made	67
EndPer	String	Last period when the artefact was possibly made	67

Table 2 – Fields of the artefacts table, their description and % of occurrence

It is also notable that most of the fields filled in 100% of the instances were created especially for this study. The field Material, even though appears complete, due to approximation still had less than 0.5% of its lines blank. Nonetheless, it had to be reworked due to the lack of standardization presented in the original fields. Only 60% of the codes used could be matched with the official code in SIKB, 2017. The earliest codification seems to be from 2011, and it suffered changes in the codes used from them, some of the reports predate even the first codes, using their acronyms to describe the materials. In other cases, the code itself changed later, as in the case of plastic, until 2015 classified as KST (*Kunststof* in Dutch), but changed from that year forward to PLA. The solution was to create a new field SIKB and translate all the original codes to the current standard. From the original 64 unique instances of code originally found, it was possible to reclassify them into 30 categories, including the unknown category that comprised 0,5% of the instances. The final list can be seen in Table 3.

The field MatAlg presented a more complicated situation, it was related to the previous field, but it had been filled in different ways for the different companies. Some of its instances had one-word descriptions and others had codes from the Artefacttype session of SIKB, 2017. While the Material field had 64 unique descriptions, the MatSpec field had 222. The amount of work for a derived field did not seem reasonable compared with the worth of most of the descriptions offered as it needed to be done manually due to the interpretation involved.

In the fields BeginPer and EndPer, a similar problem was found, as there was no standard classification for the time periodization. The periodization presented in Table 4 is the result of the compilation of the most common instances found in the reports themselves, as they did not always match with SIKB tables.

SIKB – Code	Meaning (Dutch)	Meaning (English)
GLS	*Glas*	Glass
KAW	*Aardwerk*	Pottery
KBW	*Bouwmaterial*	Building Material
KER	*Keramiek*	Ceramics
KHL	*Huttenleem*	Hut Clay
MCU	*Koper*	Copper
MFE	*Ijzer*	Iron
MPB	*Lood*	Lead
MSN	*Tin, ook legeringen*	Tin, also alloys
MXX	*Metaal*	Metal

ODB	*Dierlijk bot*	Animal Bone
ODL	*Leer*	Leather
OHO	*Hoef*	Hoof
OPHK	*Houtskool*	Charcoal
OPHT	*Hout*	Wood
OPX	*Plantaardig*	Vegetable
OXB	*Bot*	Bone
OXX	*Organisch*	Organic
PLA	*Plastic*	Plastic
SBA	*Barnsteen*	Amber
SDI	*Diabass*	Diabase
SFO	*Ijzeroer*	Bog Iron
SGR	*Graniet*	Granite
SLE	*Leisteen*	Slate
STX	*Natursteen*	Natural Stone
SVU	*Vuursteen*	Flint
SXX	*Steen*	Stone
SZA	*Zandsteen/kwartsiet*	Sandstone/Quartzite
XXX	*Onbekend*	Unknown

Table 3 - Material Table compiled from the reports for classification

PERIOD		START YEAR	CODE	Description
Recent Tijd Late Modern Period	-	1850	REC	The Late Modern Period starts in the 1850s and proceeds until the present day.
Nieuwe Tijd Early Modern Period	C	1795	NTC	The Early Modern Period (ca. 1500-1850) is a period established from the end of the Middle Ages and lasted until the Late Modern Period. It is also sometimes referred to as the New Age or as the Ancien Régime.
	B	1650	NTB	
	A	1500	NTA	
Middelleeuwen Middle Ages	Laat	1250	MEL	The Middle Ages (about 500 to about 1500) are the period in the history of Europe between antiquity and the early modern period. It is traditionally situated between the fall of the Western Roman Empire in the 5th century and the Renaissance.
	Vol	1050	MEV	
	Vroeg	450	MVR	
Romeinse Tijd Roman Period	Laat	270	RTL	The Roman Period started with the Dutch territories coming into focus and the subjugation coming from 12 BCE after the defeat of the Frisii (Frisians) and other tribes that brought areas north of the Rhine under Roman control. In the 5th century Roman legions abandoned the region after being recalled to defend the empire centre during its final decline
	Midden	70 CE	RTM	
	Vroeg	15 BCE	RTV	
Ijzertijd Iron Age	Laat	250	ITL	The Iron Age in the Netherlands was a part of the Iron Age in Europe and the last period of the prehistory of the Netherlands. In the south of the Netherlands, iron objects became common around 700 BCE, while the Northern Netherlands remained behind: the iron items only became common from the 6th century BC.
	Midden	500	ITM	
	Vroeg	800	ITV	
Bronstijd Bronze Age	Laat	1,100	BTL	The Bronze Age in the Netherlands covers the period from around 1900 to around the 8th century BCE. There were neither the raw materials for bronze nor natural resources that could be used as exchange materials. Nevertheless, there was a modest bronze industry of its own, but because bronze was scarce, the flint was never completely displaced.
	Midden	1,800	BTM	
	Vroeg	2,000	BTV	
Neolithicum	Laat	2,850	NLL	The Neolithic in the Netherlands arose from an interaction

Neolithic	Midden	4,200	NLM	between the original Mesolithic population of hunter-gatherers and drawn Neolithic farmers. In the Netherlands, the hunter-gatherers lasted longer, adopting the arts of agriculture, cattle breeding and pottery-making from the Neolithic immigrants. The Mesolithic way of life gradually fell into disuse.
	Vroeg	5,300	NLV	
Mesolithicum Mesolithic	Laat	6,450	MLL	The Mesolithic is a culture period that began after the end of the last ice age. Hunting, fishing, and gathering were the livelihoods of the people, who generally lived as itinerant hunter-gatherers. Settlements were usually temporary. Finds from the Mesolithic show that stoneworking techniques became more sophisticated, with an increased occurrence of microliths.
	Midden	8,640	MLM	
	Vroeg	9,700	MLV	
Paleolithicum Palaeolithic	Jong	35,000	PLJ	The Palaeolithic is the oldest period in the prehistory of material culture. It starts about 2.5 million years ago with the first recorded uses of stone tools and ends at the same time as the end of the last ice age. There is no clear material criterion to distinguish it from the Mesolithic.
	Midden	250,000	PLM	
	Oud		PLO	
Undetermined	-	-	IND	There is no indication of a date

Table 4 - Period Table compiled from the reports for classification

Their codification also was not standardized and there were instances of textual classification, but others used acronyms of different forms. The most common representation found was the three-letter acronym using the two levels (Period and Subperiod) of classification. That was used to integrate all recorded data included and adapted to the artefacts that had only BeginDate/EndDate pair filled. It was also created the "undetermined" category to fill the remaining records.

Following the table, the existing codes were adapted to the ones in the table, so the periods are in the same classification for all artefacts that have been dated. The dates are not precisely defined as most of the dating methods have error margins that could easily move them from one period to the next, especially in more recent ones. For instance, the average error of the Carbon14 method is ±30 years, and even more recent classifications by style can usually only pinpoint their start, as the fashion of objects can go back and forth, as can be seen in the recent "vintage" trend.

4.6. Step 4: User Assessment

The commercial archaeology data stored in the repository already achieved its original goal of assessing the archaeological value of a specific area and evaluating the disturbance of the planned construction works. If the value assessed is considered to justify the time and money needed to salvage and or document the material and structures present, it encompasses also an excavation.

This study proposes to increase these datasets' contribution by making them also able to be integrated into other research efforts. The improvement of the usability would not be possible without the assessment of how the data are being currently used, or not, and what kind of information is being sought by their users. To this end, archaeologists were asked about their experiences and strategies to deal with these datasets.

Bearing this in mind, the following questions were posed:

1. Have you used these datasets?

If yes,

a. How often these datasets are used

b. What was your experience with them?

c. What was the importance of their Geospatial component?

If not

d. What is the reason?

2. What do you think could be done to stimulate the reuse of those datasets?

3. If tools for visualizing these data were available, what kind of tools would enhance their usage

The questions were not meant to get an exhaustive statistical result but to have a qualitative evaluation of the overall view of these datasets by professionals working with Dutch Archaeology. Of the archaeologists approached, it was possible to get answers from five professionals. The way of conducting these interviews was according to the subject's availability. Two of them were done face to face, two were done by email and one was done through a Teams meeting.

Three of the archaeologists work in the archaeology department of municipalities in the Netherlands. The municipalities are major users of this kind of information, as they are responsible for managing the heritage found within their territories. One archaeologist works as a freelancer and has been contracted by different municipalities as a consultant, and the last one is a lecturer in archaeology and is doing a PhD research on archaeology at the University of Leiden.

4.6.1. The use of the datasets

Three of the archaeologists answered that they did not use these datasets, although two affirmed they had accessed the database. They pointed out the difficulties in integrating the data due to multiple, or outdated, data formats, the perception of gaps in the fieldwork done and the lack of better description and interpretation in the reports which prevents the recorded data to be useful for integration into further research.

The other two professionals extracted the information they needed from the datasets, but both of them applied different methodologies and used other platforms to overcome the perceived difficulties in searching and sorting the answers from the database.

One used the official platform for archaeological management from the Dutch Cultural Heritage department Archis (https://archis.cultureelerfgoed.nl/). The platform has its access restricted to the professionals registered in an institution that is legally able to do archaeological work in the Netherlands and allows visualization of the archaeological interventions in the whole country. All projects need to be registered there to obtain the Onderzoeksmeldingnummer (Investigation Report Number). This number is the key that integrates the databases from the municipalities within the national one. The reports of fortuitous finds from the citizens that stumble upon archaeological artefacts by accident are also stored in Archis. There, it is possible to search by location of interest and to filter it by period or type of finding, as the professional should fill in all the information when the report is concluded. The geospatial information available is a polygon that delimitates the area intervened by the archaeological work, or a point if it constitutes a find by chance. Based on this information it was possible to sort what reports were relevant for the work and to access them in the DANS database.

The other interviewed person used ArcheoDepot (https://www.archeodepot.nl/), which is a joint Archeology Data Service organized by the municipal governments to record the artefacts found in their respective territories. It contains a more detailed description and assessment of the findings, but it does not contain geospatial data associated with them. But based on the information obtained there, it was possible to sort the reports so they could be consulted in DANS.

All professionals, including the ones who responded negatively to the first question, mentioned the importance of geospatial data for archaeological work. It allows to reassess and integrate the data found in each new intervention. The interpretation of the new findings depends on their positions and of the others in the surrounding areas for being made sense of the whole assemblage. The correct positioning of structures and artefacts is crucial for the evaluation of any archaeological site.

4.6.2. Enhancing the reuse

Asked about possible suggestions to enhance the reuse of DANS datasets, the responses were mixed. On one side the professionals indicated a lack of interest in doing so, not only because of difficulties in finding the data but also for methodological reasons. One is the way that commercial archaeology research itself is conducted which restrains its use, the main issues raised were poor recording techniques and lack of deeper on-site interpretations. Both would stem from the time and money constraints affecting commercial archaeology as a whole. Another view is that if they were easier to be found, they could be integrated, but currently some municipalities have already more data than they have qualified personnel to handle them. And finally, it was pointed out that there is not much time available to make practical use of them in the current reality of Dutch commercial archaeology.

On the positive spectrum, the suggestions to increase its reuse mentioned mainly an increase in the options for advanced searches, which could allow the exploration of the content without having to add more steps in the process.

4.6.3. Tools for exploring the data

Questioned about tools that could improve the experience in a portal to visualize such datasets the answers mainly pointed out that the professionals usually download the data they need to integrate it with the data they collected in their fieldwork. The online option for data manipulation or analysis was not seen as an essential feature in the current reality of their research.

The main concern raised was about the format of the data to be found. Different preferences were posed, the main one mentioned was the need to be compatible with open-source software, like QGIS, as it would be the type commonly used by them. According to one of the professionals, there were discussions about it among the ArchaeoDepot users that the geospatial data could be exchanged there in XML format, but it was never put into practice.

The only actual functionality cited as desirable, by one archaeologist, was the ability to draw custom polygons directly on the map and use them as the boundaries for the search area. This way would favour the extraction of reports located in more restricted areas than a municipality. It was pointed out as a factor that could improve the experience of the user when searching for these datasets.

4.6.4. Overall Results

The user experiences with the commercial archaeology datasets raised concerns about the overall quality of the archaeological research done by archaeology companies. The Dutch Council for Culture (Raad voor Cultuur) in an evaluation report about archaeology development under the Heritage Act came to similar conclusions. The report indicates, for instance, that 5,189 mandatory examinations were conducted in 2020, with more than half of those performing field research. According to it, commercial archaeology is responsible for approximately 90% of all archaeological research in the Netherlands. On the other hand, it is also pointed out that the quantity of archaeological projects seems to have a detrimental effect on the quality of the reports. Most of the investigations do not present much more than a brief description of what has been found. The council questions the excavation reports' usefulness for scientific synthesis and their contribution to Dutch society. The reasons summarized in the report for that are, as the professionals also commented, the lack of time and money for doing proper research. The council considers them as results of the over-competition in the sector, which leads to the insolvency of some archaeological companies, and the lack of interest of the municipalities to ensure that the research is being properly conducted under the risk of scaring investors (Raad voor Cultuur, 2022).

Adding to the information collected by the interview and reports from the authorities, it was also possible to access first-hand legacy archaeological data during an internship performed in the municipality of Nijmegen during the timespan of this study. The excavations there were conducted through the 1970s and the 1980s, the results were recorded in excavation plans showing the structures and findings from the Roman period in the area of the municipality. Several professionals conducted the interventions during different excavation seasons, which resulted in, sometimes, neighbouring excavation floorplans not matching exactly each other due to differences in the methodological approaches, layer definitions and even recording styles. Even though those were not originated from commercial archaeology, it was possible to have a solid grasp of how hard it can be to integrate these datasets into current research and how the cascade of interpretations (from the original archaeologist working to the place, to the data recorded, to the nowadays interpretation of the data collected for the digitisation process) may hamper the synthesis of the information available. It gave a new perspective on how important is to make the data available as easily and as complete as possible so the necessary time can be devoted to the interpretation and integration within the current research being conducted.

Summarizing the results obtained from the experiences of how professionals had been dealing with these datasets, the use of a visualizing tool that allows a deeper search into the datasets is already the strategy applied to these datasets. The availability of an open access alternative to the current ones could also provide access to any researchers, not only to the ones authorized to promote excavations in the Netherlands. The current strategies also reassure the proposal of this study, as the archaeologists have already been looking for information on finds and using it to visualize their geospatial distribution.

4.7. Step 5: Geoportal Functional Requirements

Vockner et al (2013), define a geoportal as "the information broker between geospatial resources and their potential users". For Innerebner et al (2017), it is "a type of web portal used to discover, view and access spatial information using geographic services (display, editing, analysis, etc.) via the Internet". And for Jiang et al. (2019), it is "a point of access to spatial data and geo-information".

All definitions present the geoportal as a tool to allow access to geospatial content through the web, which fits very well intending to offer a centralized way for archaeologists to interact with commercial archaeology datasets available online from their repository sources.

The presentation of archaeological and heritage data through a geoportal has been used in several studies in different parts of the world (Djindjian, 2008; McKeague et al, 2012; Prinz et al, 2014; Ronzino et al, 2018; Boschetti et al, 2019; Miguel-Castro & Fernández-Pareja, 2019; Lerma et al, 2020).

For the design of the geoportal, the answers gathered from professionals were taken into account, but the specific research needs are the main factor for accessing the information and they are driven mostly by its field and objectives. To illustrate this, it will be used the hypothetical case of a researcher that wants to evaluate the presence of palaeolithic remains and their nature within the Enschede municipality.

The researcher needs to visualize the archaeological data available and explore their descriptive properties. The information will be retrieved from the file containing the archaeological data created in the previous steps. It also needed a geographical base map to be able to identify where in the current time remains were found to have a better assessment of their spatial distribution. This can be used for planning visits, new excavations or to ease the contact with the municipality about details on the landscape.

It is not expected that the final researcher will have time or knowledge to explore in depth the available technical capabilities, as the interviews showed there are instances where the lack of time and technically trained personnel can be a hindrance to data processing. For this reason, the interface and operations are aimed to focus on offering the answers to the search with minimum need for input information and as simple as possible interface.

The raster information to be presented is retrieved through a Web Map Service (WMS), using the standard ISO 19128 WMS 1.3.0 (OGC, 2006) to ensure compatibility with other possible layers added in the future versions of this proof-of-concept geoportal. The initial version uses the service from OpenStreetMap, as it can show enough detail to meet the needs of the example research user.

The integrated datasets are offered as Web Feature Service (WFS) using the standard ISO 19119 WFS 2.0.2 (OGC, 2014). In the proof-of-concept, some operations of the WFS are embedded in the interface of the geoportal. The GetCapabilities operation result is explicitly shown in the options available to the user and the GetPropertyValue operation is available through a dropdown menu to show the options and a select button to perform it. A schematic view of the expected interactions can be seen in Fig. 15.

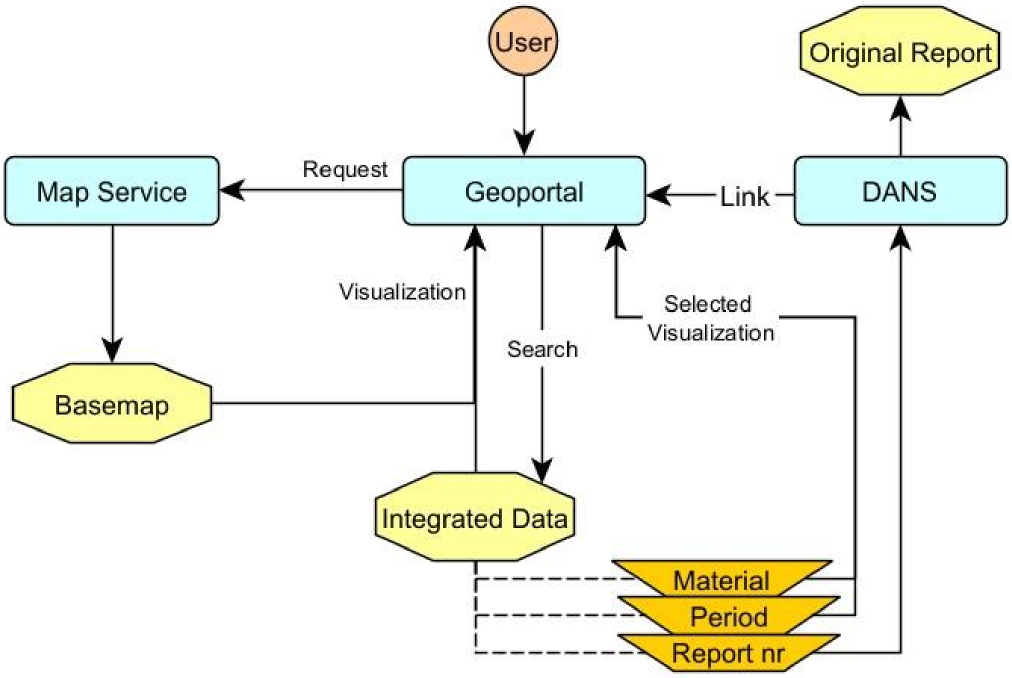


Figure 15 - Geoportal working schema

The metadata for the datasets points to the ones offered from DANS, as the construction of proper metadata falls beyond the scope of this study. The DANS metadata presents characteristics that made it very suitable for use in the context of the datasets as can be seen in Fig. 16. It uses the Dublin Core Terms (standard language for knowledge representation adopted by many institutions) that are supported by the Dublin Core Metadata Initiative (an organization aimed to promote and share standards on metadata design and best practices). The Dublin terms also follow international standards such as ISO Standard 15836:2009; ANSI/NISO Standard Z39.85-2012; IETF RFC 5013 of August 2007 [RFC5013] (E-depot for Dutch Archaeology, 2021; Dublin Core, 2022).

