Universiteit Leiden
The Netherlands

Using text mining to search for Neolithic Vlaardingen Culture sites in the Rhine-Meuse-Scheldt delta

Dikkenberg, L. van den; Brandsen, A.

Citation

Dikkenberg, L. van den, & Brandsen, A. (2025). Using text mining to search for Neolithic Vlaardingen Culture sites in the Rhine-Meuse-Scheldt delta. *Journal Of Computer Applications In Archaeology*, *8*(1), 110-124. doi:10.5334/jcaa.205

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/4249608

Note: To cite this publication please use the final published version (if applicable).

Journal of Computer Applications in Archaeology

Using Text Mining to Search for Neolithic Vlaardingen Culture Sites in the Rhine-Meuse-Scheldt Delta

RESEARCH ARTICLE

LASSE VAN DEN DIKKENBERG
ALEX BRANDSEN
*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

This paper presents a study on Vlaardingen Culture (3400–2500 BCE) sites in the Rhine-Meuse-Scheldt delta using AGNES, an intelligent search engine for Dutch and Flemish archaeological grey literature. The aims of this paper are twofold: 1) to provide an up-to-date overview of Vlaardingen Culture sites; 2) to evaluate the performance of AGNES in searching for period specific sites. Vlaardingen Culture (VLC) sites usually consist of artefact scatters without clearly discernible house plans. These scatters are often found amongst abundant remains from later periods. This type of 'by-catch' is usually not found in the metadata of archaeological reports, and can only be recovered in full text searches. AGNES uses text mining and large language models to allow searches on archaeological concepts (in this case an archaeological culture) in full texts extracted from three major repositories for Dutch (DANS and ARCHIS) and Flemish (Onroerend Erfgoed) archaeology. This paper presents a search for VLC sites, and a comparison of the retrieved information with a recently compiled overview of VLC sites in the area. Using eight queries we retrieved 4532 hits, which were subdivided into: relevant hits (n = 430), semi-relevant hits (n = 2133), and irrelevant hits (n = 1960). We recovered 30 previously unknown Vlaardingen Culture sites, amounting to 19% of the total number of VLC sites (n = 158). Not all sites could be found in AGNES; older archaeological sites are often published in scientific and semi-scientific journals, theses, or books. These publications are absent in the repositories that can be accessed through AGNES, and by extent, they cannot be retrieved. As such, AGNES does not provide an alternative to traditional search methods. Nevertheless, most of the newly found sites consist of sites that cannot be found by searching the metadata of reports in DANS and ARCHIS. Therefore, AGNES proved to be an essential and effective addition to traditional search methods. Finally, our study highlighted the fact that clear terminology to describe Vlaardingen Culture sites is presently lacking. As such, the study provided interesting insights into the terminologies employed in development-led archaeology.

CORRESPONDING AUTHOR:
Alex Brandsen
Faculty of Archaeology, Leiden University – Leiden, The Netherlands
a.brandsen@arch.leidenuniv.nl

KEYWORDS:
Grey literature; Information retrieval; Middle Neolithic; Late Neolithic; Natural language processing; Large language models

TO CITE THIS ARTICLE:
Van den Dikkenberg, L and Brandsen, A. 2025. Using Text Mining to Search for Neolithic Vlaardingen Culture Sites in the Rhine-Meuse-Scheldt Delta. *Journal of Computer Applications in Archaeology*, 8(1): 110–124. DOI: https://doi.org/10.5334/jcaa.205

1. INTRODUCTION

The field of archaeology produces large amounts of textual data, from published books and articles to grey literature reports. In the Netherlands, a large proportion of this textual data is produced by commercial archaeology units, who research and excavate in so-called development-led archaeology. It is estimated that over 5000 excavations and other investigations take place each year (RCE 2023), each producing one or more publications. On top of that, the academic world also produces books, papers and theses. The sheer size of all this data makes finding the right information for synthesising research difficult, and existing archives do not fully match the requirements of archaeologists, the end users of these systems (Habermehl 2024). Using text mining and LLMs to better handle large amounts of data has a long history (Amrani Abajian & Kodratoff 2008; Paijmans & Brandsen 2009; Paijmans & Brandsen 2010; Vlachidis & Tudhope 2012; Richards, Tudhope & Vlachidis 2015; Agapiou & Lysandrou, 2023; Cobb, 2023; Lapp & Lapp, 2024; Gonzalez-Perez et al, 2023; Brandsen 2023), but systems are often developed as proofs-of-concept, not evaluated and made fit for purpose for users, and/or not maintained in the medium to long term. In the EXALT project, we are working to create a search engine that uses text mining and large language models (LLMs) to make information retrieval in Dutch archaeology documents easier and more effective, in a user-centric interface. The system is called Archaeological Grey literature Named Entity Search (AGNES), and is described in more detail in section 3.1.

As opposed to the earlier case study, in which AGNES was used to find Merovingian cremation graves in the Netherlands (Brandsen & Lippok, 2021), this present study is aimed at finding sites attributed to the Vlaardingen Culture. In terms of subsistence strategies, Vlaardingen Culture sites characteristically yield evidence for a mixed strategy involving cereal cultivation and animal husbandry as well as hunting, fishing and gathering (Raemaekers 2005; Van Gijn and Bakker 2005). The sites are predominantly located along the coastal dunes of the western Netherlands and along the major rivers in the western and central Netherlands (Raemaekers 2003). The ceramic assemblages consist mostly of quartz tempered undecorated S-shaped pots, occasionally with a row of perforations under the rim. Furthermore, clay discs and collared flasks are a regular occurrence. The lithic assemblages are dominated by simple 'ad hoc' flake technologies. Flint axes, often found in a broken state, consist of oval axes of the 'Buren-type'. Arrowheads predominantly consist of transverse arrowheads, tanged points, and leaf-shaped arrowheads (Van Gijn 2010; Van Gijn and Bakker 2005; Van Regteren Altena et al. 1962c). Culturally, the group is closely associated with the Stein group, which is mostly located in the Limburg area (see Figure 1). It is not our aim to tackle the issue of defining what the Vlaardingen Culture is. The debate surrounding the distinction between these groups has been the subject of several studies (Louwe Kooijmans 1983; Van den Dikkenberg 2024; Van Gijn & Bakker 2005; Verhart 2010). For the present study we decided to focus on the Vlaardingen Culture and not on the Stein group. Sites attributed to the latter are thus excluded in our overview.

The vast majority of Vlaardingen Culture sites consist of artefact scatters without clear house plans (Van Gijn & Bakker 2005). Only a few sites contain clearly discernible house plans (Stokkel 2017; Van Beek 1990; Van Kampen 2013; Van Zoolingen 2021; Verhart 1992). These artefact scatters are often found as 'by-catch' on archaeological excavations. By-catch refers to 'one or a few finds that are different from the rest of the excavation' (Brandsen & Lippok 2021). Because AGNES allows for full text search, it is well equipped for finding such by-catch, unlike metadata searches. Because of the nature of Vlaardingen Culture sites, they provide an ideal case to test the efficacy of AGNES.

The case study on Merovingian cremations aimed to find something very specific (scarce cremation graves) within Merovingian contexts. The current study has a broad aim, finding anything that is attributed to the Vlaardingen Culture. In the study on Merovingian cremations it was demonstrated that AGNES excelled at finding these very specific types of finds. Brandsen and Lippok were able to recover a total of 23 previously unknown Merovingian cremation graves (2021). In the present paper we test the usefulness of AGNES in making period specific site overviews, focussing on sites attributed to the Vlaardingen Culture. To summarise, the aim of this study is twofold: 1) we aim to provide an up-to-date overview of Vlaardingen Culture (3400–2500 BCE) sites; and 2) we aim to evaluate the performance of AGNES in searching for period specific sites.

The research questions are as follows:

- Compared to previously known sites, what unknown sites can we find with AGNES?
- What does this mean for the usefulness of AGNES?
- What do the newly rediscovered sites mean for the distribution of Vlaardingen Culture sites in the Rhine-Meuse-Scheldt delta?

2. DATA

This paper relies on data on known Vlaardingen Culture sites, and the data available in AGNES. In this section we describe this data.