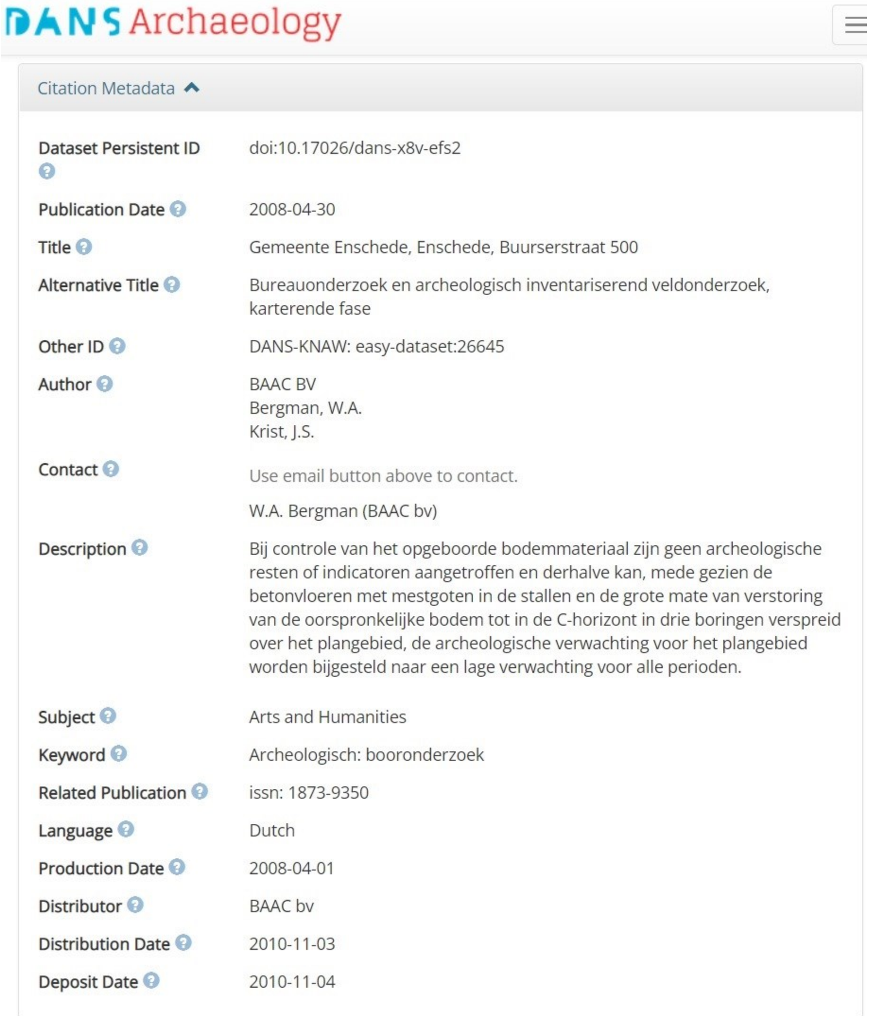


Figure 16 – Example of DANS metadata content Source: DANS

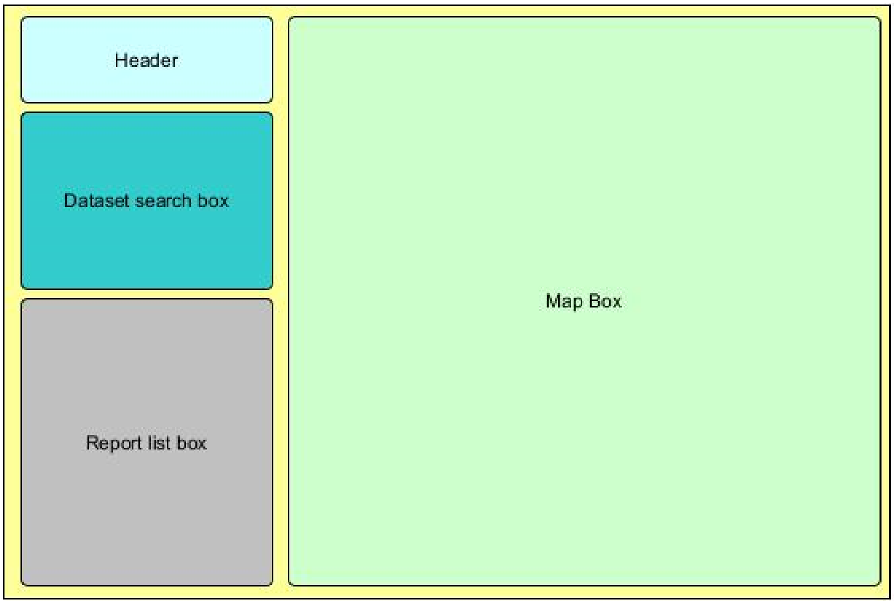


Figure 17 – Geoportal visual mockup

The structure of the geoportal is based on the OpenLayers (2019) version 6 application program interface, whose free JavaScript libraries contain tools for displaying map data in most web browsers available and contain all basic functionalities needed for the operations expected. For the handling of the data will be also used other Javascript libraries, such as Data-Driven Documents (https://d3js.org/).

Due to the requirements of OpenLayers, the file needed to be converted to one of the accepted formats so the geoportal could sort the data. The file was converted to the GeoJSON format that fulfils the requirement for OpenLayers, but keeps some of the advantages of the original CSV and adds the ability to store also the points' geographic coordinates in its geometry. This process also included the conversion of the coordinates to the WGS 84 coordinate system, more widely used in the internet environment due to its global scope. The limited scope of the dataset, containing a relatively low number of attributes that were retrieved from the datasets allows the use of a single file for the sake of simplicity in this limited proof-of-concept. Any expansion in its scope or a future prototype to be developed would necessarily need the use of a proper database solution that can be also capable of storing the geospatial information associated with the records, such as the open source object-relational database offered by PostgreSQL/PostGIS. (https://postgis.net/)

The search section allows the selection of the properties of the standardized fields to filter the data. Ideally, the fields could also be combined to narrow the visualization even further. However, due to the nature of the period data, the results would not depict the information available, giving a false impression of completeness for the results. A search using both time fields as limits for wider periods would have more or less effective results in the current dataset when searching the Palaeolithic, or the Late Modern Period. Those are the first period and the last period and are less harmed by the recording structure that keeps track of the beginning and the ending periods because they cannot be entirely skipped. But this is not true for all cases, an artefact that was made from the Palaeolithic to the Neolithic would not have any reference to Mesolithic in its record, but still should be counted as a possible match for the period. A result, in this case, would give a false impression of covering the whole dataset, which is not true. For this reason, the searches in this step can only be done for each field individually by the direct search of matching results into the dataset.

The search is performed using the button "Select" the property chosen is sought and only the points selected will be shown on the map.

It is possible to retrieve information about the artefacts' material as it is recorded in the original table, by selecting it from the dropdown menu, where the SIKB codes were represented by the materials' names in English. A similar procedure is done for the search by period. There the fields BeginPer and EndPer were combined in the search and it is possible to find the periods that are directly recorded in the table. As in the case of the materials, the codes used are also presented with their names in English.

The same reasoning was used for not including combined time/material searches.

Under the buttons, a report box will show a link to the DANS dataset where the original report(s) can be found. The default state of the dropdown menus is blank, when there are selected points, those will show in the map and the correspondent link(s) in the report box. In the case of no result, a popup box is shown with a message "no artefact found" together with a blank map and empty report box The button "Clear" is used to return the dropdown menus, the map view and the report box to their default blank status.

The limitations encountered in performing the searches showed another layer of issues in exploring the datasets. It gave one more reason to test the capabilities of implementing ontologies and knowledge graphs to search the attributes offered by the datasets.

4.8. Step 6: Ontology and Knowledge Graphs

The three ontologies mentioned earlier (Getty-AAT, CIDOC-CRM and IDAI-Chronontology) were consulted to evaluate their merits and issues for application in the current study. All have a complex structure and a comprehensive capacity for characterization within their different scopes.

CIDOC-CRM gives its focus to the events and tries to connect artefacts, images and other materials to them through a series of relationships. Its conceptual reference model (Doer et al, 2020) presents it as an event-centric model, aimed to support historical discourse by enabling the description of time-limited processes or evolutions. Its adaptation to archaeological datasets can be time-consuming, ARIADNE has a multidisciplinary project on a derivative adaptation for the field of archaeology. The initiative started in 2014 and it has been divided into different ongoing modules, the metadata-focused module is being tested in projects related to Greek heritage. The focus on events, which are not depicted on DANS datasets would require a more convoluted approach for adaptation that risks obscuring instead of making it easier to find the data sought.

IDAI-Chronontology was created to document archaeological monuments, excavations and archives of interest for the German Archaeological Institute (Deutsches Archäologisches Institut – DAI). So it gives bigger attention to the processes and organization used by the institute and it uses connections to other ontologies for many parts, including descriptions borrowed from the Getty Foundation. It seems well adapted for the institute application, although its focus on its areas of interest makes it difficult to match with areas not dealt with previously.

Getty-AAT was developed initially to meet the needs of museums and art galleries. For that reason, it emphasizes objects which hinted at interesting match possibilities with the datasets being used. However, its reliance on style, authors and current location and the characteristics classifications of the data makes it difficult to fit its categories.

The limited information contained in the datasets being used makes it difficult to reclassify them in the amount of detail and precision the ontologies require.

One way to exemplify this is by explaining the mismatch of chronologies. Even in Europe, the dates of how the different periods are classified become confusing due to regional differences. As seen in Fig. 18 from Isern, Fort and Linden (2012, 3) the starting date for the Neolithic can have three thousand years of difference between the first appearance in the Balkans until the time it reaches Scandinavia and certain areas of Northern Europe.

This uneven distribution of events generates several local chronologies and tends to fragment the classifications in many different periods. Similar names can have hundreds of years of difference and some periods can be completely skipped, as happens in Scandinavia, where there was no Roman Period, and the Iron Age was longer and was succeeded by the Middle Ages.

Some periods present in the classification of the data were not present in the Getty-AAT vocabulary, or have a completely different meaning. For instance, the Early Roman Period, which corresponds to the invasion and annexation of the lands South of the Rhein to the Roman Empire, can be equated to the Early Imperial Period for the general chronology. The subsequent periods are broken by each Roman Emperor's rule because it follows the Italian chronology, which makes it impossible to match the sequence for the Netherlands without a more precise dating of the artefacts, which is not present in the dataset.

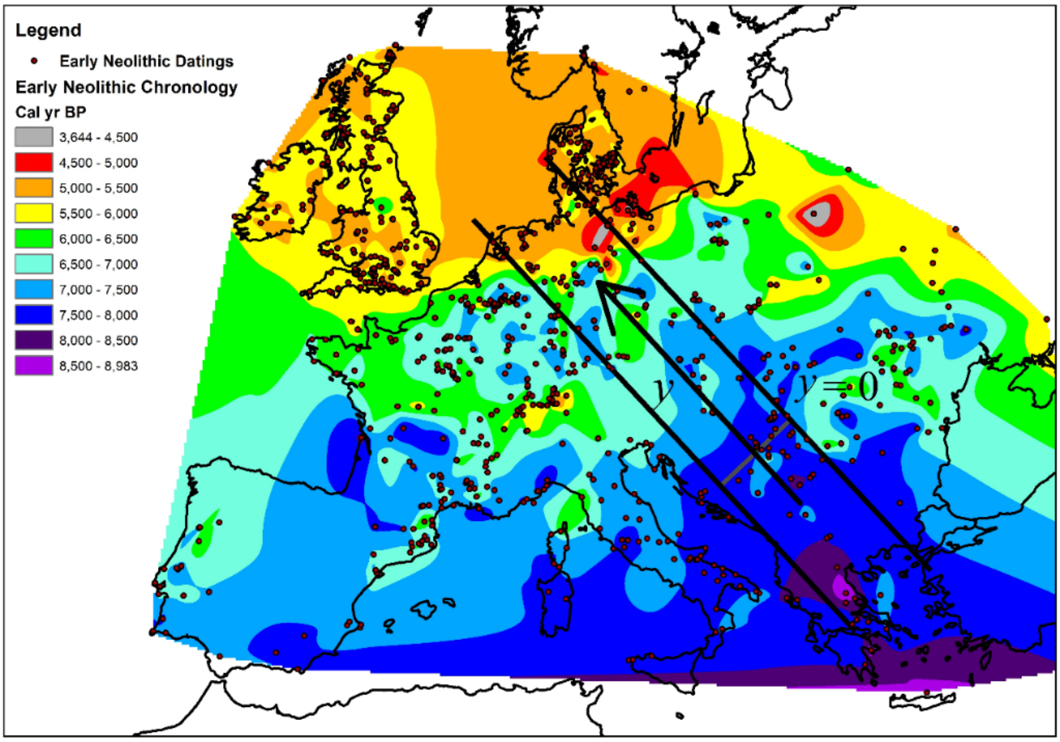


Figure 18 - Starting dates for Neolithic in Europe (Isern, Fort and Linden, 2012, 3)

The creation of an ontology that can match the local chronology (as can be seen in Fig. 2) with its different phases fitting into the general chronology of events is a daring endeavour that cannot be taken lightly and escapes completely the scope of this study. Even if there was enough time and knowledge to do it, in most cases there is not enough information about the artefacts that can allow their match within one of the local cultural traditions or precise dates that can make it possible to fit them inside more detailed chronologies than the Early-Middle-Late structure already in place.

So a simplified chronology comprehending only the terms included in the actual classification of the data was devised to make it possible to explore its relationships (Fig 19). The main interest in designing their relationship is to be able to position them in a timeline, considering the relative position between two different periods. The queries should be able to retrieve all artefacts that have their field EndPer filled with the periods that precede the selected final period.

A practical solution for the chronology was found by using partially the time ontology made available by the World Wide Web Consortium (W3C). It uses the Web Ontology Language and it was developed in a joint project involving W3C and the Open Geospatial Consortium. Each period can be represented by a ProperInterval class, and their relationships can be easily scalable to the level of detail needed. The relationships between periods can be represented, the major divisions can have their start and end marked by their subdivisions and each of them can be chained by the property intervalMetBy which marks the end of the previous period and the coincident start of the next. The connections of each subdivision to its main period were made by the property intervalIn and allow integrated searches one level above (W3C, 2020). An excerpt of the full chronology adapted to the OWL Time Ontology can be seen in Fig. 20.

The full ontology developed for the chronology is presented in the Appendix (Annex 1), and as can be difficult to visualize it in the GraphDB visualization, it is also presented in a diagram that allows a better reading.

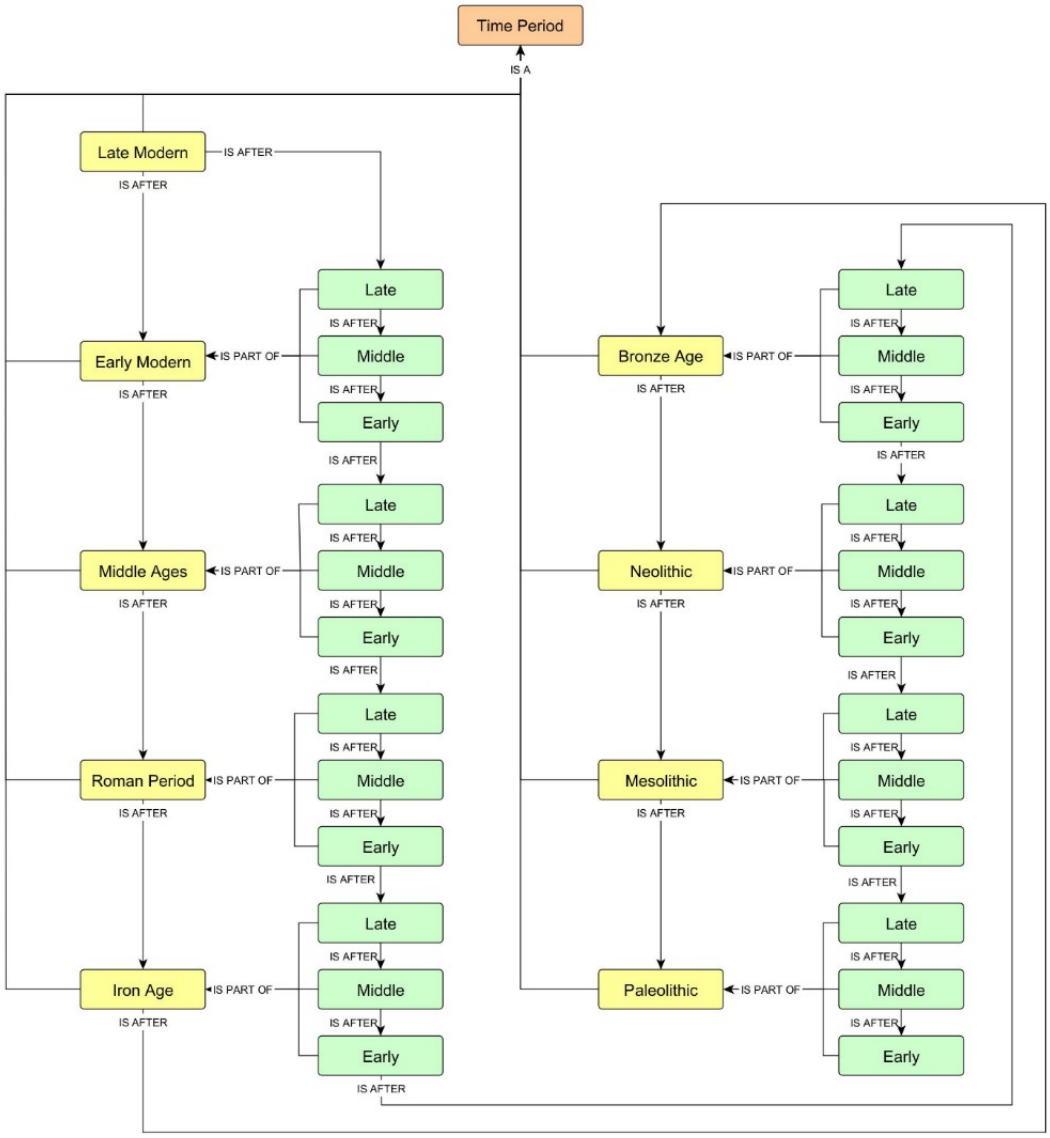


Figure 19 - Conceptual structure devised for applying an ontology to the available data – Chronology

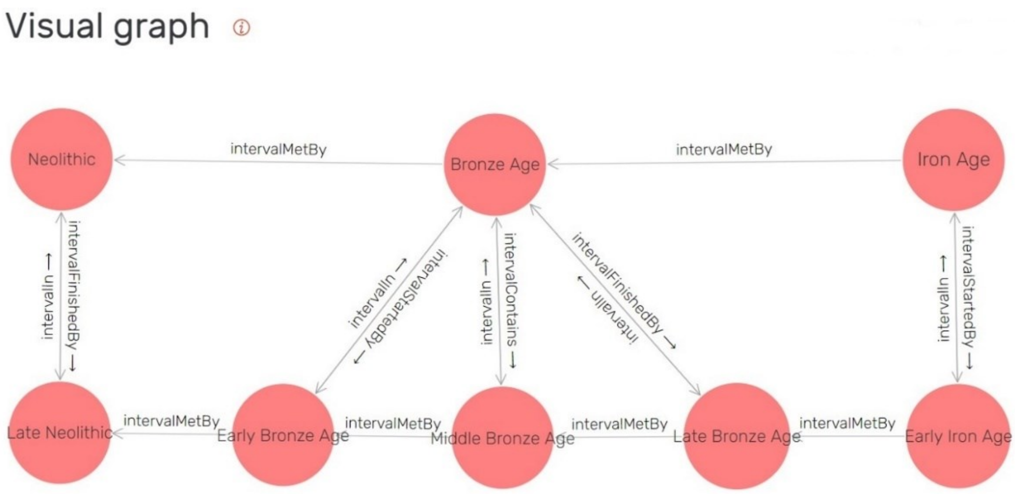


Figure 20 - Excerpt of the chronology from GraphDB

The materials can be represented with a simpler structure, which includes classes that allow searching all materials by their category. This makes it possible to look for broader subsets of data (Fig 21). The materials are currently classified by their type of material (iron, copper, tin, flint, granite, diabase). This demands a combination of those searches to obtain results for broader categories like stone (flint, granite, diabase) and metal (iron copper, tin). The codes used from SIKB already have a hint on that classification as all lithic artefacts classifications start with an "S", all metals start with an "M", and all organic materials start with an "O" and so forth.

For the materials, the RDF Schema (W3C, 2014) was used to model them as classes, with their subdivisions as subclasses and the entries that appear in the actual table as instances. This allows the desired flexible approach to selecting types of materials, not only their instances.

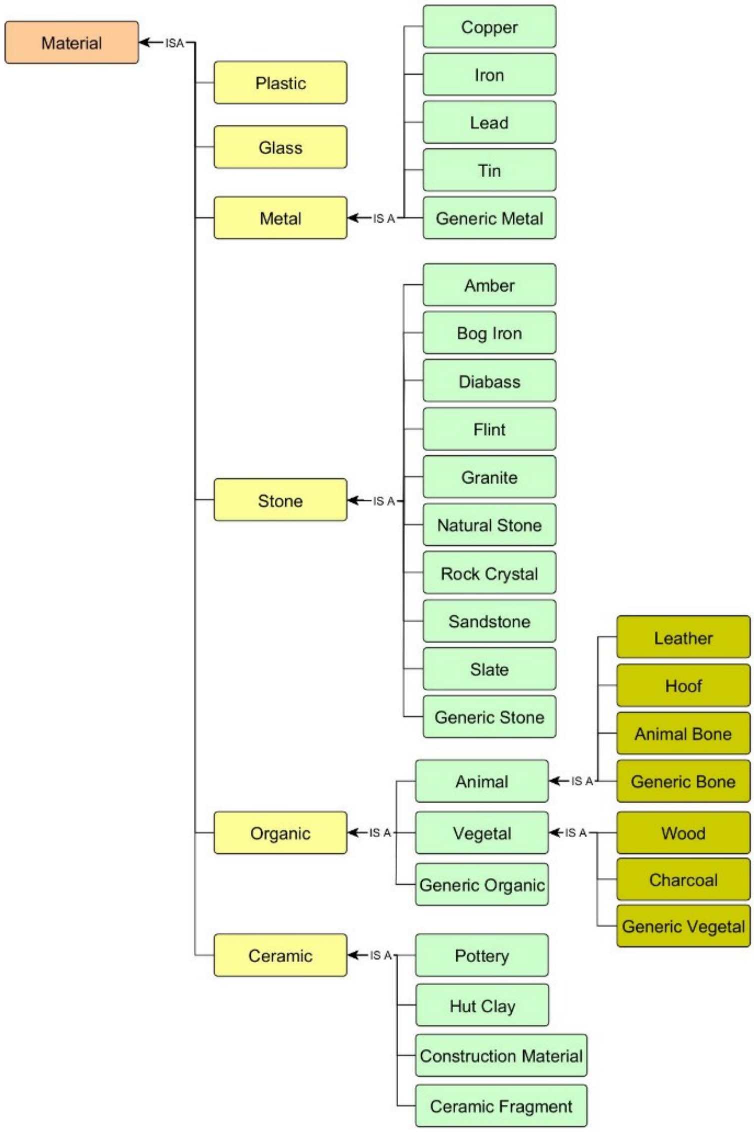


Figure 21 – Conceptual structure devised for applying an ontology to the available data - Material.

This structure also allows modelling more complex relationships like the connection of Bog Iron and Iron as different instances of the same material. (Fig 22). Bog iron is formed in shallow groundwater areas with the presence of iron minerals. These minerals are destabilized by their low pH values and their dissolved iron components are precipitated and later oxidize together with the bottom ground sediment giving origin to a more or less porous mass with high iron concentration (Thelemann et al., 2017). It was commonly exploited for early iron production as its chemical characteristics make its iron content easier to be extracted. (Thelemann et al., 2017; Brenko et al, 2020) The presence of the bog iron can be interesting for archaeologists studying the presence of early metallurgy in the region, so it would be desirable that its reports also appeared in searches focused on iron artefacts, especially in pre-historic contexts.

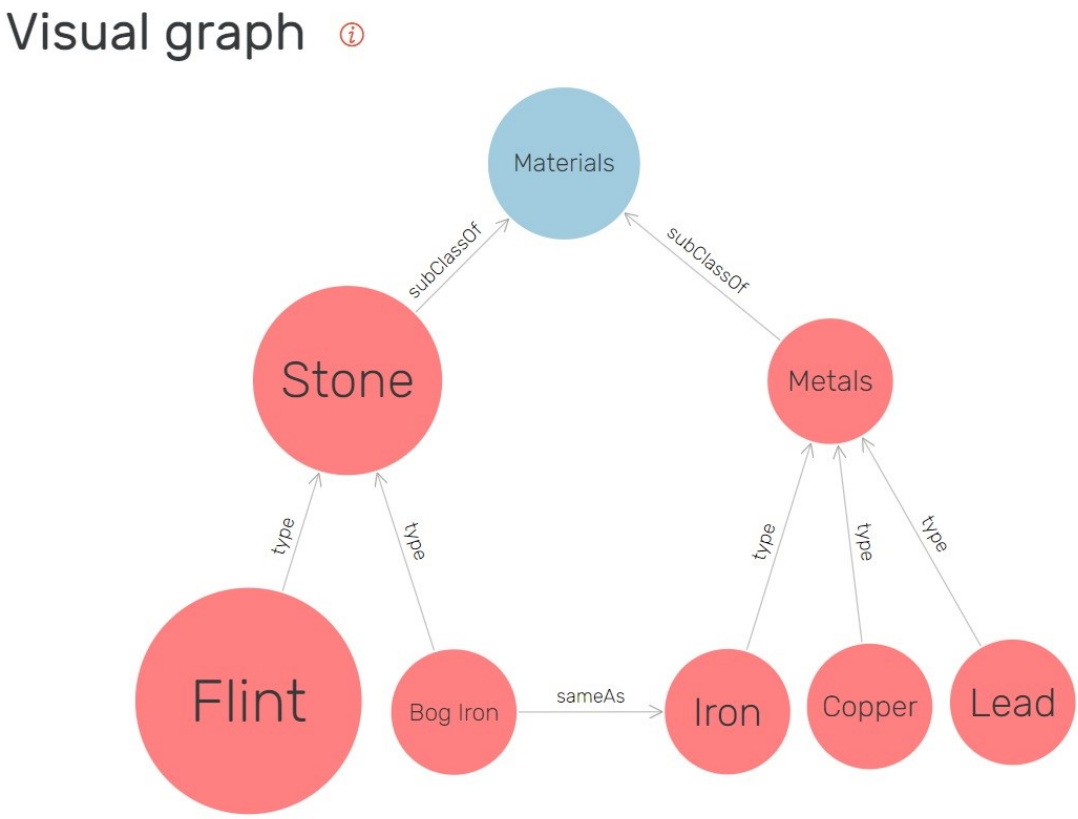


Figure 22 - Excerpt of the materials graph structure from GraphDB

Departing from those ontologies the knowledge graphs will be populated by the data entries from the tables obtained in the early phases of the work by converting the entries on the findings table to RDF triples. The resultant database will then be able to be queried differently from the data directly inserted in the geoportal. The next chapter will present and compare the different results obtained for each method.

The full ontology developed for the materials is also presented in the Appendix (Annex 2), and as can be difficult to visualize it in the GraphDB visualization, it is also presented in a diagram that allows a better reading.

5. RESULTS

After the whole process, an integrated dataset was obtained from the reports generated by the archaeological interventions in the Enschede municipality. As a preliminary way to evaluate the results, the five criteria were analysed for the usability assessment of the original way datasets. As a way to compare the procedure needed for exploring the datasets, a small case study is presented and it compares the results of three instances: the current state of the dataset exploration, the use of the geoportal with the integrated dataset and the application of knowledge graphs.

5.1. Reassessing PARTHENOS guidelines

According to the PARTHENOS guidelines, reprised in Table 5, there are five steps to enhance the usability of heritage datasets.

16. Document data systematically;
17. Follow naming conventions;
18. Use common file formats;
19. Maintain data integrity;
20. License for reuse;

Table 5 - Usability criteria extract from the guidelines at PARTHENOS et al, 2018.

For the PG-16, "Document data systematically", now it is possible to clearly describe the dataset content, as it has been unified and standardized. Metadata containing the major material types and periods discovered in the municipality can be easily derived from the content. It is not a finished work though, as many other fields need to have the same treatment to have their content as streamlined as possible, so it is a work in progress.

The PG-17, "Follow naming conventions", has been applied to the dataset content, and now the fields have a single naming table, and the same happens to the information in them. That makes the whole array of information available to be analysed and makes it easier to add new information. This standardization enhances the reuse capabilities, as the data now has defined parameters to be integrated into other databases.

For the PG-18, "Use common file formats", the format used for the table is CSV, which allows it to be open in a variety of software packages and multiple options for conversion. The TTL (Terse RDF Triple Language) format used for the knowledge graphs derived from the data is an open W3C standard, as it also happens to be the GeoJSON file used in the geoportal.

For the PG-19, "Maintain data integrity", the original fields were kept, even though the information was standardized in a different field. It helps to keep the data compatible with users that accessed previously the data directly from reports but also to make it possible to correct misinterpretations or processing mistakes.

The last guideline, PG-20 "License for reuse", is the only one that was not touched for the work, there are datasets included that ask for specific permission. That is also the reason why there is no provision for releasing the content for download, as the permission granted for the protected content was only to share its visualization.

Considering the guidelines, initially, it can be said that there was a gain in terms of the usability of the data. It results from the enhanced explorability obtained, as now it is easier to sort the reports using a wider array of criteria than before. However, to present a better perception of the difference from the current option for searching, a small case study will be presented in the three.

5.2. Case study – exploring the current state of the datasets

For this case study, we revisit the example of an archaeologist that desires to study palaeolithic sites in the Netherlands and wants to try to integrate commercial archaeology data seeking to evaluate the presence of palaeolithic remains within the Enschede municipality.

The first step is to look for the options of querying for this in the Ariadne Portal. In the filter field, we can look for Palaeolithic sites. Even though it points to more than seven thousand matches, the British spelling retrieves almost no results from continental Europe (Fig 23).

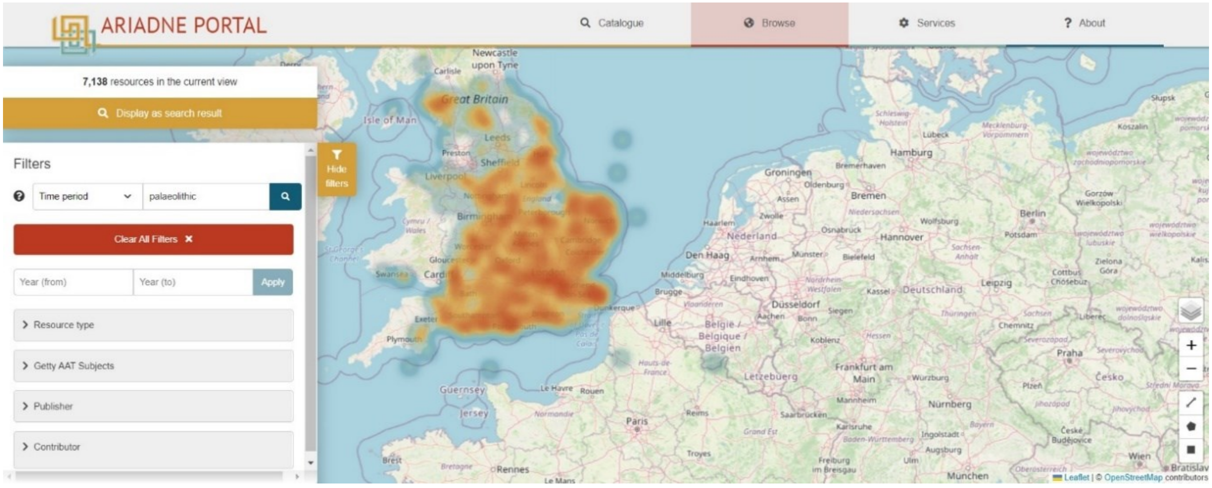


Figure 23 - Ariadne Portal search screen – Palaeolithic

Changing the spelling to Paleolithic returns a smaller amount of matches (Fig 24), but now it is possible to find them in the continent and at the same time heavily reduces the number of results in the British Isles. It is notable that due to the reduced number of results, the portal automatically changes the display from a heatmap to point groups. Nevertheless, there are still no relevant results for the Netherlands. Zooming in the results, the closest matches to the Paleolithic search are from the vicinity of Maastricht in the south of the country. No results are retrieved anywhere close to Enschede. (Fig 25)

There are further options for narrowing the search if there were any. The categories are relevant and they could be useful for retrieving relevant results as can be seen in Fig 26. Even though the commercial datasets found could be easily classified in the resource types offered, the subjects from the Getty AAT ontology are not readily available for classification. Even after processing most of the types of artefacts classification (blades, cores, flakes, axe heads, arrowheads etc.) are missing in most of the commercial records, limiting them to the raw material classification.

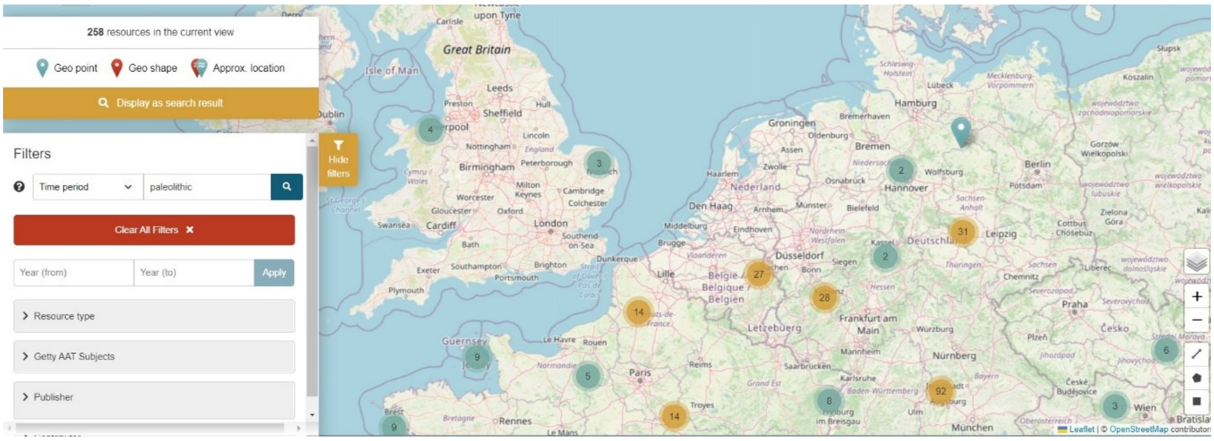


Figure 24 - Ariadne Portal search screen – Paleolithic

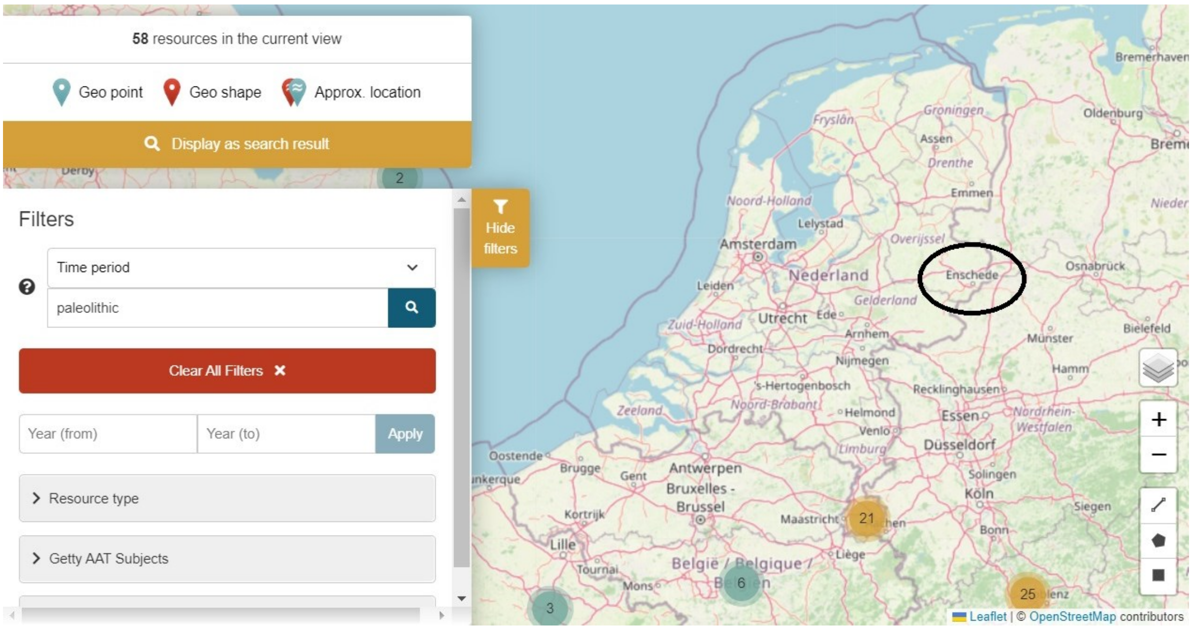


Figure 25 - Ariadne Portal search screen – Paleolithic in the Netherlands

This sends the researcher back to the spatial search that was done previously, changing the filter to place and by selecting Enschede the results are retrieved in the map, similarly to what was done in Fig. 5. Finding the reports moves the focus to the DANS website, which offers slightly different options of search.

If instead of making a spatial query, for which the Ariadne Portal is needed, a keyword query is tried directly in DANS Archaeology, some conclusions can already be drawn (Fig 27). The completely different result obtained in the DANS search shows that most of the data must lack geospatial indications, otherwise they should have been retrieved also through Ariadne. The apparent dichotomy between Palaeolithic and Paleolithic is also present in DANS datasets, but here the British spelling side retrieves seven times more entries. In any case, it is possible to be certain that the terms are not standardised to be retrieved no matter what spelling.