2.1 THE VLAARDINGEN CULTURE DATA SET

Settlements from the Vlaardingen Culture are mainly located in the coastal area of the Netherlands, notably

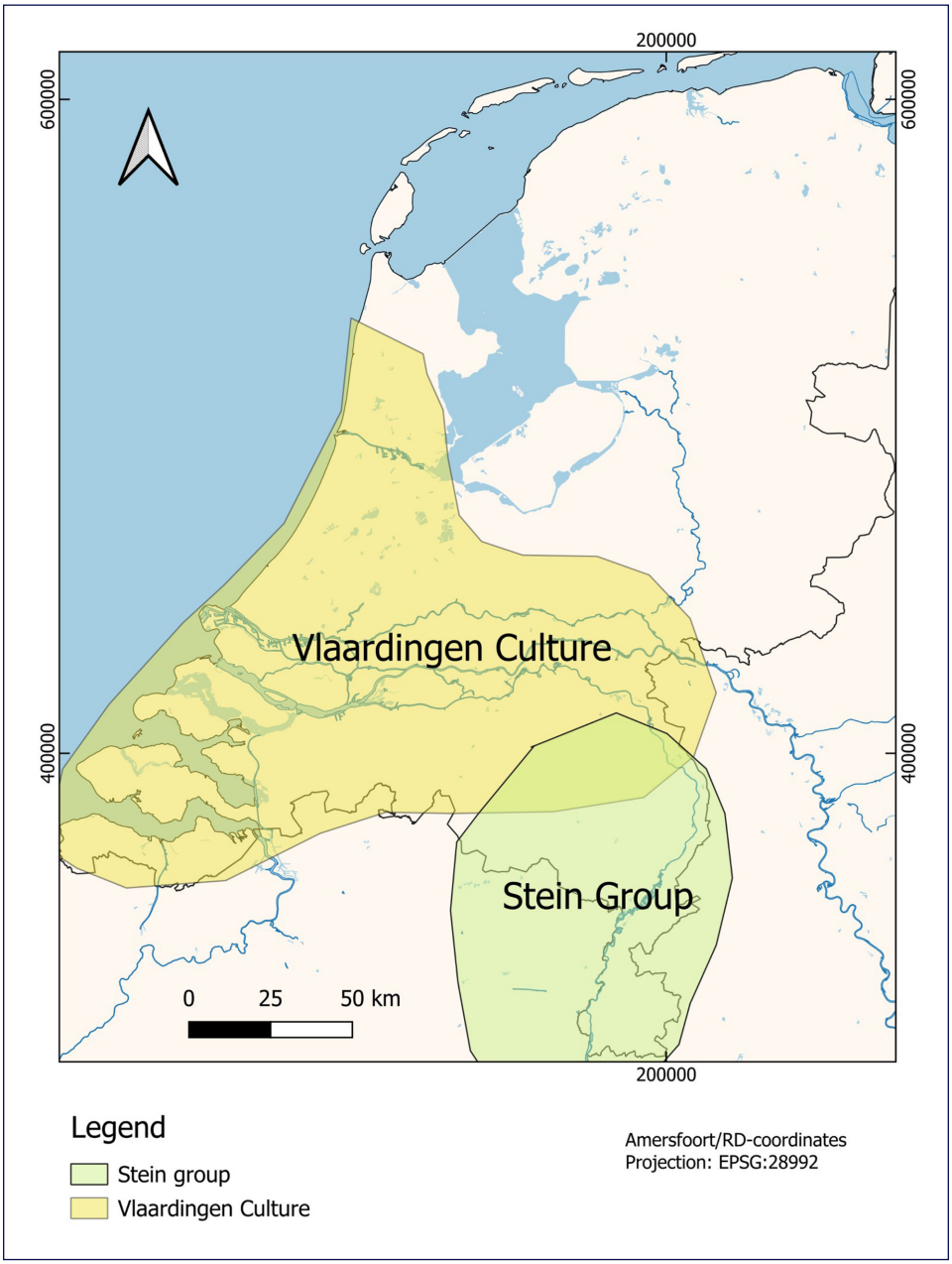

Figure 1 Distribution of the Vlaardingen Culture and Stein group (after: Verhart 2010; base map: © EuroGeographics 2024, map made in QGIS).

in wetland areas along rivers and the coastal dunes. The most complete overview of Vlaardingen Culture sites so far is presented in the distribution map by Verhart and de Ridder, an overview containing ca. 80 sites (Verhart & de Ridder 2010). This is more than the overview presented five years earlier in The Prehistory of the Netherlands, where it was mentioned that there are about 30 Vlaardingen Culture sites in the Netherlands (Van Gijn & Bakker 2005). Neither of these overviews present us with a list of sites. Furthermore, because the last overview of sites was made in 2010, newly excavated sites are lacking from these existing overviews. Recent discoveries have challenged some of the notions previously held about this period. The excavations at Den Haag Wateringse Binnentuinen zone 3 and Den Haag Noordweg 76 revealed that temporary settlements, previously thought to be a feature exclusively present on sites located on river levees, also occur in the coastal dune area of the Netherlands (Bulten & Stokkel 2017; Raemaekers 2003; Van Zoolingen & Rieffe 2023). Furthermore, the recent excavations at Den Haag Steynhof and Den Haag Wateringse Binnentuinen were amongst the largest and best documented Vlaardingen Culture settlements to date (Bulten & Stokkel 2017; Van Zoolingen & Bulten 2021). The site Veldhoven Habraken demonstrated that permanent Vlaardingen Culture sites can also be expected further inland on the sandy soils of Noord Brabant (Van Kampen 2013). These new discoveries highlight the need for a renewed overview of Vlaardingen Culture sites. An attempt to create such an overview was undertaken by Van den Dikkenberg as part of his current PhD project, which is part of the Putting Life into Late Neolithic Houses

project. Based on previously published overview studies, known site reports, and queries in ARCHIS and DANS, an overview was compiled of 129 known Vlaardingen Culture sites in the Netherlands and Belgium. This data set will be compared with the new results, as it is a fair reflection of what information can be found with tools and databases besides AGNES.

2.2 AGNES DATA SET

The aim of this data set is to incorporate all open access documents about Dutch archaeology and neighbouring countries, which is still underway. At the time of this study, just over 188,000 documents are included in AGNES, from the following sources:

- Documents from the DANS archive marked with the 'Archaeology' tag, up to December 2021 (https://dans.knaw.nl/nl/archaeology/)
- Documents from the ARCHIS database, up to December 2021 (https://archis.cultureelerfgoed.nl/)
- Documents from the Onroerend Erfgoed archive, up to June 2024 (https://www.onroerenderfgoed.be/)

For all of these sources, we only harvested and indexed PDF files, and these contain a multitude of document types. This includes excavation reports, coring reports, appendices, database descriptions, personal daily reports, maps, find lists and sometimes even photographs stored within PDFs.

3. METHODS

Below we will first introduce the AGNES search system, next we will discuss our search methodology for the case-study on Vlaardingen Culture sites. The data is visualised in distribution maps in QGIS to highlight the spatial distribution of our findings. To visualise the relationships in our data we used network visualisations which were created in Visone. We adopted network graphs because these provide a visual representation in which the results of multiple queries can be summarised in a single relational graph.

3.1 AGNES, TEXT MINING, AND LARGE LANGUAGE MODELS

As mentioned in the introduction, the original literature search for VLC sites in archives was done by metadata search; searching in e.g. the title, description, and keywords. However, this metadata can be incomplete and/or inaccurate, and is missing detailed information (Habermehl 2024). Think of an excavation of a Roman encampment; the metadata is not going to mention a single Neolithic find (by-catch), but this find is only mentioned in the excavation report. To solve this, we can apply full-text search, searching through all of the text instead of just the metadata. This would be a significant improvement, and is something the DANS archive has since implemented. However, archaeological discourse includes a lot of synonyms and homonyms; multiple words with the same or similar meaning (e.g. medieval and Middle Ages), and words with multiple meanings ('Flint' being both a material and a surname), respectively, which makes searching more difficult (Brandsen 2021b). Within AGNES, we try to solve the homonym problem using named entity recognition (NER), a natural language processing technique that finds and extracts certain entity types (Tjong Kim Sang 2002). In our research, we target artefacts, time periods, contexts, species, materials and locations. Once the word 'flint' has been identified as a material by NER, we know it's not the surname flint, and we solve the homonym problem. For time periods, we try to solve the synonym problem by taking detected time period entities, and translating them to a start and end year. This way we can search for e.g. '500 to 1500 AD' and this will return results for both 'medieval' and 'Middle Ages', as well as sub-periods and single years within this year range.

The NER is done using BERT (Bidirectional Encoder Representations from Transformers), one of the first large language models (Devlin et al. 2019). The rise of LLMs after the introduction of transformers by Vaswani et al. (2017) has had a huge impact on many fields, including archaeology. Transformers made it possible to handle language tasks with more speed and accuracy, thanks to their self-attention mechanism. Unlike traditional methods, these models can understand context, synonyms, and specific terminology, making it easier to find valuable information buried in large data sets. Similar to the newer GPT (Generative Pre-trained Transformer) models, BERT uses large amounts of unlabelled text data to pre-train a model, gaining an understanding of words and their contexts. We took these generic models, and further pre-trained them with texts from the archaeology domain (Brandsen 2021a; Brandsen 2024; Brandsen et al. 2022). This created archaeology specific BERT models for Dutch, English, and German, and finally we fine-tune these with labelled NER data to be able to predict and extract entities. Together with the full text of the documents, these entities are indexed in ElasticSearch, an open source search engine (Gormley & Tong, 2015). We built a frontend to query all this information, specifically designed and evaluated with the archaeologists' needs in mind (Brandsen et al. 2019; Brandsen et al. 2021). The system can be freely accessed via https://agnessearch.nl, and all code/models will be made available open access at the end of the AGNES project.

3.2 SEARCH METHODOLOGY

In total, eight queries were entered in AGNES resulting in a total of 4532 hits (see Table 1; see supplementary Table 1). These were exported to a single CSV file. Usually we did not use start and end dates for the queries, to expand

START DATE (BCE)	END DATE (BCE)	FREE TEXT QUERY	ENGLISH TRANSLATION	NUMBER OF HITS
		"vlaardingen cultuur"	vlaardingen culture	834
3800	2000	"vlaardingen*"	vlaardingen	2483
		"vlaardingen stein wartburg"	vlaardingen-stein-wartburg	11
		"vlaardingen stein wartberg"	vlaardingen-stein-wartburg	4
		"vlaardingen groep"	vlaardingen group	265
		"vlaardingen stein"	vlaardingen-sStein/vlaardingen stein	98
		"vlaardingencultuur"	vlaardingen culture	712
		"vlaardingengroep"	vlaardingen group	125

Table 1 List of queries entered in AGNES with number of hits.

the number of potential hits. Only for the open query 'vlaardingen*' did we include dates to limit the number of hits. We opted to use a broader date range (3800–2000 BCE) than the traditional starting and end dates for the Vlaardingen Culture (3400–2500 BCE) (Raemaekers 2005, 271). This would increase the opportunity to find VLC sites which have date ranges expanding beyond the generally accepted starting and end dates for the VLC. It is worth noting here that the free text query in Table 1 is simply doing a term match in ElasticSearch, while the start date and end date are making use of the time periods detected by BERT. For this particular study, we do not make use of the other entity types, such as artefacts or materials.

Three additional columns were added to the CSV file to include the query text, a ranking of how relevant the publication was (see Table 2), and lastly a column in which for irrelevant hits the reason was stated for the irrelevance of the hit (see Table 3). Following the methodology presented by Brandsen and Lippok we manually checked the hits. The listed categories are an adapted version of those applied by Brandsen and Lippok (Brandsen & Lippok 2021). Relevant hits consist of site reports on previously unknown Vlaardingen Culture sites (1), previously known Vlaardingen Culture sites (2), as well as reports in which previously unknown Vlaardingen Culture sites are mentioned in the text (3) or the literature lists (4). In these last two instances, the hits indirectly led to the discovery of new sites. In addition to these categories we included three semi-relevant categories. The Stein group and Vlaardingen group are closely related and as a result many reports on Stein sites also discuss the relationships between these groups. This was considered a semi-relevant hit (5) because AGNES correctly identified that the text discussed the Vlaardingen Culture, but the report itself was not about a VLC site. Similarly, other site reports also discussed the Vlaardingen Culture (6) or Vlaardingen Culture sites (7). Research plans for example often mentioned the Vlaardingen Culture, or Vlaardingen Culture sites when describing the archaeological potential of a study area.

NR.	RELEVANCE
1	Relevant (report about a Vlaardingen Culture site) unknown
2	Relevant (report about a Vlaardingen Culture site) known
3	Relevant (previously unknown Vlaardingen site mentioned in the text)
4	Relevant (previously unknown Vlaardingen site mentioned in the literature list)
5	Semi-relevant (Stein site publication, mentioning Vlaardingen Culture in discussion)
6	Semi-relevant (Vlaardingen Culture mentioned in a discussion)
7	Semi-relevant (a different Vlaardingen Culture site mentioned in the text based on previous research)
8	Not relevant (not a report about a Vlaardingen Culture site)

Table 2 Relevance of AGNES hits.

NUMBER	TYPE OF IRRELEVANT DOCUMENT
1	Wrong time period
2	Page listing abbreviations
3	Page containing research plan (plan van aanpak)
4	Unknown time period
5	Page containing list of time periods
6	Negation ('no vlaardingen culture')
7	Other
8	Literature list (only)
9	Coring chart
10	Database structure
11	Vlaardingen as a location on a map
12	Vlaardingen as place name in text

Table 3 Types of irrelevant hits.

Such hits were therefore considered to be semi-relevant hits. In the previous study by Brandsen and Lippok, hits in research plans were considered irrelevant hits (Brandsen & Lippok 2021). This makes sense for cremations as the expectation of finding cremation graves can refer to a plurality of periods. However, if research plans specifically mention the Vlaardingen Culture they do refer specifically to what we aimed to find in our queries. Therefore, in this case we often considered these as semi-relevant hits, rather than irrelevant hits. These research plans also often mentioned nearby Vlaardingen Culture sites (7).

Irrelevant hits (1–9) were classified along the typology presented by Brandsen and Lippok (Brandsen & Lippok 2021). We added three frequently occurring categories. Being a period designation, *"vlaardingen cultuur" (Vlaardingen Culture)* was frequently mentioned in documents containing a database structure. Therefore, this was added as a separate category (10). The Vlaardingen Culture is named after the type-site Vlaardingen Arij Koplaan (Van Regteren Altena et al. 1962a; 1962b, 1962c). Because of this our hits frequently contained reports which only mentioned the place name 'Vlaardingen' (12). Similarly, geographical maps which included the city of Vlaardingen were also considered as a separate category of irrelevant hits (11).

Above we presented different categories of relevant, semi-relevant, and irrelevant hits. This is based on the usefulness of the hits for the archaeological case-study. Such hits are not necessarily incorrect. When a report deals with a Stein site and it mentions the Vlaardingen Culture the hit for the query *"vlaardingen cultuur"* (Vlaardingen Culture) is correct, but it is irrelevant as the report does not deal with a Vlaardingen Culture site. Similarly, hits in which *"vlaardingen cultuur"* (Vlaardingen Culture) is only mentioned in the literature list are correct hits in the sense that AGNES correctly identified matching terms, but because it concerns a hit in the literature list it is deemed irrelevant for the case-study.

4. RESULTS

The different queries yielded in total 439 relevant hits (see Table 4). This means that 9.7% of the hits consisted of relevant hits. In addition to those 2133 (47.1%) hits were classified as semi-relevant and 1960 (43.2%) hits were classified as irrelevant. For the full data, see (supplementary Table 1, doi.org/10.5281/zenodo.14842975). The relevance of reports depended on the specific queries which returned the hits. For example, in the report of the excavation of Hellevoetssluis-Ossenhoek the site is consistently referred to as a 'Vlaardingen-groep' site (Goossens 2009). The query '*vlaardingen groep*' (Vlaardingen Group) in this case thus yielded relevant hits. The query '*vlaardingen cultuur*' (Vlaardingen Culture) however does not yield relevant hits for this publication, it does however yield two hits in the bibliography in which publications are cited which mention the term '*vlaardingen-cultuur*' (Vlaardingen Culture) (Goossens 2009: 177–179). A single publication can thus return multiple hits for multiple queries.

The irrelevant hits were further subdivided into categories (see Table 5), especially the query for 'vlaardingen' yielded a large number of irrelevant hits (n = 1506; see Figures 2 and 3). These mostly related to the fact that Vlaardingen is a place name, therefore many reports on archaeology in the city of Vlaardingen were included in this query, as well as reports mentioning Vlaardingen as a place name in the text (n = 897), or in a map (n = 8). Frequently, hits included text only contained in the bibliography, a database structure, or lists of time

RELEVANCE	COUNT	%
Relevant (report about a Vlaardingen Culture site) unknown	165	3.6
Relevant (report about a Vlaardingen Culture site) known	259	5.7
Relevant (previously unknown Vlaardingen site mentioned in the text)	9	0.2
Relevant (previously unknown Vlaardingen site mentioned in the literature list)	6	0.1
Semi-relevant (a different Vlaardingen Culture site mentioned in the text based on previous research)	1398	30.8
Semi-relevant (Stein site publication, mentioning Vlaardingen Culture in discussion)	65	1.4
Semi-relevant (Vlaardingen Culture mentioned in a discussion)	670	14.8
Not relevant (not a report about a Vlaardingen Culture site)	1960	43.2
Total	**4532**	**99.8**

Table 4 Relevance of AGNES hits totals.