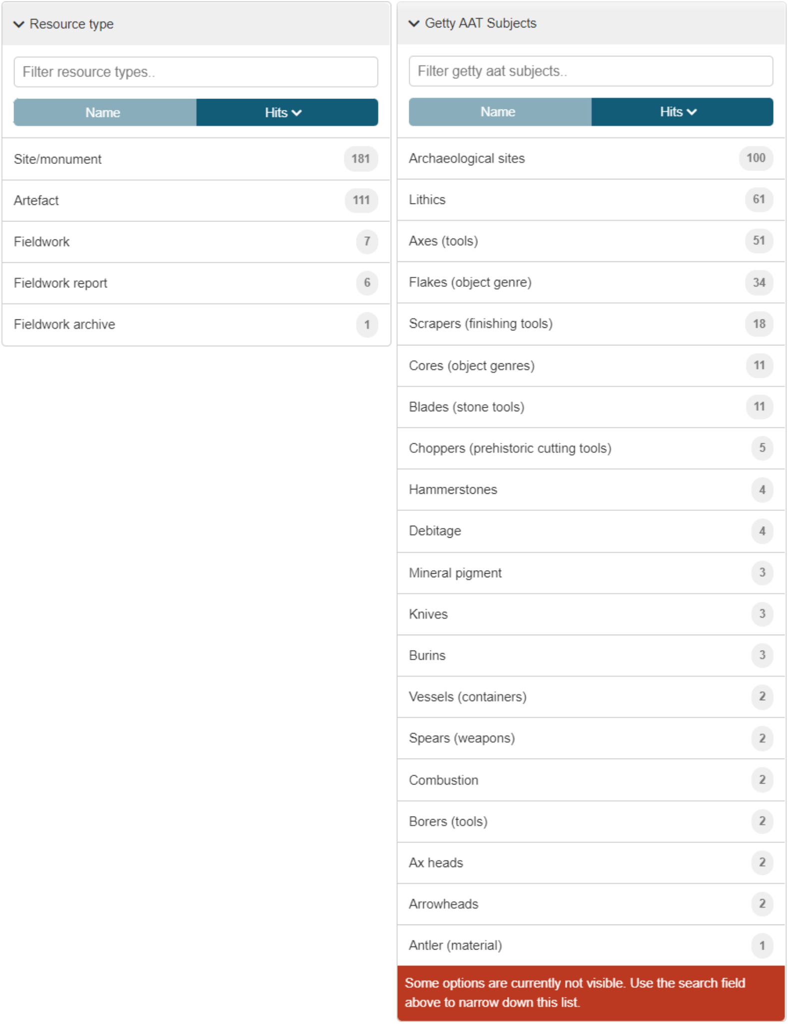


Figure 26 - Ariadne Portal filtering search options

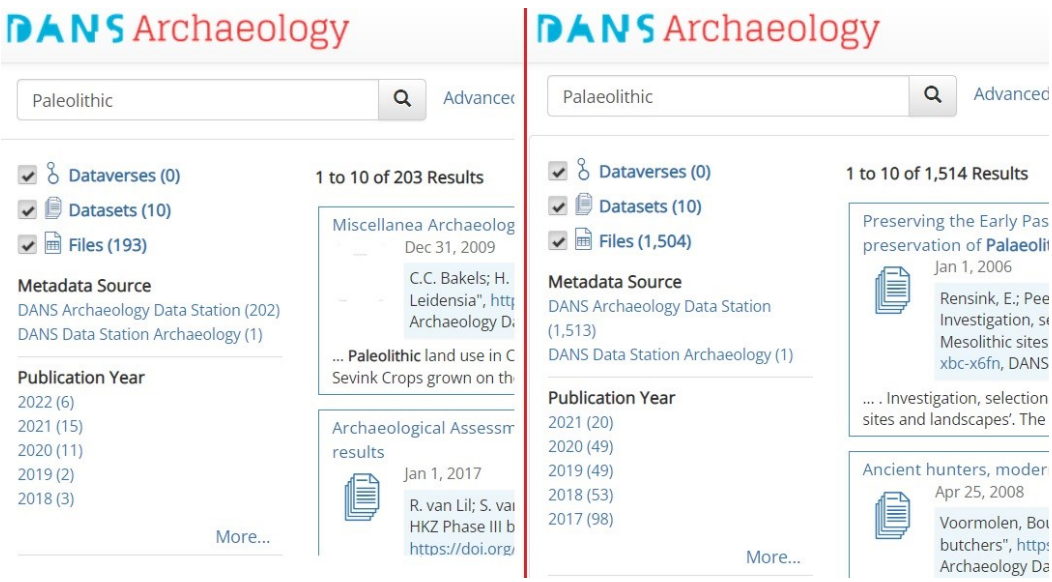


Figure 27 - Search on DANS - Paleolithic/Palaeolithic

The results also do not include any commercial archaeology reports, the results comprise only academic papers. It is not a problem by itself, as the matter of fact it is a great feat as the DANS portal makes it much easier to find the papers without having to look for them in different journals and at the same time gives access to the datasets used, if and when they are available.

However, in this specific case, the researcher is looking to integrate commercial archaeology data, and those datasets are not being retrieved through the search using *Palaeolithic/Paleolithic* as a keyword. After filtering the records by place in Ariadne and being redirected to DANS, opening their description, which is also available in the metadata is the alternative left to try to devise their potential for attending the criteria sought by the researcher. The text is often extracted from the report's abstract and it usually focuses on addressing the questions from the AMZ cycle. It can vary from one small paragraph to a longer explanation and sometimes there is no description at all. It may or may not reference the periods and artefacts encountered. When it does so, it mentions the most relevant ones and does not bring an exhaustive description of all of them. (Fig. 28)


Description

Archeodienst BV heeft een bureauonderzoek en inventariserend Veldonderzoek, karterende fase (IVO-O(verig): booronderzoek) uitgevoerd in het plangebied aan de Buurtweg 95 te Ermelo (gemeente Ermelo).
Het onderzoek is uitgevoerd voor de aanvraag van een omgevingsvergunning voor de nieuwbouw van een overdekte paddock met drie paardenboxen. De verstoringsdiepte bedraagt ca. 1.0 m –mv en voor een klein deel 1.8 m -mv. Eventueel aanwezige archeologische resten zullen daarbij verloren gaan.

Archeologische interpretatie
De natuurlijke holtpodzolgrond is in het gehele plangebied grotendeels tot geheel verstoord, wat al dar niet door het grondverzet heeft plaatsgevonden. Vuursteenvindplaatsen van jagersverzamelaars bestaan voornamelijk uit strooiing van fragmenten vuursteen en ondiepe grondsporen, zoals haardkuilen, in de bovengrond van de oorspronkelijke podzolgrond. Aangezien de bodem is verstoord, zijn eventueel aanwezige vuursteenvindplaatsen verloren gegaan. Daarnaast zijn er geen indicatoren aangetroffen, die wijzen op de aanwezigheid een vindplaats. De lage verwachting uit het bureauonderzoek voor vuursteenvindplaatsen van jagersverzamelaars uit het Laat-Paleolithicum tot er met Mesolithicum kan daarom gehandhaafd blijven.
Nederzettingsresten uit het Neolithicum tot en met de Nieuwe tijd bestaan niet alleen uit fragmenten aardewerk, maar ook uit diepere sporen zoals paalgaten en afvalkuilen. Deze sporen kunnen tot in de C-horizont reiken en zijn mogelijk nog intact. Tijdens het booronderzoek zijn echter geen archeologische resten of indicatoren aangetroffen, die wijzen op de anwezigheid een vindplaats uit deze periode. Daarom kan de middelhoge verwachting uit het bureauonderzoek om archeologische resten uit de perioden Neolithicum tot en met de Vroege-Middeleeuwen en de aan te treffen worden bijgesteld naar laag, terwijl de lage verwachting voor de perioden Late- Middeleeuwen tot en met Nieuwe tijd kan worden gehandhaafd. Hoewel de aanwezigheid van begravingen door middel van boringen vrijwel niet is vast te stellen is het gegeven dat er voor de herinrichting van het plangebied geen verhogingen te zien waren een aanwijzing dat deze waarschijnlijk ook niet te verwachten zijn. Daarnaast is de bodem door de herinrichting aanzienlijk verstoord. Daarom wordt de middelhoge verwachting op het aantreffen van grafheuvels en begravingen uit de perioden Neolithicum tot en met de Vroege-Middeleeuwen bijgesteld naar laag.

Op grond van de resultaten van het onderzoek acht Archeodienst BV een archeologisch vervolgonderzoek niet noodzakelijk.

Bureauonderzoek en inventariserend veldonderzoek karterende fase

Description

In opdracht van de gemeente Enschede heeft RAAP Archeologisch Adviesbureau van 8 juni t/m 15 juli 2009 een archeologische opgraving uitgevoerd in verband met de voorgenomen ontwikkeling van plangebied 't Vaneker te Lonneker, gemeente Enschede. De vindplaats bestaat uit twee delen die van elkaar worden gescheiden door de onverharde Sprakelweg.

Op vindplaats 5a aan de westzijde van de Sprakelweg zijn de undamenten van een boerenerf uit zijn laatste gebruiksfase blootgelegd. Het betreft de plaats van het voormalige erf Sprakel dat van oorsprong uit de Late Middeleeuwen dateert. Hoewel op historische kaarten meerdere erven met de naam Sprakel zijn aangeduid, is de hier besproken locatie op basis van historisch kaartmateriaal en de teruggevonden resten het oorspronkelijke erf. Daarbij behoort ook de huisakker of woerd ten oosten van de Sprakelweg, waaronder uiteindelijk alleen prehistorische sporen zijn aangetroffen. De voor woerden kenmerkende dikke plaggendekken zijn ook aan de overzijde van de Sprakelweg aangetroffen. Wanneer de voorganger van het huidige erf Sprakel, in de nabijheid van de hier besproken locatie, is gebouwd, is tijdens de opgraving niet gebleken.

Het onderzoek op vindplaats 5b was in eerste instantie geïnitieerd om eventuele voorlopers van de historische boerderij te lokaliseren. Behalve de vondst van enkele middeleeuwse (kogelpot) scherven, een enkele kuil en greppel is deze niet op deze plaats aangetroffen. De wel opgetekende, vage grondsporen maken waarschijnlijk deel uit van een klein nederzettingsterrein uit de IJzertijd. Het nederzettingsterrein heeft een duidelijke relatie met een lichte verhoging in het landschap die zich uitstrekt naar het zuiden. Dat deze verhoging al in de Vroege Prehistorie door de mens werd bezocht, wordt bewezen door de vondst van enkele vuurstenen artefacten en aardewerk uit het Neolithicum en de Bronstijd. Deze vondsten zijn vooral afkomstig uit de fossiele cultuurlaag of uit kuilen uit de IJzertijd. Uit de aangetroffen kuilen en paalkuilen uit de IJzertijd kon een mogelijke, slecht geconserveerde structuur herleid worden. Het betreft een rechthoekig gebouw van ongeveer 13,5 x 7 m.

Description

Het onderzoeks- en adviesbureau voor Bouwhistorie, Archeologie, Architectuur- en Cultuurhistorie (BAAC bv) heeft in maart 2008 een bureauonderzoek en een veldinspectie uitgevoerd in het plangebied Velve-Lindenhof te Enschede (gemeente Enschede). In één boring is een fragment bewerkt vuursteen aangetroffen in de top van het keizand. Deze mogelijke vindplaats dient gewaardeerd te worden door een waarderend booronderzoek om de aard en omvang van de vindplaats te bepalen.

Figure 28 - Examples of the length of descriptions in the DANS datasets

Assuming a limited time for the researcher, the most probable course of action would be to scan briefly the descriptions and look for the terms indicating the presence of finds from the Palaeolithic. Those reports would be then selected for a more detailed review and depending on their content could be integrated into the research. Due to the limitations of the descriptions pointed out previously, this approach is not only time-consuming but also risks missing any reports that do not mention the period in the text.

5.3. Case study – exploring the geoportal with the integrated dataset

Considering the same researcher, but now visiting the geoportal created for this study, whose final appearance is shown in Fig 29. The search to be performed can be selected on the left side of the screen, allowing the use of both standardised fields from the previous steps. The criteria for the searches are standardised drop-down lists containing each material or period that is present in the data acquired.

To obtain the answers, the researcher would select the three subperiods into which the Palaeolithic is divided in the dataset, Early Palaeolithic, Middle Palaeolithic and Late Palaeolithic. The combined results of those searches can be seen in Fig 30. The results point to eight different reports stored in the DANS datasets 21387,21653, 26646, 36895, 40237, 45911, 48482, and 59182. Checking their descriptions, the keyword Palaeolithic only appears in two of them (21387 and 59182), and a third mentions Mesolithic (21653) in the text, which would also be of interest to our researcher.

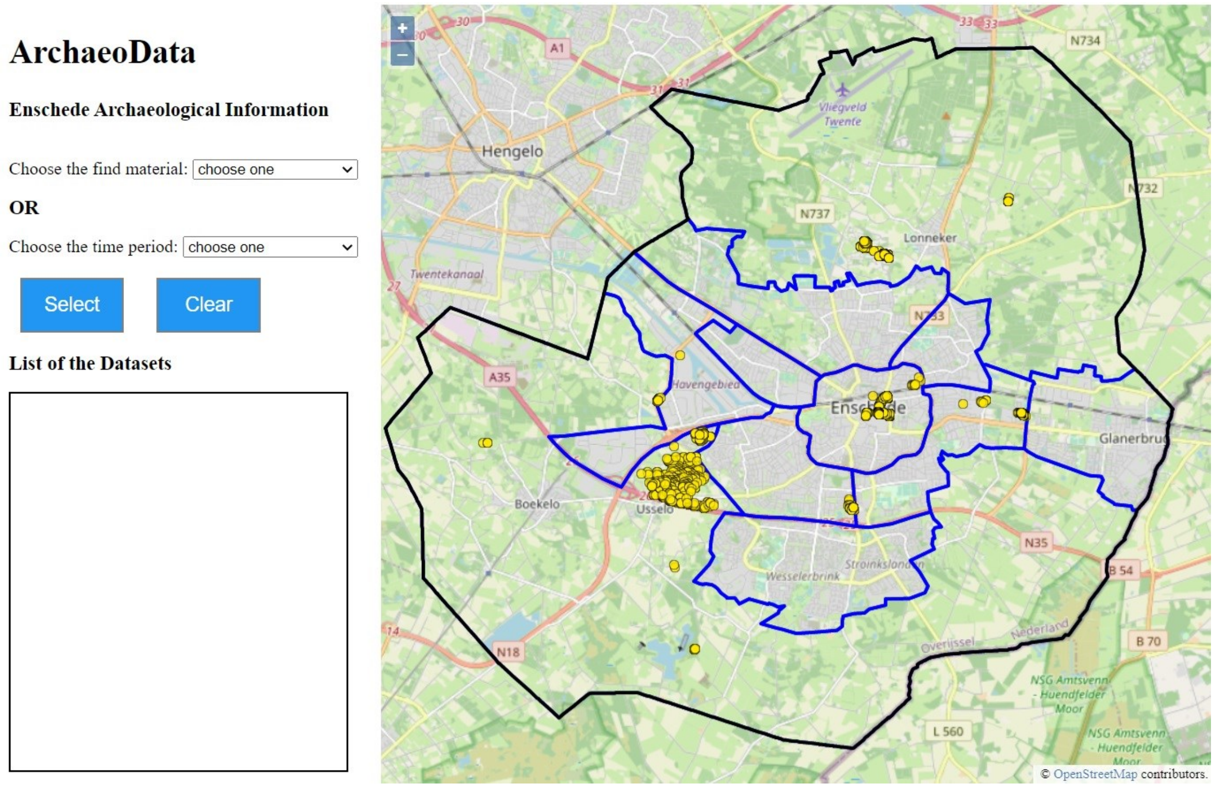


Figure 29 – Geoportal appearance

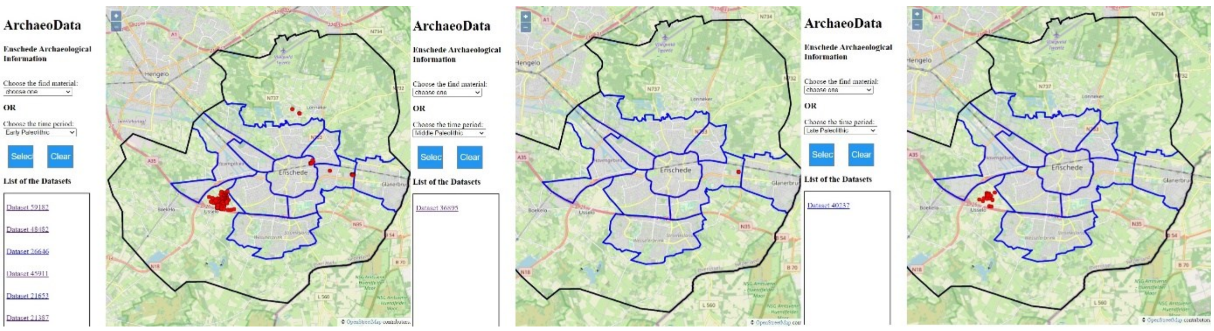


Figure 30 - Search examples of Early/Middle/Late Palaeolithic

It means that 60% of the results do not present any indication of period in their description and they would likely be missed by the researcher. Even this simple integration of the artefact data allowed an increase in the results available to be reached by the research.

As a way to explore further the datasets, the researcher also could look for reports that found flint artefacts, which are the main material whose characteristics allow the periodization in this period. Due to the lack of expert knowledge, some non-dated artefacts could be indeed from the Palaeolithic. This search retrieves 19 reports, from which seven were already in the Palaeolithic, adding twelve extra reports that could be also evaluated for this study (Fig 31).

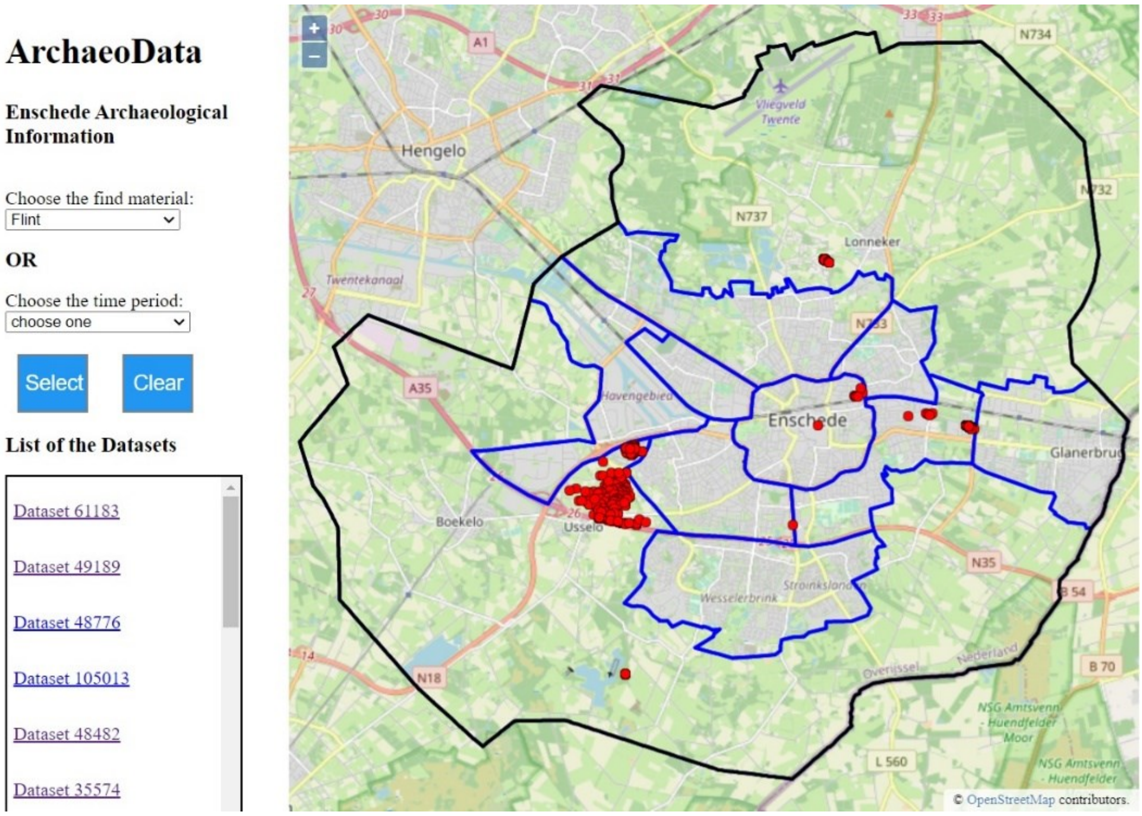


Figure 31 - Search examples of Flint artefacts

The portal can suggest reports as a possibility for evaluation, it will depend on the information delivered by the reports if it will be possible to identify them in a better way than the original archaeologists were able to do.

If the intermediary periods (between BeginPer and EndPer) could be better integrated, more precise results could be reached for all periods and different combinations could be tried. This can be done by inserting a full timeline structure into the code, but in case of further expansion for other areas, it could demand major fixes and quickly become cumbersome, and different for almost every area. The other option would be to add it to the database structure, but that would generate extra steps in the integration process and create several empty attributes, as it is not common to have a big gap between the beginning and the ending period recorded.

As an alternative to this, knowledge graphs can deal which such relationships through the ontology structures, which will be explored in the next section.

5.4. Case study – exploring the contribution of knowledge graphs

The goal of using knowledge graphs in the geoportal setting is to enhance the possibilities of searches without having to manipulate the original table itself. The creation of the schemas adapting the ontologies allows to ask not only for a specific material, or period, for instance, the Early Roman Period or flint, but to get the whole Roman Period, or any period desired. The materials relationship would make it possible to use aggregations levels to obtain all reports with all artefacts made of stone (lithics), or metal, and for organic it is possible to break in animal and vegetal remains.

For this task, GraphDB was used to create the triples based on the adapted ontologies adapted and populated by the dataset available. GraphDB is a semantic graph database, compliant with W3C standards used during the Master's course. It allows writing the queries in SPARQL, which is the standard query language for RDF triples (W3C, 2008).

The actual knowledge graph is not planned to be visible to the user, but it could be implemented through a similar system of drop-down menus, with wider options for searches. The visualization of the actual knowledge graphs by the user would need to be tested by test users to measure if it would contribute to the information being sought or distract from it. In the background, the search is made into triples using the relationships established by the ontologies. The answers identify the artefacts and this information can allow the display of points on the map the same way it is done in the first geoportal. It also generates a list of reports from where those points came from is offered for obtaining more details directly from the original reports.

Using the knowledge graph query structure (Fig. 32) is possible to extract the whole list of results of artefacts belonging to the class "Stone", the database identifier that points back to the reports stored at DANS and it can count the how many artefacts are recorded of each class by the report. It starts by stating the prefixes of the ontologies used and declaring the relationships that are being required to retrieve the information needed.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX adt: <http://example.com/base/ArchaeoData>
SELECT DISTINCT  ?mat ?report (count(?artefact) as ?quant)
WHERE {
    ?artefact adt:isMadeOf ?material;
              adt:isStoredAt ?EasyDb.
    ?EasyDb rdfs:label ?report.
    ?material rdfs:type ?var.
    ?material rdfs:label ?mat.
    FILTER (?var = adt:Stone)
}
Group by ?report ?mat
Order by ?mat
```

Filter query results Showing results from 1 to 49 of 49. Query took 0.1s, minutes ago.

	mat	report	quant
1	"Amber"	"61183"	"1"^^xsd:integer
2	"Bog Iron"	"104905"	"1"^^xsd:integer
3	"Bog Iron"	"59182"	"1"^^xsd:integer
4	"Bog Iron"	"59500"	"2"^^xsd:integer
5	"Bog Iron"	"35574"	"10"^^xsd:integer
6	"Bog Iron"	"33823"	"14"^^xsd:integer
7	"Diabase"	"61183"	"1"^^xsd:integer
8	"Flint"	"26646"	"1"^^xsd:integer
9	"Flint"	"35574"	"1"^^xsd:integer
10	"Flint"	"49189"	"1"^^xsd:integer
11	"Flint"	"59485"	"1"^^xsd:integer
12	"Flint"	"21872"	"4"^^xsd:integer
13	"Flint"	"26873"	"4"^^xsd:integer
14	"Flint"	"48776"	"5"^^xsd:integer

Figure 32 - Query on the material using knowledge graphs

Our Palaeolithic researcher would be able to extract here the list of all reports that contain stone finds, and, if needed due to time constraints, also weigh which reports should be accessed first due to their quantities, or to decide how many artefacts from other stone materials are present. In the classification of the whole dataset, there were nine different types of stone artefacts. This would mean that the same search would demand the aggregation of nine individual queries.

Accessing the time ontology is a little different as it depends on how many periods are located between the initial and final markers included in the dataset. These variations can be extracted using the query in Fig 33. The results show the distance between the nodes and allow structuring queries able to extract the information desired. It is also possible to see how many of the records have at least one step skipped (678). It does not look much considering the number of total records (4235), however only 67% of them were filled (Table 2), and even in this case, many were already undetermined. This reduces the total of queryable records to 1704. This points to almost 40% of the records having at least one extra step that is not explicit in the data. The remaining records have subsequent periods as initial and final, or they are the same.

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?artefact ?report ?Bname ?Ename (count(?mid) as ?distance) {
    ?artefact adt:wasMadeFrom ?begin;
              adt:wasMadeUntil ?end;
              adt:isStoredAt ?EasyDb.
    ?EasyDb rdfs:label ?report.
    ?begin rdfs:label ?Bname.
    ?end rdfs:label ?Ename.

    ?end time:intervalMetBy* ?mid .
        filter(!sameTerm(?end, ?mid))
    ?mid time:intervalMetBy+ ?begin.
        filter(!sameTerm(?begin, ?mid))
}
group by ?artefact ?report ?Bname ?Ename
order by ?distance
```

Filter query results | Showing results from 1 to 678 of 678. Query took 0.2s, on 2022-07-22 at 11:37.

	artefact	report	Bname	Ename	distance
1	adt:1609	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
2	adt:1610	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
3	adt:1611	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
4	adt:1757	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
5	adt:1759	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
6	adt:1760	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
7	adt:1762	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
8	adt:1763	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
9	adt:1764	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
10	adt:1765	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
11	adt:1766	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
12	adt:1767	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
13	adt:1769	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
14	adt:1770	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer
15	adt:1771	"48482"	"Early Iron Age"	"Late Iron Age"	"1"^^xsd:integer

Figure 33 - Querying the number of steps between the beginning and final period for each artefact

The queries were then built using the number of steps skipped (Table 6) to retrieve the information. For instance, is shown in Fig 34 that query that creates variables for each of them and allows the filtering of the period required on all three steps. it is possible to record the missing periods using the variables ?timeX, so later they can be filtered.

Distance (in nodes)	Occurrence
1	324
2	46
3	7
4	89
5	2
6	6
7	13
9	29
10	155
13	7

Table 6 - Distance between Beginning and Ending periods in nodes.

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
select DISTINCT ?artefact ?report ?begin ?time2 ?time3 ?time4 ?end
where {
  ?artefact adt:wasMadeFrom ?begin;
            adt:wasMadeUntil ?end;
            adt:isStoredAt ?EasyDb.
  ?EasyDb rdfs:label ?report.

  ?time1 time:intervalMetBy ?begin.
  ?time2 time:intervalMetBy ?time1.
  ?time3 time:intervalMetBy ?time2.
  ?end time:intervalMetBy ?time3.
}
```

Filter query results

Showing results from 1 to 7 of 7. Query took 0.1s, moments ago.

	artefact	report	begin	time1	time2	time3	end
1	adt:171	"189862"	adt:MEV	adt:MEL	adt:NTA	adt:NTB	adt:NTC
2	adt:173	"189862"	adt:MEV	adt:MEL	adt:NTA	adt:NTB	adt:NTC
3	adt:4	"198193"	adt:MEL	adt:NTA	adt:NTB	adt:NTC	adt:REC
4	adt:125	"189862"	adt:MEL	adt:NTA	adt:NTB	adt:NTC	adt:REC
5	adt:2888	"21387"	adt:BTM	adt:BTL	adt:ITV	adt:ITM	adt:ITL
6	adt:2785	"21387"	adt:RTV	adt:RTM	adt:RTL	adt:MVR	adt:MEV
7	adt:2908	"21387"	adt:RTV	adt:RTM	adt:RTL	adt:MVR	adt:MEV

Figure 34 - Querying the skipped steps with distance = 3 nodes.

The combination of queries for each distance allows the assessment of the entire dataset. In our case study for the palaeolithic research, the comparison with the previous one using each subperiod of the Palaeolithic brings a slightly different result (Fig 35). Due to its length, the full code for the query is reproduced in the Appendix (Annex 3). Recapping from Fig 30, the Early Palaeolithic had recovered 6 reports, and it has the same result using knowledge graphs. However, the other two periods recovered not only the single report presented there but also the other six, as they were comprised in the timespan, but not explicitly recorded in the dataset.

1 Showing results from 1 to 6 of 6. Query took 0.2s, moments ago.

Filter query results

	report	quant
1	"21387"	"151"^^xsd:integer
2	"21653"	"19"^^xsd:integer
3	"26646"	"1"^^xsd:integer
4	"45911"	"7"^^xsd:integer
5	"48482"	"2"^^xsd:integer
6	"59182"	"1"^^xsd:integer

2 Showing results from 1 to 7 of 7. Query took 0.2s, moments ago.

Filter query results

	report	quant
1	"21387"	"151"^^xsd:integer
2	"21653"	"19"^^xsd:integer
3	"26646"	"1"^^xsd:integer
4	"36895"	"5"^^xsd:integer
5	"45911"	"7"^^xsd:integer
6	"48482"	"2"^^xsd:integer
7	"59182"	"1"^^xsd:integer

3 Showing results from 1 to 7 of 7. Query took 0.2s, moments ago.

Filter query results

	report	quant
1	"21387"	"151"^^xsd:integer
2	"21653"	"19"^^xsd:integer
3	"26646"	"1"^^xsd:integer
4	"40237"	"29"^^xsd:integer
5	"45911"	"7"^^xsd:integer
6	"48482"	"2"^^xsd:integer
7	"59182"	"1"^^xsd:integer

Figure 35 - Results using knowledge graphs: **1***. Early Palaeolithic;* **2***. Middle Palaeolithic;* **3***. Late Palaeolithic*

This effect can be more prominent in periods not located in the extremities of the timeline, as is the case for the Palaeolithic. Using the Neolithic as an example, The direct search returns five reports for the Early Neolithic, one report for the Middle Neolithic and seven reports for the Late Neolithic. The knowledge graph search of the same periods has brought as results the same ten reports for each one of the three subdivisions. What initially could seem a heterogeneous distribution end up revealing that at least one of the artefacts found in each of those ten reports could be dated from any of those subperiods.

The query also includes two reports whose artefacts have no explicit mention of the Neolithic at all, one of them has its artefacts placed between the Early Paleolithic and Late Bronze Age (26646), and the other (45911) from Early Palaeolithic and Late Iron Age. If such broad periodizations are due to the lack of specialist knowledge, the data structure would contribute even further to preventing them from being found by someone who could give them a more precise date.

Beyond querying the data subperiod, using knowledge graphs is also possible to use the time ontology setting to include the whole period in one query. Using the same code from Fig34 as a base, Fig 36 shows the changes needed for doing it. Lines 16 to 20 were added to retrieve the period related to each subdivision and assign it to a new variable (?perX). On the final query, those variables are used to filter the desired period.

To query the dataset about the reports with finds dated from the Palaeolithic retrieves a unified result with all eight reports in it (Fig 37). Due to its length, the full code for the query is reproduced in Appendix (Annex 4). It also gives the total number of artefacts dated on the desired period for each report. It is an extra layer of information about the possible relevance of each report for the research being done.

Now that is possible to retrieve information about the whole timeline of the artefacts, the combination of searches between them and the material can retrieve a complete picture of what is contained in the dataset.

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

select DISTINCT ?artefact ?perB ?per1 ?per2 ?per3 ?perE
where {
  ?artefact adt:wasMadeFrom ?begin;
            adt:wasMadeUntil ?end;
            adt:isStoredAt ?EasyDb.
  ?EasyDb rdfs:label ?report.
  ?time1 time:intervalMetBy ?begin.
  ?time2 time:intervalMetBy ?time1.
  ?time3 time:intervalMetBy ?time2.
  ?end time:intervalMetBy ?time3.

  ?begin time:intervalIn ?perB.
  ?time1 time:intervalIn ?per1.
  ?time2 time:intervalIn ?per2.
  ?time3 time:intervalIn ?per3.
  ?end time:intervalIn ?perE.
}
```

Filter query results

Showing results from 1 to 7 of 7. Query took 0.1s, minutes ago.

	artefact	perB	per1	per2	per3	perE
1	adt:4	adt:Middle_Ages	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Late_Modern_Period
2	adt:125	adt:Middle_Ages	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Late_Modern_Period
3	adt:171	adt:Middle_Ages	adt:Middle_Ages	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Early_Modern_Period
4	adt:173	adt:Middle_Ages	adt:Middle_Ages	adt:Early_Modern_Period	adt:Early_Modern_Period	adt:Early_Modern_Period
5	adt:2785	adt:Roman_Period	adt:Roman_Period	adt:Roman_Period	adt:Middle_Ages	adt:Middle_Ages
6	adt:2908	adt:Roman_Period	adt:Roman_Period	adt:Roman_Period	adt:Middle_Ages	adt:Middle_Ages
7	adt:2888	adt:Bronze_Age	adt:Bronze_Age	adt:Iron_Age	adt:Iron_Age	adt:Iron_Age

Figure 36 - Querying the periods from the skipped steps with distance = 3 nodes.

Fig 38 shows the results of the same search for Palaeolithic, but now including all material that is a type of stone. Due to decay, stones are usually the only material recovered from that period. The full code for the query is found in Appendix (Annex 5), but its main difference is the appending of a material query similar to the one found in fig 32 to it. It retrieves 215 artefacts, one undetermined type of stone artefact and 214 artefacts made of flint.

In other periods a broader array of materials is more common to be found, making the combination much more important. As an example, using the Iron Age, it is possible to find 224 artefacts, from those only one is made of metal (artefact 1609/ reported stored at dataset 48482)

Using semantic integration for these tasks makes it also easier to implement and maintain the whole system as these classes can be interrelated without having to change the basic data itself. The original data classifications can be kept as they were interpreted by the archaeologists in the field, but their relations and how they are recovered can be rearranged by adapting the underlying concepts. Instead of having to rework the tables every time a subclass or subperiod is created or changed, it allows a much more practical approach. A new element can be added to the current framework in a more "natural" way, as the new concepts are indeed added to the theoretical bases of the science itself.

Showing results from 1 to 8 of 8. Query took 0.2s, today at 11:16.

Filter query results

	report	quant
1	"21387"	"302"^^xsd:integer
2	"21653"	"38"^^xsd:integer
3	"26646"	"2"^^xsd:integer
4	"36895"	"5"^^xsd:integer
5	"40237"	"29"^^xsd:integer
6	"45911"	"14"^^xsd:integer
7	"48482"	"4"^^xsd:integer
8	"59182"	"2"^^xsd:integer

Figure 37 - Result using knowledge graphs for the whole Palaeolithic

⚠ Showing results from 1 to 215 of 215. Query
k 0.2s, today at 11:56.

Filter query results

	artefact	report	material
1	adt:714	"59182"	adt:SXX
2	adt:1541	"48482"	adt:SVU
3	adt:1593	"48482"	adt:SVU
4	adt:2231	"26646"	adt:SVU
5	adt:2232	"45911"	adt:SVU
6	adt:2233	"45911"	adt:SVU
7	adt:2234	"45911"	adt:SVU
8	adt:2236	"45911"	adt:SVU
9	adt:2237	"45911"	adt:SVU
10	adt:2239	"45911"	adt:SVU
11	adt:2240	"45911"	adt:SVU
12	adt:2365	"21653"	adt:SVU

Figure 38 - Result of material/period query (Stone/Palaeolithic)

Major reworks or paradigm shifts will still occur and they will always need to deal with adapting previous datasets, however, on a day-to-day basis, it presents itself as a more flexible and adaptable approach that can be used to integrate different data into the core dataset. If it was required to make changes in periodization or the further subdivision of the existing periods or for the archaeological cultures. For instance, recapping Fig. 2 periodization, breaking Middle Neolithic into two subdivisions, A and B. It could be inserted through its insertion into the structure of the ontology currently in use. Adjusting accordingly the queries would allow the integration of data with different granularities without the need for major reworks on the existing tables. The data classified as Middle Neolithic A would still appear in the queries about Neolithic, or Middle Neolithic, but keep its more detailed definition to be queried with the new data being classified in the future.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

The whole process described in this study, from acquiring the datasets to making them ready and available through the geoportal shows how important is to have reusability in mind when the data are being produced. Maybe in other scientific fields, the matter is dealt with better, or with a different approach, but at least in archaeology, in most cases, not enough attention is given to the future professional who will need to make sense of the information collected today. All attention is turned from the present moment to the past.

In centralized institutions, like universities, heritage- and research-centres the procedures for recording information are documented and mostly followed, which makes it easier to understand the reasoning behind the data.

Decentralized environments, as is the commercial archaeology in the Netherlands, have a harder time standardizing interpretation vocabularies and recording systems. Each entity has its own set of rules and standards derived from the legal requirements. Furthermore, the market competition, recognized even by governmental sources, has put a financial strain on the companies. This strain can reflect directly on the professionals and in the results of the fieldwork. The time for analysis becomes limited, as simultaneous projects being worked on are paramount to keep the companies afloat.

Geospatial tools have been used in many past studies not only to visualize spatially the data but also to perform several types of analysis. However, they are bound to the quality of the data available, not only the quality of the data at the time it was produced, but also the quality of how it is stored and shared. The adoption of FAIR principles and their derived guidelines are an important step in making the data produced today, still relevant for the future.

In Archaeology, data is a non-renewable resource, the information obtained from the excavation cannot be replicated in any other way. There are no alternatives, like ice cores or radiometric analysis that can provide the same information and can be collected somewhere else. Any information contained in excavated sites that are not properly recorded, or whose records were not well kept, is lost forever. In many places, there are still entire archaeological record collections solely on paper.

Paradoxically, the archaeological information more endangered are the loads of data that are digitized but kept in formats that fell into disuse, or the media in what they were recorded is no longer accessible due to hardware evolution. For those, this small study hopefully will serve as a reminder they can be recovered and that is possible to apply new ideas of how it can be processed and made available.

The data produced by the Dutch commercial archaeology sector is commonly stored in text files, as expected, but in the last few years, a trend in making it available in more easily machine-readable formats has become popular. Even though the number of different formats being used can pose a threat to future integration as technology advances.

The current study did not manage to extract all the information available in the datasets available. Part of it is due to a lack of specific knowledge of the nature of findings that are usually excavated in the Netherlands, local typologies and periodizations. As is pointed out by professionals and authorities, they also lack depth in the artefact description, which would contribute to better usability. But the discussion about the data falling short in information is only possible if the data properties are available for comparison. If these datasets can be explored by more professionals, a list of needs can be compiled. The companies cannot do it by themselves, because their clients are not willing to pay for this extra information, so the demand needs to come from someone else.

The use of knowledge graphs allows the data to be queried in different and more useful ways without having to manipulate the original data. The ontologies used in this study for classifying the data can still be improved, but in their current state already enhanced the results. Better ontologies demand a multidisciplinary approach to ensure their correctness and effectiveness. In any case, the results have shown how the knowledge graph format can allow queries to reach more results and enrich the exploration of the dataset content.

The mishaps found during the work also hinted at possible research avenues that could lead to solutions for some of them. The result of the interviews was meant to evaluate possible uses or analysis for the geospatial information because it lacked professionals interested in exploring these datasets for more than visualization. Perhaps due to the already big enough challenge the exploration of the data poses, and how the perceived lack of depth of the content of the dataset contributed to limiting their re-use.

The SPARQL queries become longer due to the need of following the chain of time periods not explicitly described in the datasets. The process to streamline the code would need a dedicated effort on finding better solutions especially if the steps to create a prototype from the proof-of-concept are meant to be pursued.

6.2. Recommendations and future work

The expansion of the current proof-of-concept geoportal into a fully working prototype could ensure a better exploration of those datasets. A first step would be to standardise the remaining descriptive fields of the datasets. This would demand contacting at least some of the archaeological companies to have a better understanding of what information those fields are meant to convey, and the coding system used to fill them. The work would allow a fine tunning in the searches, even though it would reach a smaller subset of the data as not all findings are classified in such detail.

The searches on the knowledge graphs will also need to be integrated into the geoportal future prototype, and the expanded options added to the interface. The presentation of the interface would also need a rework for attending to the needs of future users, which would need another round of interviews and user evaluations.

The ontologies proposed here needed to be expanded and integrated into the local chronology as presented in Fig. 2. Even if the granularity of the current data is not enough to take advantage of the whole classification, it could ensure their integration into other researchers' works. The effort would need a project on its own, as it would encompass integrating different areas and chronologies, but it would be crucial for any plans to expand the area coverage of the prototype.

The queries performed to the knowledge graphs also needed to be streamlined as now they are lengthy and it would grow even more if all the possible options (not only the ones present in the current dataset) are added. A better way to construct them needs to be sought, probably by contacting more experienced developers to combine efforts in seeking solutions.