IRRELEVANCE REASON	COUNT	PERCENTAGE
Page listing abbreviations	4	0.2%
Page containing research plan (*plan van aanpak*)	2	0.1%
Page containing list of time periods	257	13.1%
Negation ('no vlaardingen culture')	16	0.8%
Literature list (only)	465	23.7%
Database structure	302	15.4%
Vlaardingen as place name in text	905	46.2%
Vlaardingen as place name on a map	8	0.4%

Table 5 Reasons for irrelevant hits AGNES.

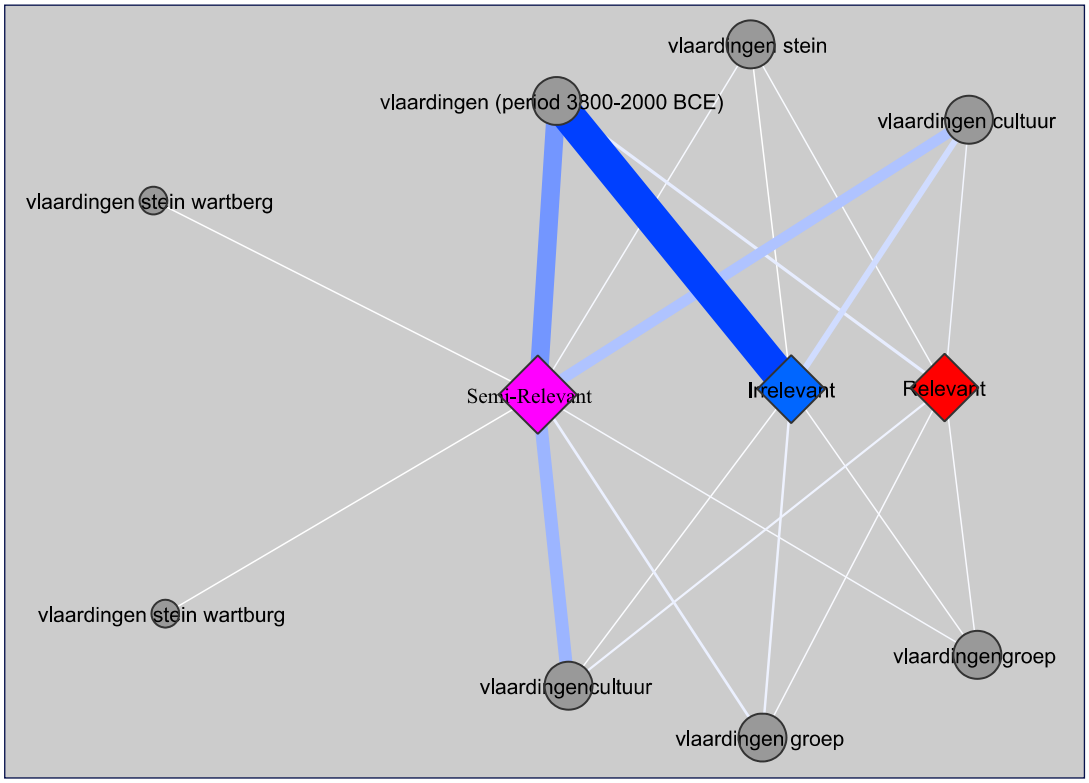

Figure 2 Network representation; two-mode network visualising the relevance of different queries. Network visualising the different queries (grey) and relevant (red), irrelevant (blue), and semi-relevant hits. Nodes are scaled according to their centrality degree (std), links are ranked by weight, visualized in stress minimization layout (graph made in Visone).

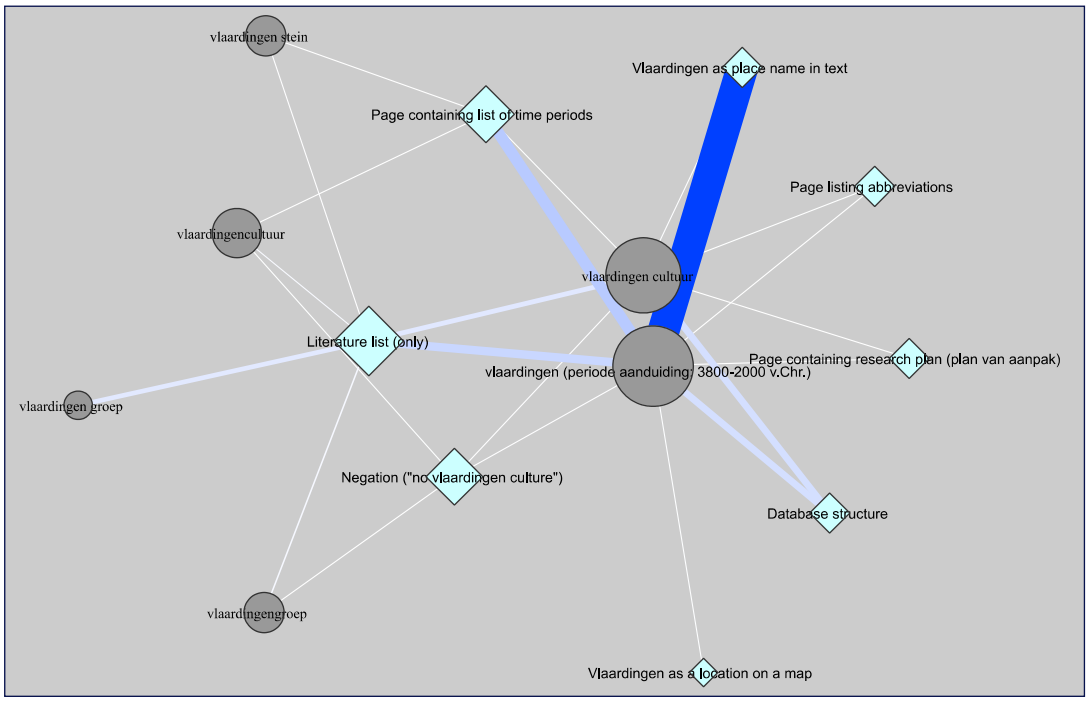

Figure 3 Network representation; two-mode network visualising irrelevance types (blue) for different queries (grey), Nodes are scaled according to their centrality degree (std), links are ranked by weight, visualized in stress minimization layout (graph made in Visone).

periods. These hits were deemed irrelevant but they were generally not incorrect as AGNES correctly matched the search terms in the documents.

It is interesting that the queries for "*vlaardingen stein wartburg*" and "*vlaardingen stein wartberg*" only yielded semi-relevant hits (see Figure 2). These terms were thus not used in reports about Vlaardingen Culture sites but they were used in discussions in other site reports.

In terms of irrelevant hits it is interesting to note that hits relating to database structures and lists of time

periods never contained hits for *"vlaardingengroep"* (Vlaardingen Group) or *"vlaardingen groep"* (Vlaardingen Group). Database structures systematically use the terms "vlaardingen" or *"vlaardingen cultuur"* (Vlaardingen Culture) (see Figure 3). Standardised lists of time periods use the terms "vlaardingen", *"vlaardingen culture"* (Vlaardingen Culture), or *"vlaardingen stein"* (see Figure 3). As such the irrelevant hits provide additional information relating to the terminology frequently employed in development-led archaeology.

4.1 NEWLY DISCOVERED SITES

The queries yielded information on a total of thirty sites (19% of the total number of sites) which were not previously included in the overview (see Table 6; see: supplementary Table 2 and supplementary file 3 doi.org/10.5281/zenodo.14842975). In 27 instances this included hits on site reports of previously unknown Vlaardingen Culture sites. In three instances the hits consisted of indirect hits. These were site reports or research plans which mentioned previously unknown Vlaardingen Culture sites in their respective study areas.

For 13 of the 89 sites which were not found in AGNES, the publications were present in DANS or ARCHIS. In seven cases this concerned sites published in the monograph by Louwe Kooijmans (1974). These sites were found in the appendix where they were listed under the abbreviation 'VL'. As this abbreviation was not part of our queries these sites were not found in AGNES, even though the file in which they were listed was available. In one instance the report did not mention a cultural attribution, but the material could be attributed to the VLC based on the characteristics of the finds. For four sites the documents were present in DANS but these were not imported in AGNES due to errors during the PDF text extraction process (De Koning 2010; Eijskoot 2004; Eimermann 2008; Van den Broeke 1993). In one instance the BERT model did not recognise a date correctly (Hiddink 2000: 11). The publication in this case mentions the term 'Vlaardingen' along with several other Neolithic cultures and a date range (4000–2000 BCE). But, as the date range was not recognised as a time period, the query incorrectly did not match this page.

RESULT PER SITE	COUNT	PERCENTAGE
Found exclusively in AGNES	27	17.1%
Found exclusively indirectly in AGNES	3	1.9%
Found previously and in AGNES	39	24.7%
Not found in AGNES queries (pdf not present in DANS or ARCHIS)	76	48.1%
Not found in AGNES queries (pdf is present in DANS)	13	8.2%
Total	**158**	**100%**

Table 6 Results per site, newly found sites, previously known sites and sites of which the reports are not in AGNES.

In one instance the unknown site was only mentioned in the bibliography of a report. The article concerned a previously unknown site on the island of Texel (Van Noort 1998). The site was discovered by an amateur archaeologist and it was published in a local historical journal, which explains why it remained unknown despite being published in 1998. This is an interesting find because the site is located much further north (see Figure 4, the most northern purple dot on the map) than the most northern Vlaardingen Culture site known so far: Zandwerven (Van Gijn & Bakker 2005; Verhart & de Ridder 2010). It is located about twenty kilometres north of what is traditionally assumed to be the limit of the distribution of the Vlaardingen Culture (Van Gijn & Bakker 2005). As such the site presents a significant discovery.

Interestingly, many of the newly discovered sites are located in the eastern Netherlands in the area of Nijmegen (see Figure 5). In the 2010 overview it was also noted that this area yielded a high number of Vlaardingen Culture sites (Verhart 2010). At the time this concentration could mostly be attributed to the tireless efforts of the local AWN (Association of Archaeology Volunteers) dependance. Between 1970 and 2000 this group discovered a great deal of sites in the Nijmegen and Wijchen area (Teubner & Tuijn 2010; Verhart & de Ridder 2010). The sites which were newly found in AGNES mostly consist of recent (post 2010) excavations in the area. Often these reports mention the previous studies conducted by local volunteers ('t Hart, Norde & Tuinstra 2019:12; Heirbaut 2010, 12; Janssen 1989; Janssen & Tuijn 1978). These studies thus led to a better formulation of archaeological expectations, and in turn to a better formulation of research plans. In recent years this led to new excavations which in turn led to new discoveries. It is an excellent example of how citizen science contributed to development-led archaeology.

5. DISCUSSION

Below we will discuss two main themes which popped up during our investigation. The first is the plurality of terms used to describe Vlaardingen Culture sites. The second part will discuss the data sources accessed in this study, with the aim of explaining on the one hand why we were able to find previously unknown Vlaardingen Culture sites through AGNES, while we will also discuss the types of sources which were missed during this study. This section will also provide recommendations on how to make these sources more accessible for future studies.

5.1 VLAARDINGEN OR STEIN GROUP?

Regarding the cultural attributions of Vlaardingen Culture sites, we decided to adhere to the conclusions presented

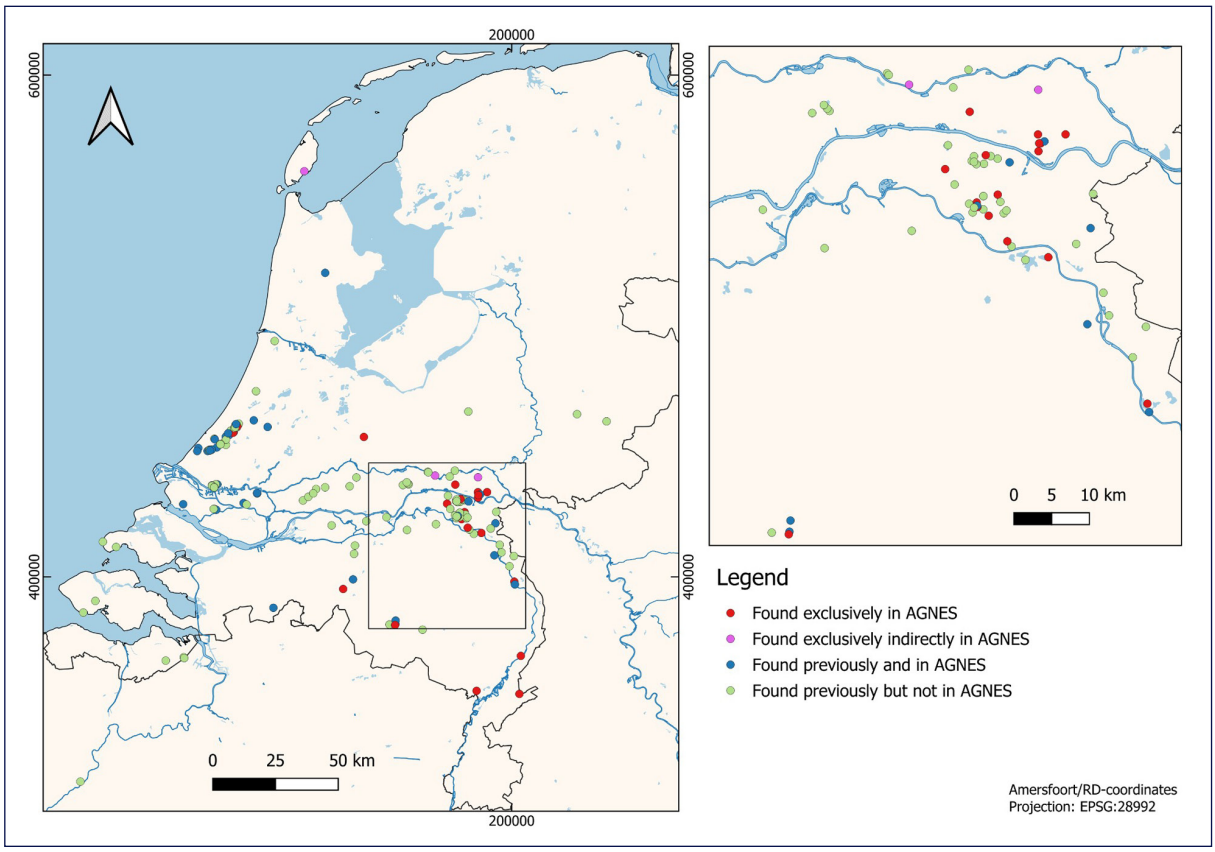

Figure 4 Vlaardingen Culture sites plotted according to whether or not they were found in AGNES (basemap: © EuroGeographics 2024, map made in QGIS).