Another expansion that could be foreseen would be the inclusion of not only the exploratory pits and trenches excavated digitized as polygons. These polygons can be extracted from almost every report and they can offer information about the soil's layer structure. They contain information about the depth and composition of each layer. It would demand a much bigger effort for the digitization as it would need to draw all areas that were excavated. But it could allow for extrapolating the similar layers through larger areas and modelling the past landscapes in their vicinity.

Another possible approach could offer the integrated dataset as Linked Data and allow its information to be connected with the different databases found during the research (Archis, Archeodepot etc). This integration into a single portal would contribute to making the research process more agile and complete. Even though, it would need to tackle questions of data licensing, authorization for accessing governmental databases, and also possible issues with the companies about the commercial leverage which could be gained if these data are released for public access.

The result of the sum of all these efforts could be a tool that presents all the knowledge integrated into a geographic information system. It would offer a living archaeological map that would allow assessment of the archaeological interest of the areas for the municipalities, construction companies and researchers and it would be continuously improved as new knowledge would be constantly added.

However, all initiatives come up against the need for standardization of the produced datasets. A better quality of the data output obtained from excavations is a prerequisite for the enhancement of their usage. The standards that could be used are already available, but they have not been consistently applied by the companies, at least in the sample used for this study.

The evaluation published about the archaeological sector (Raad voor Cultuur, 2022) complained about the municipalities not enforcing rules on the contractors, and therefore also not ensuring the quality of the data acquired. The municipalities' archaeology departments in turn do not have enough personnel or funds to perform the task, especially the smaller ones that are more prone to suffer pressure from big contractors. A solution proposed by them is the creation of a national archaeology research centre that could centralize these tasks and establish a fund for innovation in the sector.

The same report (Raad voor Cultuur, 2022) was vague about the role of this newly created institute in the improvement of data. An interesting first step could be a programme promoting the use of the naming standards already in existence and unifying the data processing outputs. It would greatly enhance the value of the data produced and influence its posterior usability.

LIST OF REFERENCES

Literary References

Aitchison, K., Rock-Macqueen, D., 2020. State of the Archaeological Market 2019. Federation of Archaeological Managers and Employers, London. https://www.archaeologists.net/sites/default/files/ State-of-the-Archaeological-Market-2019-1.pdf.

Allison, P., 2008. Dealing with Legacy Data - an introduction. Internet Archaeology 24. https://doi.org/10.11141/ia.24.8.

Ariadne Plus Project, 2021. Ariadne Portal - a data infrastructure serving the archaeological community worldwide. http://portal.ariadne-infrastructure.eu/.

Bakker, R.R., 1987. Knowledge Graphs: representation and structuring of scientific knowledge. Ph.D. thesis, University of Twente, Enschede, The Netherlands, ISBN 90-9001963-4, 1987.

Bergamaschi, S., Castano, S., Vincini, M., 1999. Semantic integration of semistructured and structured data sources. ACM SIGMOD Record. 28, pp.54-59. https://doi.org/10.1145/309844.309897.

Bergman, M., 2019. A Common Sense View of Knowledge Graphs. AI³ - Adaptive Information, Adaptive Innovation, Adaptive Infrastructure. July 1st 2019. https://www.mkbergman.com/2244/a-common-sense-view-of-knowledge-graphs/#kg23 (accessed 14 June 2021).

Binding, C., May, K., Tudhope, D., 2008. Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction via the CIDOC CRM. Research and Advanced Technology for Digital Libraries, vol. 5173 of Lecture Notes in Computer Science, pp. 280–290. Springer. http://dx.doi.org/10.1007/978-3-540-87599-4_30.

Binding, C., Charno, M., Jeffrey, S., May, K., & Tudhope, D., 2015. Template-Based Semantic Integration. International Journal on Semantic Web and Information Systems, 11(1), pp. 1–29. https://doi.org/10.4018/ijswis.2015010101.

Boschetti, A. Schimmelpfennig, W., Prado, E., 2019. From the “listing of antiquities” to the Geoportal. NIKE-Bulletin, 34(3), pp.16-21. https://doi.org/10.5167/uzh-213640.

Brandsen, A., Verberne, S., Lambers, K., Wansleeben, M., 2021. Usability Evaluation for Online Professional Search in the Dutch Archaeology Domain. https://arxiv.org/abs/2103.04437v1.

Calvanese, D., Liuzzo, P., Mosca, A., Remesal, J. Rezk, M., Rull, G., 2016. Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire. Engineering Applications of Artificial Intelligence, 51, pp. 212-229. https://doi.org/10.1016/j.engappai.2016.01.005.

Caraher, W., 2016. Slow Archaeology: Technology, Efficiency, and Archaeological Work. In: Averet, E.W.; Gordon, J. M. Counts, D. B. (Eds.), Mobilizing the Past for a Digital Future. Digital Press: Grand Forks/US, pp. 421-441.

Caraher, W., 2019. Slow Archaeology, Punk Archaeology, and the ‘Archaeology of Care’. European Journal of Archaeology 22 (3) 2019, pp. 372–385. https://doi.org/10.1017/eaa.2019.15.

Clarke, D., 1973. Archaeology: the loss of innocence. Antiquity 47(185), pp. 6–18.https://doi.org/ 10.1017/S0003598X0003461X.

Cultural Heritage Agency, 2022, Archeologie, https://www.cultureelerfgoed.nl/domeinen/archeologie (accessed 8 April 2022).

Demoule, J-P., 2012. Rescue Archaeology: A European View. Annual Review of Anthropology 41(1), pp. 611-626 https://doi.org/10.1146/annurev-anthro-092611-145854.

Djindjian, F., 2008. Webmapping in the historical and archaeological sciences. An introduction. Archeologia e Calcolatori, 19. pp. 9-16. http://eprints.bice.rm.cnr.it/id/eprint/15311.

Doerr, M., Bruseker, G., Bekiari, C., Ore, C. E., Velios, T., Stead, S., 2020. Definition of the CIDOC Conceptual Reference Model. Available online at: https://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v6.2.9%2030-4-2020%20.pdf (accessed 14 June 2022).

Dublin Core™ Metadata Initiative, 2022. A community of metadata professionals aimed to share practices, bridging established metadata practices and graph-based solutions for integration across data silos within a Linked Data ecosystem. https://www.dublincore.org/ (accessed 13 July 2022).

E-Depot on Dutch Archaeology, 2021. Repository to archive and make available archaeological research data. https://dans.knaw.nl/en/about/services/easy/e-depot-for-dutch-archaeology (accessed 13 June 2021).

Ehrlinger, L, Wöß, W, 2016. Joint Proceedings of the Posters and Demos Track of 12th International Conference on Semantic Systems - SEMANTiCS2016 and 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS16) at Leipzig, Germany. Volume: 1695, pp. 13-16. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1054.8298&rep= rep1&type=pdf (accessed 14 June 2021).

Eide, Ø., Felicetti, A., Ore, C.E., D'Andrea, A., Holmen, J., 2008. Encoding Cultural Heritage Information for the Semantic Web. EPOCH Conference on Open Digital Cultural Heritage Systems, Rome, Italy. http://public-repository.epoch-net.org/rome/05%20Procedures%20Data%20 Integration.pdf (accessed 10 July 2022).

Gagnon, M., 2007. Ontology-based integration of data sources. In 2007 10th International Conference on Information Fusion (pp. 1-8). IEEE July. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4408086.

Garstki, K., 2022. Challenges of a Critical Archaeology in the Modern World In: Garstki, K.(ed.) Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visiting Scholar's Conference. UCLA: Los Angeles, pp. 1-10. https://escholarship.org/uc/item/0vh9t9jq.

German Archaeological Institute (Deutsches Archäologisches Institut - DAI) (online). About iDAI.chronontology. https://chronontology. dainst.org/info/about (accessed 10 June 2022).

Guarino, N. 1998. Formal Ontology and Information Systems. Proceedings of the 1st International Conference on Formal Ontologies in Information Systems (FOIS 1998), pp. 3–15. https://doi.org/10.1006/ijhc.1995.1066.

Harrison, L. K., 2022. Closing the Loop on the Digital Data Lifecycle: Reviving a Salvage Archaeology Dataset. In: Garstki, K.(ed.) Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visiting Scholar's Conference. UCLA: Los Angeles, pp. 79-96. https://escholarship.org/uc/item/0vh9t9jq.

Hedges, M., Jordanous, A., Lawrence, K. F., Rouech, C., Tupman, C., 2017. Computer-Assisted Processing of Intertextuality in Ancient Languages. Journal of Data Mining & Digital Humanities, Numéro spécial sur le traitement assisté par ordinateur de l'intertextualité dans les langues anciennes. https://doi.org/10.46298/jdmdh.1375.

Heritage Act of 9 December 2015, Staatsblad 2015, 511. https://english.cultureelerfgoed.nl/publications/publications/2016/01/ 01/heritage-act-2016.

Huggett, J., 2015. A Manifesto for an Introspective Digital Archaeology. Open Archaeology 2015; 1, pp. 86–95. https://doi.org/10.1515/opar-2015-0002.

Huggett, J., 2022. Is Less More? Slow Data and Datafication in Archaeology. In: Garstki, K.(ed.) Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visiting Scholar's Conference. UCLA: Los Angeles, pp. 97-110, https://escholarship.org/uc/item/0vh9t9jq.

Innerebner, M., Costa, A., Chuprikova, E., Monsorno, R., Ventura, B., 2017. Organizing earth observation data inside a spatial data infrastructure. Earth Sci Inform 10, pp. 55–68. https://doi.org/10.1007/s12145-016-0276-0.

Isern, N., Fort J., Vander Linden M., 2012. Space Competition and Time Delays in Human Range Expansions. Application to the Neolithic Transition. PLoS ONE 7(12):e51106. https://doi.org/10.1371/journal.pone.0051106.

Jiang, H.; van Genderen, J.; Mazzetti, P.; Koo, H.; Chen, M., 2019. Current status and future directions of geoportals, International Journal of Digital Earth, https://doi.org/10.1080/17538947.2019.1603331.

Kadaster, (n.d.). Rijksdriehoeksstelsel. Available at https://www.kadaster.nl/zakelijk/registraties/basisregistraties/ rijksdriehoeksmeting/rijksdriehoeksstelsel (accessed 6 July 2022).

Kejriwal, M., 2019. What Is a Knowledge Graph? Springer Briefs Computer Science. 0, 1–7. https://doi.org/10.1007/978-3-030-12375-8_1.

Kletter, R., De-Groot, A., 2001. Excavating to Excess? Implications of the Last Decade of Archaeology in Israel. Journal of Mediterranean Archaeology 14(1), pp. 76–85. http://dx.doi.org/10.1558/jmea.v14i1.76.

Lauwerier, R.C.G.M, Eerden, M.C., Groenewoudt, B.J., Lascaris, M.A., Rensink, E., Smit, B.I., Speleers, B.P., van Doesburg, J., 2017. Knowledge for informed choices. Tools for decision makingin archaeological heritage management in the Netherlands, Rijksdienst voor het Cultureel Erfgoed/Cultural Heritage Agency of the Netherlands: Amersfoort 222p Available at: https://english.cultureelerfgoed.nl/binaries/cultureelerfgoed-en/documenten/publications/ 2017/01/01/nar-55-knowledge-for-informed-choices/NAR055_Knowledge_for_informed_choices.pdf (acessed 12 April 2022).

Lerma, J.L. Heras, V., Mora-Navarro, G., Rodas, P., Matute, F., 2020. Geoportal proposal for the inventory of cultural heritage in Nabón (Ecuador), ISPRS, pp. 1415-1418. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1415-2020.

McCoy, M. D., 2017. Geospatial Big Data and archaeology: Prospects and problems too great to ignore. Journal of Archaeological Science 84, pp. 74–94. https://doi.org/10.1016/j.jas.2017.06.003.

McCoy, M. D., Ladefoged, T. N., 2009. New developments in the use of spatial technology in archaeology. Journal of Archaeological Research 17(3), pp. 263–295. https://doi.org/10.1007/s10814-009-9030-1.

McKeague, P. et al, 2012. Developing a Spatial Data Infrastructure for Archaeological and Built Heritage. International Journal of Spatial Data Infrastructures Research, Vol.7, pp. 38-65. https://ijsdir.sadl.kuleuven.be/index.php/ijsdir/article/view/239.

Miguel-Castro, Ó. L., Fernández-Pareja, T., 2019. Hacia una Infraestructura de Datos Espaciales Arqueológica. Geoportal de la excavación arqueológica de Geoktchik-depe, Turkmenistán. ISIMU, vol 22, pp. 213-223. https://doi.org/10.15366/ isimu2019.22.012.

Meroño-Peñuela, A., Ashkpour, A., van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., van Harmelen, F., 2015. Semantic Technologies for Historical Research: A Survey. Semantic Web, 6(6), pp. 539-564. https://doi.org/10.3233/SW-140158.

Miller, E., 1998. An Introduction to the Resource Description Framework. Bul. Am. Soc. Info. Sci. Tech., 25, pp. 15-19. https://doi.org/10.1002/bult.105.

Mountantonakis, M., Tzitzikas, Y., 2019. Large-scale Semantic Integration of Linked Data: A Survey. ACM Comput. Surv. 52, 5, Article 103 (September 2019). https://doi.org/10.1145/3345551

Ontotext, n.d. What are Ontologies? Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies/ (accessed 12 June 2022).

Open Geospatial Consortium (OGC), 2006. OpenGIS® Web Map Server Implementation Specification. http://www.opengeospatial.org/standards/wms (accessed 6 July 2022)

Open Geospatial Consortium (OGC), 2014. OpenGIS® OpenGIS Web Feature Service 2.0 Interface Standard – With

Corrigendum. http://docs.opengeospatial.org/is/09-025r2/09-025r2.html (accessed 6 July 2022)
OpenLayers, 2019. OpenLayers Version 6 Main Page. http://openlayers.org.

PARTHENOS et al., 2018. PARTHENOS Guidelines to FAIRify data management and make data reusable. Zenodo. https://doi.org/10.5281/zenodo.2668479.

Pearson, E., 2019. Commercial Environmental Archaeology: are we back in the dark ages or is environmental archaeology a potential agent of change? Internet Archaeology 53. https://doi.org/10.11141/ia.53.4.

Prinz, T et al, 2014. GeoArchaeology Web 2.0: Geospatial Information Services Facilitate New Concepts of Web-Based Data Visualization Strategies in Archaeology—Two Case Studies from Surveys in Sudan (Wadi) and Turkey (Doliche). Archaeological Discovery, 2, pp. 91-106. https://doi.org/10.4236/ad.2014.24011.

Raad voor Cultuur, 2022. Archeologie bij de tijd. Aanbevelingen voor aanpassingen. Raad voor Cultuur: Den Haag. https://www.raadvoorcultuur.nl/binaries/raadvoorcultuur/documenten/adviezen/2022/02/10/archeologie-bij-de-tijd/ Archeologie+bij+de+tijd.pdf (acessed 12 April 2022).

Richards, J. D., 1997. Preservation and re-use of digital data: the role of the Archaeology Data Service, Antiquity 71(274), pp. 1057–1059. https://doi.org/10.1017/S0003598X00086014.

van de Riet, R., Meersman, R., 1992. Knowledge Graphs, in: Linguistic Instruments in Knowledge Engineering: Proceedings of the 1991 Workshop on Linguistic Instruments in Knowledge Engineering, Tilburg, the Netherlands, 17-18 January 1991, North-Holland, pp. 97-117.

Ronzhin, S., Folmer, E.; Maria, P.; Brattinga, M.; Beek, W.; Lemmens, R.; van't Veer, R., 2019. Kadaster Knowledge Graph: Beyond the Fifth Star of Open Data. Information, 10, 310. https://doi.org/10.3390/info10100310.

Ronzino, P., Acconcia, V., Falcone, A., 2018. Towards the integration of spatial data through the Italian Geoportal for Archaeological Resources. 3rd Digital Heritage International Congress (DigitalHERITAGE) with 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), 2018, pp. 1-5, http://dx.doi.org/10.1109/DigitalHeritage.2018.8810132.

Sinclair, P. Addis, M., Choi, F., Doerr, M., Lewis, P., Martinez, K., 2006. The use of CRM Core in Multimedia Annotation. In Proceedings of First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006), part of the 15th World Wide Web Conference, Edinburgh, Scotland. https://eprints.soton.ac.uk/262828/.

Stenborg, P., 2018. On the potentials and limitations of digital mediation of archaeological information. In: Huvila, I. (Ed.), Archaeology and Archaeological Information in the Digital Society. Routledge: London. https://doi.org/ 10.4324/9781315225272.

SIKB, 2007. Hoe stel ik eisen aan archeologisch vooronderzoek? SIKB: Gouda, 12p. https://www.sikb.nl/doc/archeo/ Brochure%20Hoe%20stel%20ik%20eisen%20aan%20archeologisch%20vooronderzoek.pdf (accessed 6 April 2022).

SIKB, 2017. Domeintabellen IM SIKB0102. SIKB: Gouda, 271p. https://www.sikb.nl/doc/sikb0102/SIKB0102%20versie% 204.2.0%20domeintabellen.pdf (acessed 19 April 2022).

SIKB – BRL 4000, 2018. Beoordelingsrichtlijn Archeologie. SIKB: Gouda, 97p. https://www.sikb.nl/doc/BRL4000/Protocol%20 4002%20 Bureauonderzoek%204_1.pdf (accessed 8 April 2022).

SIKB – Protocol 4002, 2018. Bureauonderzoek. SIKB: Gouda, 25p. https://www.sikb.nl/doc/BRL4000/Protocol%204002%20 Bureauonderzoek%204_1.pdf (accessed 6 April 2022).

SIKB – Protocol 4003, 2018. Inventariserend Veldonderzoek (landbodems). SIKB: Gouda, 40p https://www.sikb.nl/doc/ BRL4000/Protocol%204003%20Inventariserend%20VeldOnderzoek-lb%204_1.pdf (accessed 6 April 2022).

SIKB – Protocol 4004, 2018. Opgraven (landbodems). SIKB: Gouda, 92p https://www.sikb.nl/doc/BRL4000/ Protocol%204004%20Opgraven-lb%204_1.pdf (accessed 6 April 2022).

Singhal, A., 2012. Introducing the knowledge graph: things, not strings, Official Google Blog, May - 2012. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html (accessed 13 June 2022).

Stokman, F.N. & Vries, P.H. de, 1988. Structuring Knowledge in a Graph. In: Human-Computer Interaction, Springer Science + Business Media, pp. 186–206. http://dx.doi.org/10.1007/978-3-642-73402-1_12.

Thelemann, M., Bebermeier, W., Hoelzmann, P., Lehnhardt, E., 2017. Bog iron ore as a resource for prehistoric iron production in Central Europe – a case study of the Widawa catchment area in eastern Silesia, Poland Catena, 149 (1) (2017), pp. 474-490 https://doi.org/ 10.1016/j.catena.2016.04.002.

UN General Assembly, 2015. Transforming our world: the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/1. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1& Lang=E (accessed 20 June 2021)

Uzdanaviciute, V., & Butleris, R., 2011. Ontology-based foundations for data integration. In BUSTECH The First International Conference on Business Intelligence and Technology, pp. 34-39). http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.467.3399&rep=rep1&type=pdf

Varnienė-Janssen, R., Kuprienė, J., 2021. Linked Open Data as a Strategic Orientation of Lithuanian Libraries. Relevant Tomorrow, 1(19), pp. 14–34. https://doi.org/10.51740/RT.1.19.2.

Vockner, B., Richter, A., Mittlböck, M., 2013. From Geoportals to Geographic Knowledge Portals. ISPRS International Journal of Geo-Information 2, no. 2, pp. 256-275. https://doi.org/10.3390/ijgi2020256.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al, 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

World Wide Web Consortium, 2008. SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query/ (accessed on 24 June 2022).