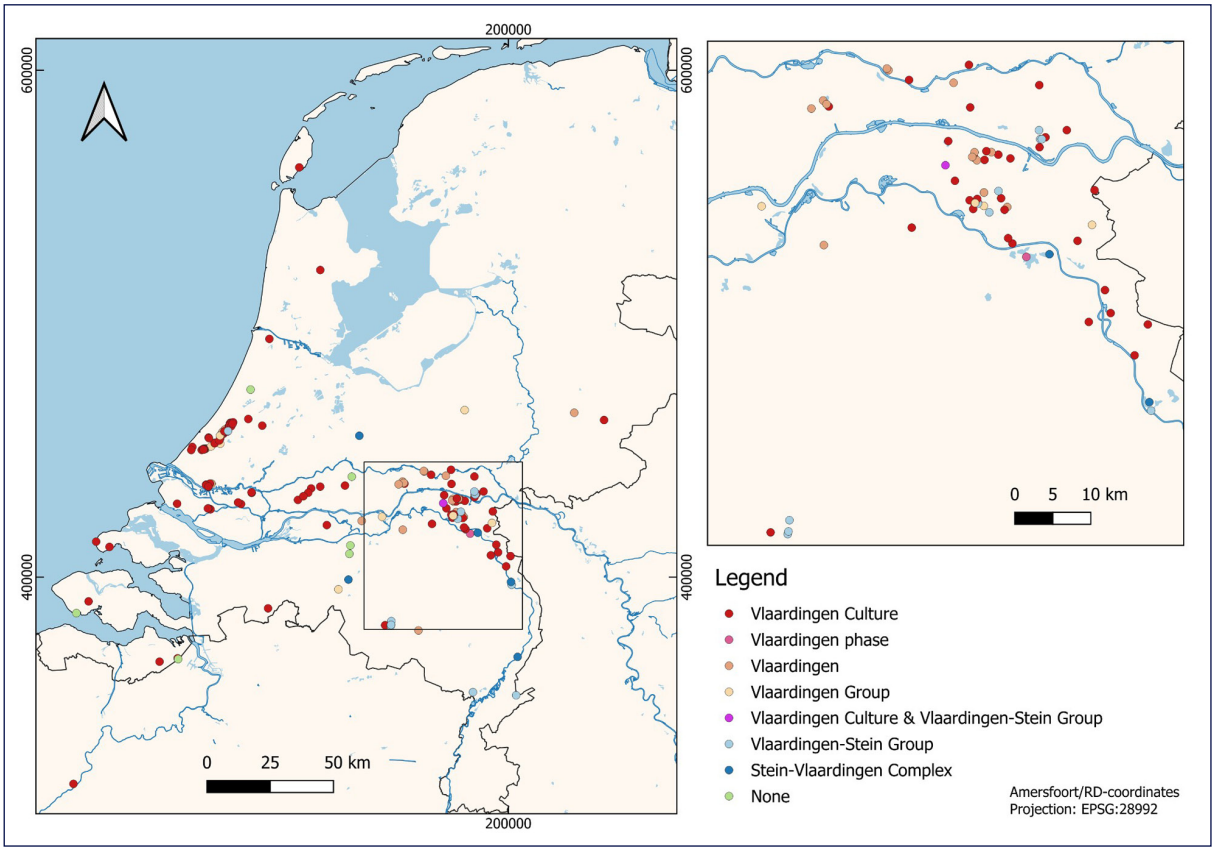

Figure 5 Vlaardingen Culture sites plotted according to their cultural attribution, on the right the supposed border area between the Vlaardingen and Stein group (basemap: © EuroGeographics 2024, map made in QGIS). The sites are plotted according to their cultural attribution, on the left the supposed border area between the Vlaardingen and Stein group.

by the excavators. In seven cases an exception was made. These consist of sites without a cultural attribution but where the find material is dated to this period; sites where the material is consistent with Vlaardingen Culture material; and sites which are geographically located in the area of the Vlaardingen Culture. This is for example the case with the single find of a flint oval axe in Elshout. Similar finds of single oval flint axes in the western Netherlands are consistently attributed to the Vlaardingen Culture (Dorenbos & Koot 2010; Groenman-van Waateringe & Van Regteren Altena 1966). For the other sites a plurality of cultural attributions was used; Vlaardingen Culture, Vlaardingen group, Vlaardingen-Stein group (or Stein-Vlaardingen group), and Stein-Vlaardingen Complex (see Figure 5). It seems that different terms here are generally not applied based on differing archaeological characteristics. Rather, it seems that terms are regionally dependent. Sites in Zuid Holland (and more generally in the western Netherlands) are usually referred to as Vlaardingen Culture sites. Sites in, and around, the border zone between the Vlaardingen and Stein group are often referred to as "*vlaardingen-stein groep*" (Vlaardingen-Stein group) or "*stein-vlaardingen complex*" (see Figure 5).

The fact that these cultural attributions in the literature depend more on the geographical location of these sites rather than the archaeological material is occasionally made explicit. The Stein site Schoolstraat in Thorn is for example attributed to the Stein group because *'the Vlaardingen Culture predominantly occurs in the coastal area of the western and southern Netherlands, as well as the riverine area in the central Netherlands, the site probably represent remains from the Stein group'*[1] (De Ridder 2011: 27). This problem was already envisioned in 1983 by Louwe Kooijmans when he first defined the characteristics of the Stein Group '*The here discussed Late Neolithic find groups display a high degree of affinity with the Vlaardingen Culture, so much so that we believe that, had they been found in the delta region, they would have been, without much trouble, classified as Vlaardingen*'[2] (Louwe Kooijmans 1983: 64). The supposed differences between these groups; the presence of Lousberg axes, blade technologies, and axe production for the Stein group vs. the presence of ceramic baking plates and pottery with perforations under the rim for the Vlaardingen Culture, might be real to some extent (Louwe Kooijmans 1983; Van den Dikkenberg 2024; Verhart 2010). Nevertheless, they are too often absent, making cultural attributions for sites discovered in development-led archaeology difficult.

It is noteworthy that the term Vlaardingen-Stein-Wartberg complex is not applied in any of the reports. Based on the irrelevant hits this term also seems to be avoided in lists of abbreviations, periods, and in database structures (see Figure 3). Such lists systematically employ either the term Vlaardingen, or a variant of Vlaardingen Culture. Overall it seems that these are the dominant terms used in development-led archaeology (see Figures 3 and 5).

5.2 DATA SETS

There are clear regional differences in terms of which sites were found by AGNES. As mentioned before a lot of sites found in AGNES were located in the Nijmegen area. Furthermore, clearly a lot of previously known sites in Zuid Holland were found as well (see Figure 4). Several other areas are however largely missed. None of the sites in Zeeland were for example found in AGNES. This is not entirely surprising as all but one of these sites were excavated in the twentieth century. As such their reports were not deposited in DANS or Archis, and by extension they could not be found in AGNES. Older sites are often published as articles in archaeological journals, rather than as site reports deposited in DANS. This is for example the case with many of the key Vlaardingen Culture sites including the type site Vlaardingen Arij Koplaan and the key sites of Zandwerven, Haamstede Brabers, and Leidschendam Prinsenhof (Clason 1962; Glasbergen, Groenman-van Waateringe & Hardenberg-Mulder 1967; Van Iterson Scholten 1988; Van Regteren Altena 1958; Van Regteren Altena & Bakker 1961; Van Regteren Altena et al. 1962a; Verhart 1992). Similarly, archaeological sites discovered by volunteers are also generally absent in DANS. They are often published in either the Westerheem (now Archeologie in Nederland) or in local AWN reports. This is for example the case with many of the previously mentioned sites discovered by volunteers in the Nijmegen area between 1970 and 2000 (De Jong 1986; 1988; Janssen 1976; 1980; 1989; 1993; Janssen & Tuijn 1978; Koolen 1976). A similar problem applies to many of the sites found in the central part of the Netherlands and eastern parts of Zuid Holland. For example, a series of sites in the municipality Molenlanden were discovered by local AWN volunteers in the 1960's. These sites were published in the dissertation of Louwe Kooijmans, but they do not have formal published excavation reports (Louwe Kooijmans 1974).

A core strength of AGNES is that it allows us to find 'by-catch' in archaeological reports. Many of the previously unknown Vlaardingen Culture sites found in AGNES can be considered to be 'by-catch'. This is for example the case with the site Nijmegen Park Waaijenstein. The excavation focussed on a Roman period settlement in the area. The metadata of the report in DANS only mention the Roman period settlement. The Vlaardingen Culture remains at the site consist of three ceramic sherds and a few flint artefacts (Daniël 2018). The site Bergharen de Weem presents a similar case, the report titled 'On the edge of a medieval settlement'[3] is focussed on medieval finds. The metadata in DANS mention Neolithic remains but those are not specified. As such they would not be found with the queries such as "*vlaardingen cultuur*" (Vlaardingen

Culture). Only in the full text of the report is it specified that these remains consist of flint and ceramics from the Vlaardingen Culture (Diepeveen & Van Enckevort 2009). It is not surprising that many of these by-catch finds are located in the region of Nijmegen. Nijmegen is the heart of the Roman Netherlands, it is the oldest city in the country, and a major centre during the medieval period. Archaeological excavations frequently yielded large quantities of finds from these periods. It is perhaps unsurprising that a handful of Vlaardingen Culture sherds or flint artefacts on these excavations do not end up in the metadata of these reports. This is also no longer problematic as we were now able to retrieve this kind of information through AGNES.

Unfortunately, we cannot calculate the recall (and by extension, the F1 score), as the total amount of relevant documents in the collection is unknown. We can however make an estimation of the recall of the already known sites, i.e. how many of the known sites that are indexed in AGNES were actually retrieved. This would be a recall of 0.75. The current case-study yielded a precision of 9.7%, in terms of relevant hits. This is a much higher precision than that of the previous case-study which dealt with Merovingian cremation graves. In this case-study only 2.1% of the hits consisted of relevant hits (Brandsen & Lippok 2021). This is a significant improvement, which can partly be attributed to improvements following the recommendations made in 2021, and partly due to the type of queries. As part of the EXALT project several other case-studies will be carried out to further assess the efficiency of the system. It will be interesting to see whether the improved precision is indeed a constant factor or whether this is largely case-dependent.

The original overview which was compiled of Vlaardingen Culture sites took several months to be completed and it has been further refined over the past years. Going through the AGNES queries took about two to three weeks. It is clear that AGNES vastly speeds up the process of compiling such overviews. This can mostly be attributed to the fact that AGNES provides direct access to the relevant literature, and more specifically to the relevant pages contained within those documents.

Finally, two main problems in AGNES were identified during the case study; PDFs not being imported into AGNES due to PDF text extractions errors, and BERT missing a date range. The PDF text extraction process has been updated after the case study to solve this problem, by using a different tool less prone to errors (PyMuPDF[4]). In a future version of AGNES, we will re-index all the documents again, which means these missed documents will be available in the future. Regarding the BERT error, the models currently have an F1 score of around 84% for detecting time periods (Brandsen et al. 2022), meaning that around 16% of time periods are missed or incorrectly classified. In future work, we want to improve on this performance by improving the BERT model, potentially with more training data, or by using newer techniques such as GPT models.

6. CONCLUSION

In the present study we aimed to test how well AGNES was equipped for generalised queries aimed at finding period specific sites, in this case sites attributed to the Vlaardingen Culture. As such our tests deviated from an earlier case-study which attempted to find highly specific information, in this case on Merovingian cremation graves (Brandsen & Lippok 2021). We can conclude that AGNES also greatly contributes to more general queries. Through AGNES we found thirty (19%) previously unknown Vlaardingen Culture sites. This included the most northern Vlaardingen Culture site ever found. As such, the study also contributed to our understanding of the geographical spread of the Vlaardingen Culture phenomenon.

Newly discovered sites often consisted of by-catch on excavations in which the majority of finds were dated to other (later) periods. Although AGNES greatly contributed to our overview the program also missed a large number of known Vlaardingen Culture sites. This is mainly the case because older sites are usually not published in site reports which have been deposited in DANS or Archis. Often such sites are published in scientific and semi-scientific publications. Sites by amateur archaeologists are usually published in local archaeological and historical journals. It is recommended that such publications are digitised and deposited in DANS. This will increase the visibility and usefulness of citizen science.

Another problem we observed is that there is no consensus amongst authors on which terms to use to describe Vlaardingen Culture sites. They are referred to as: Vlaardingen Culture, Vlaardingen group, Vlaardingen-Stein group (or Stein-Vlaardingen group), or Stein-Vlaardingen Complex. It seems that the terms are applied rather arbitrarily, when sites are located in the border area between the Stein group and Vlaardingen Culture they are often referred to as Stein-Vlaardingen or Vlaardingen-Stein sites. In the western Netherlands the term Vlaardingen Culture is more systematically applied. The problem seems to stem from the fact that a distinction between the Stein group and Vlaardingen Culture is problematic, especially when dealing with a small number of finds, discovered by chance, as is often the case in development-led archaeology. While we were not able to provide a satisfactory solution to this problem, the study provided valuable insights into the terminologies employed in development-led archaeology.

We can conclude that AGNES cannot be used as an alternative to established search methods for creating thematic or temporal site overviews, because archaeological sites are not exclusively published in excavation reports. Nevertheless, it provides an effective tool for finding archaeological 'by-catch'. In this case thirty (19%) of the Vlaardingen Culture sites in our overview were discovered exclusively through AGNES. This type of 'by-catch' cannot be effectively found through other means, therefore it is recommended that AGNES is used systematically in tandem with established search methods.

DATA ACCESSIBILITY STATEMENT

The data used in this research is available in this Zenodo archive (doi.org/10.5281/zenodo.14842975), together with the supplementary materials (1–3). The BERT models and code are also available (Brandsen, 2023).

NOTES

1 "Omdat de Vlaardingencultuur voornamelijk in het holocene kustgebied van west en zuid-west Nederland, alsmede het midden-Nederlandse rivierengebied voorkomt, betreft het hier waarschijnlijk de culturele nalatenschap van de Stein-groep. Vindplaatsen hiervan kennen we uit Limburg (met een concentratie in Midden-Limburg aan weerszijden van de Maas), het aangrenzende Rijnland, Noord-Brabant en het oostelijk rivierengebied" (De Ridder 2011: 27).

2 "De hier besproken, laat-neolithische vondstgroepen tonen een grote verwantschap met die van de Vlaardingen-cultuur, zodanig zelfs, dat wij het idee hebben dat zij, waren ze in de delta gevonden, zonder veel moeite „VL" waren genoemd" (Louwe Kooijmans 1983: 64).

3 "Aan de rand van een middeleeuwse nederzetting" (Diepeveen & Van Enckevort 2009).

4 https://pymupdf.readthedocs.io/en/latest/.

ACKNOWLEDGEMENTS

We like to thank Annelou van Gijn, PI of the Putting Life into Late Neolithic Houses project, for supporting the collaboration between the Putting Life and the EXALT projects. We would like to thank all the people who have provided access to difficult to find literature for the compilation of the original overview. A preprint version of this article has been peer-reviewed and recommended by PCIArchaeology (https://doi.org/10.24072/pci.archaeo.100547).

PEER REVIEW

This article has been reviewed & recommended by PCI Archaeology (https://doi.org/10.24072/pci.archaeo.100547).

FUNDING INFORMATION

This research was funded by an NWO grant for the project: EXALT: Excavating Archaeological Literature – Enabling scientific discovery by applying multilingual text mining to archaeological documents (grant number: AIB.19.011). The original overview of Vlaardingen Culture sites in the Netherlands was created as part of the NWO project Putting Life into Late Neolithic Houses: Investigating domestic craft and subsistence activities through experiments and material analysis (grant number: AIB.19.020).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Lasse Van den Dikkenberg orcid.org/0000-0003-1576-1577
Faculty of Archaeology, Leiden University – Leiden, The Netherlands

Alex Brandsen orcid.org/0000-0003-1623-1340
Faculty of Archaeology, Leiden University – Leiden, The Netherlands

REFERENCES

Agapiou, A, and **Lysandrou, V.** 2023. 'Interacting with the artificial intelligence (AI) language model ChatGPT: A synopsis of earth observation and remote sensing in archaeology'. *Heritage*, 6(5): 4072–4085. DOI: https://doi.org/10.3390/heritage6050214

Amrani, A, Abajian, V and **Kodratoff, Y.** 2008. 'A chain of text-mining to extract information in archaeology'. In: *Information and communication technologies: From theory to applications, ICTTA 2008, Damascus, Syria*. Damascus. pp. 1–5. DOI: https://doi.org/10.1109/ICTTA.2008.4529905

Brandsen, A. 2021a. 'ArcheoBERTje – A Dutch BERT model for the Archaeology domain'. *Zenodo Repository*. DOI: https://doi.org/10.5281/zenodo.4739063

Brandsen, A. 2021b. *Digging in Documents – Using Text Mining to Access the Hidden Knowledge in Dutch Archaeological Excavation Reports*. Thesis (PhD), Leiden University. https://hdl.handle.net/1887/3274287

Brandsen, A. 2023. 'Information extraction and machine learning for archaeological texts'. In: Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches. Cham: Springer International Publishing. pp. 229–261. DOI: https://doi.org/10.1007/978-3-031-37156-1_11

Brandsen, A. 2024. 'Archaeology specific BERT models for English, German, and Dutch. Amsterdam'. *Proceedings of*

the CAA2023 Conference. DOI: https://doi.org/10.5281/zenodo.8296920

Brandsen, A, Lambers, K, Verberne, S and **Wansleeben, M.** 2019. 'User Requirement Solicitation for an Information Retrieval System Applied to Dutch Grey Literature in the Archaeology Domain'. *Journal of Computer Applications in Archaeology,* 2(1): 21–30. DOI: https://doi.org/10.5334/jcaa.33

Brandsen, A and **Lippok, F.** 2021. 'A burning question – Using an intelligent grey literature search engine to change our views on early medieval burial practices in the Netherlands'. *Journal of Archaeological Science*, 133: 105456. DOI: https://doi.org/10.1016/j.jas.2021.105456

Brandsen, A, Verberne, S, Lambers, K and **Wansleeben, M.** 2021. 'Usability Evaluation for Online Professional Search in the Dutch Archaeology Domain'. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2103.04437

Brandsen, A, Verberne, S, Lambers, K and **Wansleeben, M.** 2022. 'Can BERT Dig It? – Named Entity Recognition for Information Retrieval in the Archaeology Domain'. *Journal on Computing and Cultural Heritage,* 15(3): 1–18. DOI: https://doi.org/10.1145/3497842

Bulten, EEB and **Stokkel, PJA.** 2017. 'Synthese'. In: Stokkel, PJA and Bulten, EEB (eds.), *De Wateringse Binnentuinen, Gemeente Den Haag: Een vlaardingennederzetting in het Wateringse Veld*. Haagse Oudheidkundige Publicaties 20. Den Haag: Afdeling Archeologie & Natuur- en Milieueducatie Dienst Stadsbeheer. pp. 283–305.