World Wide Web Consortium, 2020. RDF Schema 1.1. https://www.w3.org/TR/rdf-schema/ (accessed on 24 June 2022).

World Wide Web Consortium, 2020. Time Ontology in OWL. https://www.w3.org/TR/2020/CR-owl-time-20200326/ (accessed on 24 June 2022).

Yan, J., Wang, C., Cheng, W., Gao, M., Zhou, A., 2018. A retrospective of knowledge graphs. Front. Comput. Sci., 2018, 12(1), pp. 55–74. https://doi.org/10.1007/s11704-016-5228-9

Zhang, L., 2002. Knowledge graph theory and structural parsing. Twente University Press (TUP), Enschede. https://research. utwente.nl/en/publications/knowledge-graph-theory-and-structural-parsing.

Dataset References

ADC ArcheoProjecten, Huisman, N., 2011. Enschede Usseler-es fase 3, 2a en 2b Begeleiding. DANS Data Station Archaeology, V1. https://doi.org/10.17026/ans-xj7-3fzn.

BAAC BV., Bergman, W.A., Buesink, A., 2008. Enschede, Plangebied Velve-Lindenhof. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xc4-28p.

BAAC B.V., Bergman, W.A., Mooren, J., 2008. Enschede, Plangebied Velve-Lindenhof. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xkw-b6hu.

Benerink, G.M.H. , 2016. Inventariserend Veldonderzoek door middel van proefsleuven 'Plangebied Helmerstraat 310', Enschede, Gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zgq-cec3.

Bergman, W.A., Krist, J.S., 2008. Enschede Laares. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-x3w-s9jr

Bot, M.C.J., 2014. Enschede Beckum Ganzebos Begeleiding. DANS Data Station Archaeology, V2. https://doi.org/10.17026/dans-xjx-yqdc.

Brouwer, E.W., Luytelaar, P., van; Hesseling, I., 2012. IVO-P Zuidkamp te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xe4-94sh.

Diependaal, S., 2013. Archeologische begeleiding de Klomp te Enschede gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xyn-xg44.

Diependaal, S., Hagens, D., Koeman, S.M., Hesseling, I., 2009. Een archeologische begeleiding onder protocol opgraven, Walstraat 45-49 te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xgs-465s.

Dijkstra, J., 2002. Aanvullend archeologisch onderzoek op de Josink Es te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-2zg-vswv.

Eeltink, N.T.D., 2006. Enschede Plangebied Cascade. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zds-rwb5.

Fijma, P., Boon, H., Osinga, M., 2010. Archeologisch onderzoek Kotmanpark Oost te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xv8-g7xf.

Groot, B. de, 2005. Enschede, plangebied Cascade Inventariserend archeologisch veldonderzoek: karterende fase. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zzz-g25x.

Hebinck, K.A., Verboom-Jansen, M., 2011. Een waarderend inventariserend veldonderzoek door middel van boringen voor project Belevingsplek Laag van Usselo, gemeente Enschede (Ov). DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-x39-gv45.

Jans, J.E.A., Hielkema, J.B., 2015. Sporen uit IJzertijd - Nieuwe tijd op de Usseler Es nabij Usselo, gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-x7q-k9my.

Kalisvaart, C.C. 2012. Enschede, Auke vleerstraat Fase 2a Enschede, Plangebied Auke Vleerstraat Fase 2a. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zcz-6y6f.

Kemme, W.; Krist, J.S.; BAAC BV, 2009. Enschede Laares. Inventariserend veldonderzoek d.m.v. proefsleuven aangevuld met megaboringen, waarderende fase (IVO-P en IVO-O). DANS Data Station Archaeology, V2. https://doi.org/10.17026/dans-zzz-kjsm.

Kenemans, M.C., 2005. Enschede, Usseler Es. DANS Data Station Archaeology, V2. https://doi.org/10.17026/dans-xwc-rkvv.

Mousch, R.G. van, 2007. Enschede Plangebied Cascade II. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-x59-rc9m.

Oude Rengerink, J.A.M., 2018. Opgraving Enschede Zuidkamp Opgraving Erve Overmaat op de Zuidkamp, gemeente Enschede (OV). DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xvv-tjjr.

Oude Rengerink, J.A.M., 2019. Archeologische begeleiding Oldenzaalsestraat te Enschede, gemeente Enschede (OV) Archeologische begeleiding Oldenzaalsestraat te Enschede, gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-z88-bkqy.

Oude Rengerink, J.A.M., 2019. IVO-P Heetkampweg Zuidkampweg IVO-P Heetkampweg Zuidkampweg. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zrq-dgx8.

Oude Rengerink, J.A.M., 2020. Archeologische begeleiding Van Lochemstraat te Enschede, gemeente Enschede (OV) Archeologische begeleiding Van Lochemstraat te Enschede, gemeente Enschede (OV). DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zvd-b7mj.

Pronk, E.C., 2006. Plangebied Welmanweg 20 te Lonneker, gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-28r-wgyg.

Putten, M.J. van, 2006. Plangebied Cascade, Sleutelweg te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zm6-nmer.

Schabbink, M., 2008. t Vraneker vindplaats 5. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zhj-ehf7.

Schabbink, M., 2010. Opgraving Erve Sprakel: een historische boerderij. Plangebied ’t Vaneker, gemeente Enschede. DANS Data Station Archaeology, V2. https://doi.org/10.17026/dans-x2n-r2yu.

Schorn, E.A., 2004. Plangebied Usseleres-Noord (gemeente Enschede) Inventariserend archeologisch veldonderzoek : karterende fase. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xf2-epcd.

Spitzers, T.A., Wemerman, P.J.L., 2008. Klokkenplas, Gemeente Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-zvs-cd76.

Veldhuis, J.R. , 2011. Kampbewoners uit het Mesolithicum: een aanvullend IVO-P, op de locatie van het plangebied Cascade (vindplaats IV) te Enschede (O). DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xp5-68sf.

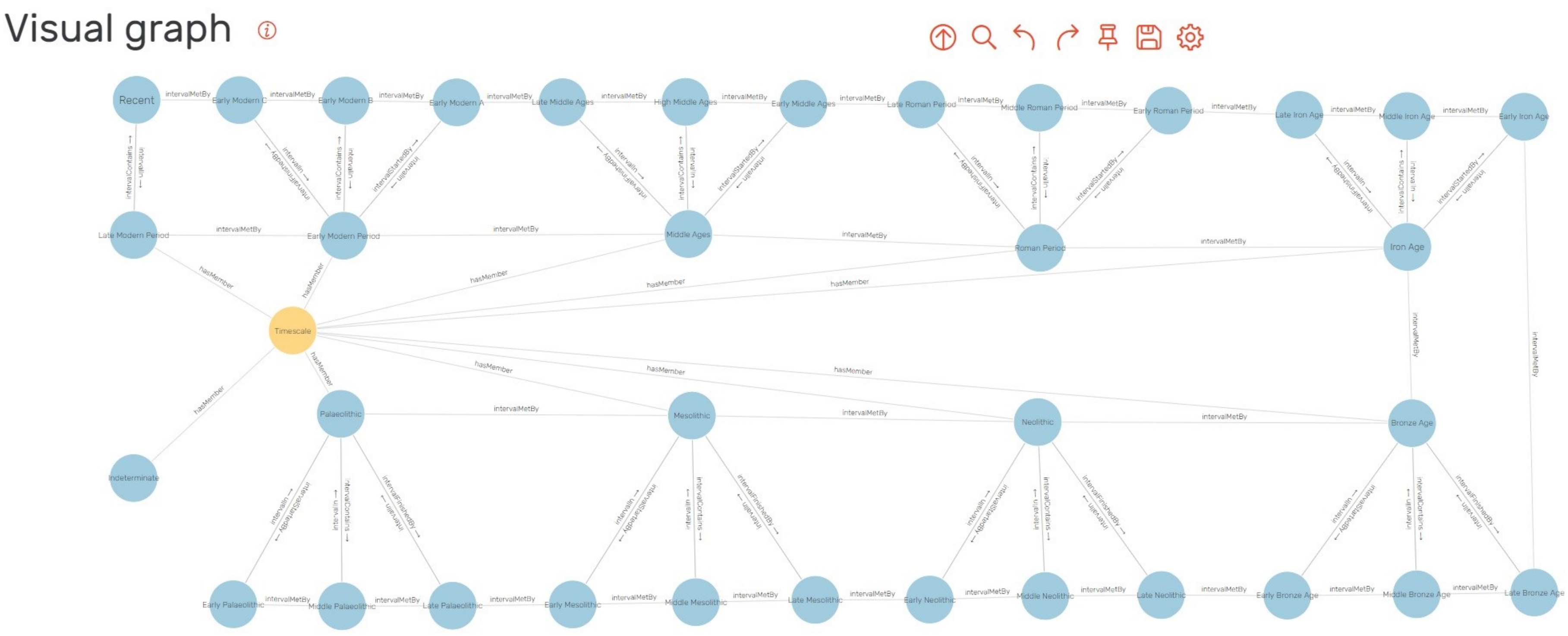
Weerden, J.F. van der; Wemerman, P.J.L., 2008. Enschede Usseler Es, Aanvullend Archeologisch Onderzoek d.m.v. proefsleuven. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-x48-trwd.

Wemerman, P.J.L.; Koeman, S.M., 2009. Auke Vleerstraat te Enschede. DANS Data Station Archaeology, V1. https://doi.org/10.17026/dans-xty-k6fw.

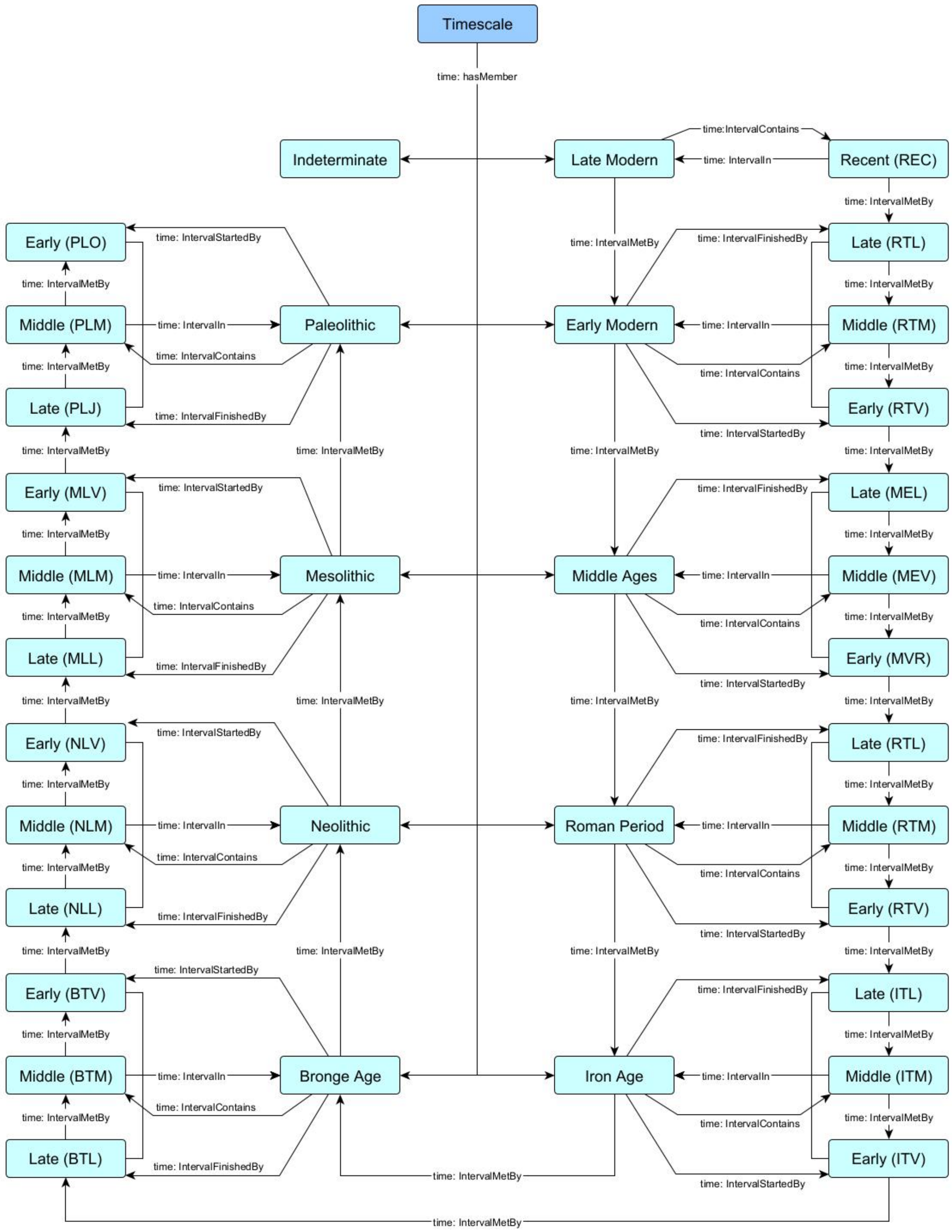
APPENDIX

Annex 1: Visual representation of the ontology for the chronology used in the study.

Visualization obtained in GraphDB

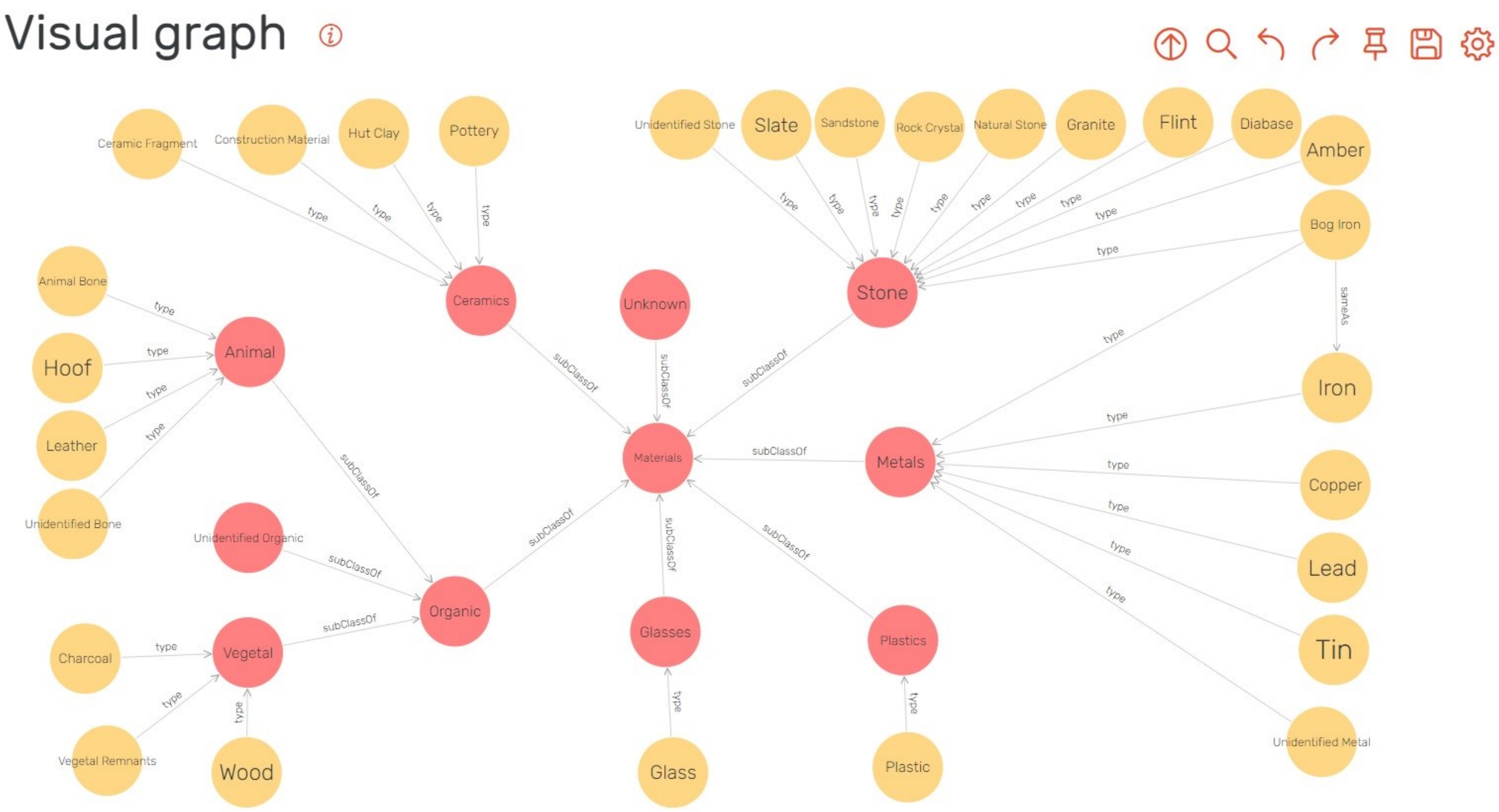


Schematic visualization for better reading

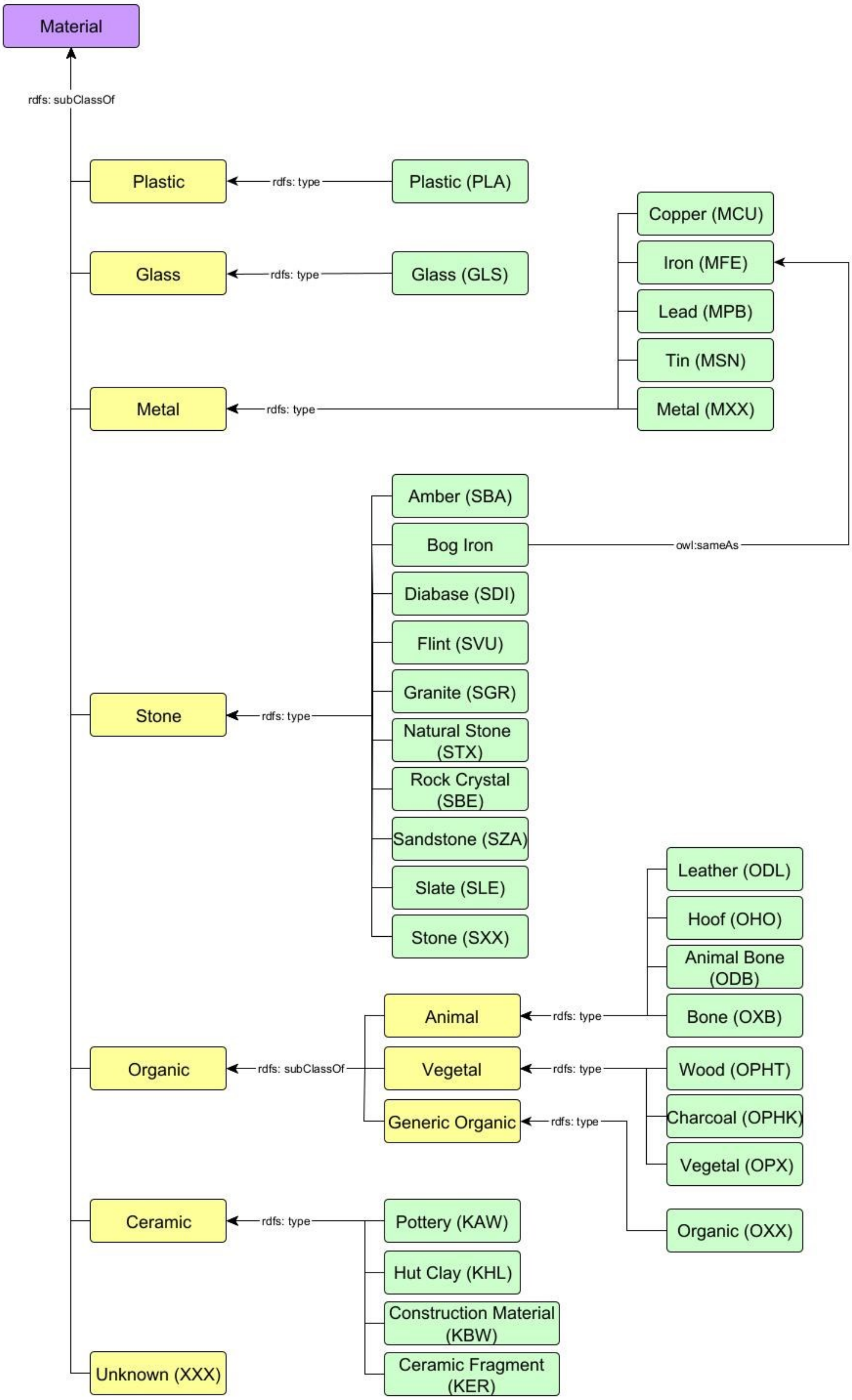


Annex 2: Visual representation of the ontology for the materials used in the study.

Visualization obtained in GraphDB



Schematic visualization for better reading



Annex 3: SPARQL code for querying for a single period's subdivision.

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?report (count(?artefact) as ?quant) WHERE
{
  {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:isStoredAt ?EasyDb;
                      adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end.
                      ?EasyDb rdfs:label ?report.

            ?begin time:intervalIn ?perB.
                      ?end time:intervalIn ?perE

            BIND(adt:PLM as ?search).
                      FILTER (?begin = ?search || ?end = ?search).
            }
    }
UNION
  {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.

                      ?EasyDb rdfs:label ?report.

                      ?time2 time:intervalMetBy ?begin.
                      ?end time:intervalMetBy ?time2.

                      BIND(adt:PLM as ?search).
                      FILTER (?time2 = ?search).
            }
  }
UNION
    {
     select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

            ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?end time:intervalMetBy ?time3.

            BIND(adt:PLM as ?search).
                      FILTER (?time2 = ?search||?time3 = ?search).

            }
    }
UNION
    {
      select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

                      ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?time4 time:intervalMetBy ?time3.
```

```
                        ?end time:intervalMetBy ?time4.

              BIND(adt:PLM as ?search).
                        FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search).

              }
       }
UNION
       {
          select DISTINCT ?artefact ?report where {
              ?artefact adt:wasMadeFrom ?begin;
                        adt:wasMadeUntil ?end;
                        adt:isStoredAt ?EasyDb.
                        ?EasyDb rdfs:label ?report.

                        ?time2 time:intervalMetBy ?begin.
                        ?time3 time:intervalMetBy ?time2.
                        ?time4 time:intervalMetBy ?time3.
                        ?time5 time:intervalMetBy ?time4.
                        ?end time:intervalMetBy ?time5.

                        BIND(adt:PLM as ?search).
                        FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search).

              }
       }
UNION
       {
          select DISTINCT ?artefact ?report where {
              ?artefact adt:wasMadeFrom ?begin;
                        adt:wasMadeUntil ?end;
                        adt:isStoredAt ?EasyDb.
                        ?EasyDb rdfs:label ?report.

                        ?time2 time:intervalMetBy ?begin.
                                  ?time3 time:intervalMetBy ?time2.
                                  ?time4 time:intervalMetBy ?time3.
                                  ?time5 time:intervalMetBy ?time4.
                                  ?time6 time:intervalMetBy ?time5.
                                  ?end time:intervalMetBy ?time6.

                BIND(adt:PLM as ?search).
                                  FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search).

                        }
       }
 UNION
       {
          select DISTINCT ?artefact ?report where {
              ?artefact adt:wasMadeFrom ?begin;
                        adt:wasMadeUntil ?end;
                        adt:isStoredAt ?EasyDb.
                                  ?EasyDb rdfs:label ?report.

                                  ?time2 time:intervalMetBy ?begin.
                                  ?time3 time:intervalMetBy ?time2.
                                  ?time4 time:intervalMetBy ?time3.
                                  ?time5 time:intervalMetBy ?time4.
                                  ?time6 time:intervalMetBy ?time5.
                                  ?time7 time:intervalMetBy ?time6.
                                  ?end time:intervalMetBy ?time7.

                        BIND(adt:PLM as ?search).
                        FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search||?time7 = ?search).

```

```
                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                             ?EasyDb rdfs:label ?report.

                             ?time2 time:intervalMetBy ?begin.
                             ?time3 time:intervalMetBy ?time2.
                             ?time4 time:intervalMetBy ?time3.
                             ?time5 time:intervalMetBy ?time4.
                             ?time6 time:intervalMetBy ?time5.
                             ?time7 time:intervalMetBy ?time6.
                             ?time8 time:intervalMetBy ?time7.
                             ?end time:intervalMetBy ?time8.

                             BIND(adt:PLM as ?search).
                             FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search||?time7 = ?search||?time8 = ?search).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                             ?EasyDb rdfs:label ?report.

                             ?time2 time:intervalMetBy ?begin.
                             ?time3 time:intervalMetBy ?time2.
                             ?time4 time:intervalMetBy ?time3.
                             ?time5 time:intervalMetBy ?time4.
                             ?time6 time:intervalMetBy ?time5.
                             ?time7 time:intervalMetBy ?time6.
                             ?time8 time:intervalMetBy ?time7.
                             ?time9 time:intervalMetBy ?time8.
                             ?time10 time:intervalMetBy ?time9.
                             ?end time:intervalMetBy ?time10.

                             BIND(adt:PLM as ?search).
                             FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search||?time7 = ?search||?time8 = ?search||?time9 = ?search||?time10 = ?search).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                             ?EasyDb rdfs:label ?report.

                             ?time2 time:intervalMetBy ?begin.
                             ?time3 time:intervalMetBy ?time2.
                             ?time4 time:intervalMetBy ?time3.
                             ?time5 time:intervalMetBy ?time4.
                             ?time6 time:intervalMetBy ?time5.
                             ?time7 time:intervalMetBy ?time6.
                             ?time8 time:intervalMetBy ?time7.
                             ?time9 time:intervalMetBy ?time8.
                             ?time10 time:intervalMetBy ?time9.
```

```
                              ?time11 time:intervalMetBy ?time10.
                              ?end time:intervalMetBy ?time11.

                              BIND(adt:PLM as ?search).
                              FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search||?time7 = ?search||?time8 = ?search||?time9 = ?search||?time10 = ?search||?time11 = ?search).

                    }
        }
UNION
        {
            select DISTINCT ?artefact ?report where {
                ?artefact adt:wasMadeFrom ?begin;
                          adt:wasMadeUntil ?end;
                          adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?time9 time:intervalMetBy ?time8.
                              ?time10 time:intervalMetBy ?time9.
                              ?time11 time:intervalMetBy ?time10.
                              ?time12 time:intervalMetBy ?time11.
                              ?time13 time:intervalMetBy ?time12.
                              ?time14 time:intervalMetBy ?time13.
                              ?end time:intervalMetBy ?time14.

                              BIND(adt:PLM as ?search).
                              FILTER (?time2 = ?search||?time3 = ?search||?time4 = ?search||?time5 = ?search||?time6 = ?search||?time7 = ?search||?time8 = ?search||?time9 = ?search||?time10 = ?search||?time11 = ?search||?time12 = ?search||?time13 = ?search||?time14 = ?search).

                    }
        }
}
Group by ?report
```

Annex 4: SPARQL code for querying the whole period.

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?report (count(?artefact) as ?quant) WHERE
{
  {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:isStoredAt ?EasyDb;
                      adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end.
                      ?EasyDb rdfs:label ?report.

            ?begin time:intervalIn ?perB.
                      ?end time:intervalIn ?perE

                      BIND(adt:Paleolithic as ?persearch).
                      FILTER (?perB = ?persearch || ?perE = ?persearch).
            }
    }
UNION
  {
        select DISTINCT ?artefact ?report where {
          ?artefact adt:wasMadeFrom ?begin;
                    adt:wasMadeUntil ?end;
                    adt:isStoredAt ?EasyDb.
  ?EasyDb rdfs:label ?report.

  ?time2 time:intervalMetBy ?begin.
  ?end time:intervalMetBy ?time2.

  ?time2 time:intervalIn ?per2.

BIND(adt:Paleolithic as ?persearch).
FILTER (?per2 = ?persearch).

            }
  }
    UNION
    {
     select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

            ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?end time:intervalMetBy ?time3.

                                ?time2 time:intervalIn ?per2.
                                ?time3 time:intervalIn ?per3.


                      BIND(adt:Paleolithic as ?persearch).
                      FILTER (?per2 = ?persearch || ?per3 = ?persearch).
            }
    }
UNION
    {
      select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.
```

```

                        ?time2 time:intervalMetBy ?begin.
                        ?time3 time:intervalMetBy ?time2.
                        ?time4 time:intervalMetBy ?time3.
                        ?end time:intervalMetBy ?time4.

                                  ?time2 time:intervalIn ?per2.
                                  ?time3 time:intervalIn ?per3.
                                  ?time4 time:intervalIn ?per4.

                        BIND(adt:Paleolithic as ?persearch).
                        FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch).
              }
     }
UNION
     {
         select DISTINCT ?artefact ?report where {
              ?artefact adt:wasMadeFrom ?begin;
                        adt:wasMadeUntil ?end;
                        adt:isStoredAt ?EasyDb.
                        ?EasyDb rdfs:label ?report.

                        ?time2 time:intervalMetBy ?begin.
                        ?time3 time:intervalMetBy ?time2.
                        ?time4 time:intervalMetBy ?time3.
                        ?time5 time:intervalMetBy ?time4.
                        ?end time:intervalMetBy ?time5.

                                  ?time2 time:intervalIn ?per2.
                                  ?time3 time:intervalIn ?per3.
                                  ?time4 time:intervalIn ?per4.
                                  ?time5 time:intervalIn ?per5.

                        BIND(adt:Paleolithic as ?persearch).
                                  FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch).

                        }
     }
 UNION
     {
         select DISTINCT ?artefact ?report where {
              ?artefact adt:wasMadeFrom ?begin;
                        adt:wasMadeUntil ?end;
                        adt:isStoredAt ?EasyDb.
                                  ?EasyDb rdfs:label ?report.

                                  ?time2 time:intervalMetBy ?begin.
                                  ?time3 time:intervalMetBy ?time2.
                                  ?time4 time:intervalMetBy ?time3.
                                  ?time5 time:intervalMetBy ?time4.
                                  ?time6 time:intervalMetBy ?time5.
                                  ?end time:intervalMetBy ?time6.

                                            ?time2 time:intervalIn ?per2.
                                            ?time3 time:intervalIn ?per3.
                                            ?time4 time:intervalIn ?per4.
                                            ?time5 time:intervalIn ?per5.
                                            ?time6 time:intervalIn ?per6.

                                  BIND(adt:Paleolithic as ?persearch).
                                  FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch).
                        }
     }
 UNION
     {
```

```
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?end time:intervalMetBy ?time7.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?end time:intervalMetBy ?time8.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
```

```
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?time9 time:intervalMetBy ?time8.
                              ?time10 time:intervalMetBy ?time9.
                              ?end time:intervalMetBy ?time10.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.
                                        ?time9 time:intervalIn ?per9.
                                        ?time10 time:intervalIn ?per10.
                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch|| ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch).
                     }
        }
UNION
        {
            select DISTINCT ?artefact ?report where {
                ?artefact adt:wasMadeFrom ?begin;
                          adt:wasMadeUntil ?end;
                          adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?time9 time:intervalMetBy ?time8.
                              ?time10 time:intervalMetBy ?time9.
                              ?time11 time:intervalMetBy ?time10.
                              ?end time:intervalMetBy ?time11.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.
                                        ?time9 time:intervalIn ?per9.
                                        ?time10 time:intervalIn ?per10.
                                        ?time11 time:intervalIn ?per11.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch || ?per11 = ?persearch).

                     }
        }
UNION
        {
            select DISTINCT ?artefact ?report where {
                ?artefact adt:wasMadeFrom ?begin;
                          adt:wasMadeUntil ?end;
```

```
                        adt:isStoredAt ?EasyDb.
                                ?EasyDb rdfs:label ?report.

                                ?time2 time:intervalMetBy ?begin.
                                ?time3 time:intervalMetBy ?time2.
                                ?time4 time:intervalMetBy ?time3.
                                ?time5 time:intervalMetBy ?time4.
                                ?time6 time:intervalMetBy ?time5.
                                ?time7 time:intervalMetBy ?time6.
                                ?time8 time:intervalMetBy ?time7.
                                ?time9 time:intervalMetBy ?time8.
                                ?time10 time:intervalMetBy ?time9.
                                ?time11 time:intervalMetBy ?time10.
                                ?time12 time:intervalMetBy ?time11.
                                ?time13 time:intervalMetBy ?time12.
                                ?time14 time:intervalMetBy ?time13.
                                ?end time:intervalMetBy ?time14.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.
                                        ?time9 time:intervalIn ?per9.
                                        ?time10 time:intervalIn ?per10.
                                        ?time11 time:intervalIn ?per11.
                                        ?time12 time:intervalIn ?per12.
                                        ?time13 time:intervalIn ?per13.
                                        ?time14 time:intervalIn ?per14.
                                        ?end time:intervalIn ?perE

                                BIND(adt:Paleolithic as ?persearch).
                                FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch || ?per11 = ?persearch || ?per12 = ?persearch || ?per13 = ?persearch || ?per14 = ?persearch).
                        }
        }
}
Group by ?report
order by ?quant
```

Annex 5: SPARQL code for querying the combination of material and period

```
PREFIX adt: <http://example.com/base/ArchaeoData>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?artefact ?report ?material WHERE
{
  ?artefact adt:isMadeOf ?material.
    ?material rdfs:type adt:Stone.

    {
  select DISTINCT ?artefact ?report where {
            ?artefact adt:isStoredAt ?EasyDb;
                      adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end.
                      ?EasyDb rdfs:label ?report.

            ?begin time:intervalIn ?perB.
                      ?end time:intervalIn ?perE

                      BIND(adt:Paleolithic as ?persearch).
                      FILTER (?perB = ?persearch || ?perE = ?persearch).
            }
    }
UNION
  {
        select DISTINCT ?artefact ?report where {
          ?artefact adt:wasMadeFrom ?begin;
                    adt:wasMadeUntil ?end;
                    adt:isStoredAt ?EasyDb.
  ?EasyDb rdfs:label ?report.

  ?time2 time:intervalMetBy ?begin.
  ?end time:intervalMetBy ?time2.

  ?time2 time:intervalIn ?per2.

BIND(adt:Paleolithic as ?persearch).
FILTER (?per2 = ?persearch).

            }
  }
    UNION
    {
     select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

            ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?end time:intervalMetBy ?time3.

                                ?time2 time:intervalIn ?per2.
                                ?time3 time:intervalIn ?per3.


                      BIND(adt:Paleolithic as ?persearch).
                      FILTER (?per2 = ?persearch || ?per3 = ?persearch).
            }
    }
UNION
    {
      select DISTINCT ?artefact ?report where {
```

```
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

                      ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?time4 time:intervalMetBy ?time3.
                      ?end time:intervalMetBy ?time4.

                              ?time2 time:intervalIn ?per2.
                              ?time3 time:intervalIn ?per3.
                              ?time4 time:intervalIn ?per4.

                      BIND(adt:Paleolithic as ?persearch).
                      FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch).
            }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                      ?EasyDb rdfs:label ?report.

                      ?time2 time:intervalMetBy ?begin.
                      ?time3 time:intervalMetBy ?time2.
                      ?time4 time:intervalMetBy ?time3.
                      ?time5 time:intervalMetBy ?time4.
                      ?end time:intervalMetBy ?time5.

                              ?time2 time:intervalIn ?per2.
                              ?time3 time:intervalIn ?per3.
                              ?time4 time:intervalIn ?per4.
                                      ?time5 time:intervalIn ?per5.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch).

                      }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?end time:intervalMetBy ?time6.

                                      ?time2 time:intervalIn ?per2.
                                      ?time3 time:intervalIn ?per3.
                                      ?time4 time:intervalIn ?per4.
                                      ?time5 time:intervalIn ?per5.
                                      ?time6 time:intervalIn ?per6.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch).
```

```
                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?end time:intervalMetBy ?time7.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?end time:intervalMetBy ?time8.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch).

                    }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
```

```
                        adt:isStoredAt ?EasyDb.
                                ?EasyDb rdfs:label ?report.

                                ?time2 time:intervalMetBy ?begin.
                                ?time3 time:intervalMetBy ?time2.
                                ?time4 time:intervalMetBy ?time3.
                                ?time5 time:intervalMetBy ?time4.
                                ?time6 time:intervalMetBy ?time5.
                                ?time7 time:intervalMetBy ?time6.
                                ?time8 time:intervalMetBy ?time7.
                                ?time9 time:intervalMetBy ?time8.
                                ?time10 time:intervalMetBy ?time9.
                                ?end time:intervalMetBy ?time10.

                                          ?time2 time:intervalIn ?per2.
                                          ?time3 time:intervalIn ?per3.
                                          ?time4 time:intervalIn ?per4.
                                          ?time5 time:intervalIn ?per5.
                                          ?time6 time:intervalIn ?per6.
                                          ?time7 time:intervalIn ?per7.
                                          ?time8 time:intervalIn ?per8.
                                          ?time9 time:intervalIn ?per9.
                                          ?time10 time:intervalIn ?per10.
                                BIND(adt:Paleolithic as ?persearch).
                                FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch|| ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch).
                      }
    }
UNION
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?time9 time:intervalMetBy ?time8.
                              ?time10 time:intervalMetBy ?time9.
                              ?time11 time:intervalMetBy ?time10.
                              ?end time:intervalMetBy ?time11.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.
                                        ?time9 time:intervalIn ?per9.
                                        ?time10 time:intervalIn ?per10.
                                        ?time11 time:intervalIn ?per11.

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch || ?per11 = ?persearch).

                      }
    }
UNION
```

```
    {
        select DISTINCT ?artefact ?report where {
            ?artefact adt:wasMadeFrom ?begin;
                      adt:wasMadeUntil ?end;
                      adt:isStoredAt ?EasyDb.
                              ?EasyDb rdfs:label ?report.

                              ?time2 time:intervalMetBy ?begin.
                              ?time3 time:intervalMetBy ?time2.
                              ?time4 time:intervalMetBy ?time3.
                              ?time5 time:intervalMetBy ?time4.
                              ?time6 time:intervalMetBy ?time5.
                              ?time7 time:intervalMetBy ?time6.
                              ?time8 time:intervalMetBy ?time7.
                              ?time9 time:intervalMetBy ?time8.
                              ?time10 time:intervalMetBy ?time9.
                              ?time11 time:intervalMetBy ?time10.
                              ?time12 time:intervalMetBy ?time11.
                              ?time13 time:intervalMetBy ?time12.
                              ?time14 time:intervalMetBy ?time13.
                              ?end time:intervalMetBy ?time14.

                                        ?time2 time:intervalIn ?per2.
                                        ?time3 time:intervalIn ?per3.
                                        ?time4 time:intervalIn ?per4.
                                        ?time5 time:intervalIn ?per5.
                                        ?time6 time:intervalIn ?per6.
                                        ?time7 time:intervalIn ?per7.
                                        ?time8 time:intervalIn ?per8.
                                        ?time9 time:intervalIn ?per9.
                                        ?time10 time:intervalIn ?per10.
                                        ?time11 time:intervalIn ?per11.
                                        ?time12 time:intervalIn ?per12.
                                        ?time13 time:intervalIn ?per13.
                                        ?time14 time:intervalIn ?per14.
                                        ?end time:intervalIn ?perE

                              BIND(adt:Paleolithic as ?persearch).
                              FILTER (?per2 = ?persearch || ?per3 = ?persearch || ?per4 = ?persearch || ?per5 = ?persearch || ?per6 = ?persearch || ?per7 = ?persearch || ?per8 = ?persearch || ?per9 = ?persearch || ?per10 = ?persearch || ?per11 = ?persearch || ?per12 = ?persearch || ?per13 = ?persearch || ?per14 = ?persearch).
                    }
    }
}
Group by ?artefact ?report ?material
```