Clason, AT. 1962. 'Nieuwe opgravingen van de neolithische nederzetting te Zandwerven, Gem.Opmeer, III'. *Varia Bio-Archaeologica,* 18: 201–219.

Cobb, PJ. 2023. 'Large Language Models and Generative AI, Oh My!: Archaeology in the Time of ChatGPT, Midjourney, and Beyond'. *Advances in Archaeological Practice*, 11(3), 363–369. DOI: https://doi.org/10.1017/aap.2023.20

Daniël, AAWJ. 2018. *Een inheems-Romeinse nederzetting in plangebied Park Waaijenstein te Nijmegen Noord. Proefsleuvenonderzoek project Npw2*. Archeologische Berichten Nijmegen 83. Nijmegen: Gemeente Nijmegen, Bureau Archeologie en Monumenten.

De Jong, JPM. 1986. 'Op de Craayenberg'. In: *Jaarverslag 1986 AWN afd. Nijmegen e.o.* pp. 24–26.

De Jong, JPM. 1988. 'Op de Craayenberg II'. In: *Jaarverslag 1988 AWN afd. Nijmegen e.o.* pp. 11–13.

De Koning, MWA. 2010. *Onder de rook van Wijchen : vondsten van de Vlaardingen-Groep, de klokbekercultuur, sporen van ovens en resten van een mogelijk grafveldje uit de late bronstijd-Romeinse tijd*. Archeologische Berichten Wijchen 11. Nijmegen: Gemeente Nijmegen, Bureau Archeologie en Monumenten.

De Ridder, JAA. 2011. *Middenneolithische vondsten aan de Schoolstraat en Molenweg te Thorn, gemeente Maasgouw Een Inventariserend Veldonderzoek in de vorm van proefsleuven*. ADC-Rapport 2908. Amersfoort: ADC.

Devlin, J, Chang, MW, Lee, K and **Toutanova, K.** 2019. 'BERT: Pre-training of deep bidirectional transformers for language understanding'. *arXiv*. DOI: https://doi.org/10.48550/arXiv.1810.04805

Diepeveen, MC and **Van Enckevort, H.** 2009. *Aan de rand van een middeleeuwse nederzetting. Archeologisch onderzoek in de bouwput van het Kulturhus in Berharen, gemeente Wijchen*. Archeologische Berichten Wijchen 9. Nijmegen: Gemeente Nijmegen, Bureau Archeologie en Monumenten.

Dorenbos, O and **Koot, H.** 2010. 'Wonen bij de kust, Vlaardingen bewoners van de strandwal'. *Westerheem,* Special 2010: Vlaardingen-cultuur (2): 96–113.

Eijskoot, YDRT. 2004. *Freegolf 6.161. Basisverslag van de archeologische begeleiding van de uitbreiding van het Freegolfterrein*. VLAK-verslag 18. Vlaardingen: Vlaardings Archeologisch Kantoor.

Eimermann, E. 2008. *Gemeente Hulst, Grote Bagijnestraat nr. 21–29. Sporen van laatmiddeleeuwse activiteiten. Een Archeologische Begeleiding en Inventariserend Veldonderzoek in de vorm van een proefsleuf*. ADC-Rapport 1225. Amersfoort: ADC.

Glasbergen, W, Groenman-van Waateringe, W and **Hardenberg-Mulder, GM.** 1967. 'Settlements of the Vlaardingen culture at Voorschoten and Leidschendam (I)'. *Helinium* 7: 4–31.

Gonzalez-Perez, C, Martin-Rodilla, P, and **Pereira-Fariña, M** (eds.). 2023. 'Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches'. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-031-37156-1

Goossens, TA (Ed.) 2009. *Opgraving Hellevoetsluis-Ossenhoek. Een nederzetting van de Vlaardingen-groep op een kwelderrug in de gemeente Hellevoetsluis*. Archol-Rapport 87. Leiden: Archol.

Gormley, C and **Tong, Z.** 2015. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. Cambridge: O'Reilly Media.

Groenman-van Waateringe, W and **Van Regteren Altena, JF.**1966. 'Voorschoten'. *Bulletin van de Koninklijke Nederlandse Oudheidkundige Bond,* 65(5): 88–89.

Habermehl, D. 2024. *Over zaaien en oogsten II. Onderzoek naar de kwaliteit en bruikbaarheid van archeologische rapporten voor synthetiserend onderzoek. Een aanvulling (2024) op eerder onderzoek (2019)*. Amersfoort: Rijksdienst voor Cultureel Erfgoed.

Heirbaut, ENA. 2010. 'Inleiding'. In: Heirbaut, ENA, Hendriks, J and Hermsen, I (eds.), *Onder een deken van zand. Archeologisch onderzoek op de Meshallen en de Kraanvogel te Wijchen*. Archeologische Berichten Wijchen 10. Nijmegen: Gemeente Nijmegen, Bureau Archeologie en Monumenten. pp. 1–18.

Hiddink, HA. 2000. *Groesbeek-Klein Amerika. Aanvullend archeologisch onderzoek van bewoning uit de prehistorie en de Romeinse tijd en een mogelijk pre-Flavisch openluchtheiligdom*. Rapportage Archeologische Monumentenzorg 74. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Janssen, AJ. 1976. 'Wychen, randweg oost'. In: *Jaarverslag 1976 AWN afd. Nijmegen e.o.* pp. 16–18.

Janssen, AJ. 1980. 'Vlaardingen-kultuur op de Blankenburg te Beuningen'. In: *Jaarverslag 1980 AWN afd. Nijmegen e.o.* pp. 28.

Janssen, AJ. 1989. 'Een neolithische nederzetting in het Ewijkse Veld'. *Westerheem*, 38(3): 133–145.

Janssen, AJ. 1993. 'Neolithische vondsten van het Woerdje aan de Ooigraaf te Beuningen'. *Westerheem* 42(6): 287–294.

Janssen, AJ and **Tuijn, WN.** 1978. 'De Homberg te Wijchen'. *Westerheem* 27(4–5): 238–255.

Koolen, M. 1976. 'Vondsten op "Hoogbroek" te Wijchen'. In: *Jaarverslag 1976 AWN afd. Nijmegen e.o.* pp. 11–15.

Lapp, EC and **Lapp, LW.** 2024. 'Evaluating ChatGPT as a viable research tool for typological investigations of cultural heritage artefacts—Roman clay oil lamps'. *Archaeometry*, 66(3), 696–717. DOI: https://doi.org/10.1111/arcm.12937

Louwe Kooijmans, LP. 1974. 'The Rhine/Meuse delta. Four studies on its prehistoric occupation and Holocene geology'. *Analecta Praehistorica Leidensia,* 7: 169–445. DOI: https://doi.org/10.1163/9789004673649

Louwe Kooijmans, LP. 1983. 'Tussen SOM en TRB, enige gedachten over het laat-neolithicum in Nederland en België'. *Bulletin van de Koninklijke Musea voor Kunst en Geschiedenis* 54(1): 55–67.

Paijmans, H and **Brandsen, A.** 2009. 'What is in a Name: Recognizing Monument Names from Free-Text Monument Descriptions'. In: Van Erp, M G J, Stehouwer, J H and Van Zaanen, M (eds.), *Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning (Benelearn).* Tilburg: Tilburg Center for Creative Computing. pp. 2–6.

Paijmans, H and **Brandsen, A.** 2010. 'Searching in archaeological texts: Problems and solutions using an artificial intelligence approach'. *PalArch's Journal Of Archaeology Of Egypt/Egyptology,* 7(2): 1–6.

Raemaekers, D. 2003. 'Cutting a long story short? The process of neolithization in the Dutch delta re-examined'. *Antiquity,* 77(298): 740–748. DOI: https://doi.org/10.1017/S0003598X00061688

Raemaekers, D. 2005. 'Het vroeg en midden Neolithicum in Noord, Midden en West Nederland'. In: Deeben J, Drenth, E, van Oorsouw, M F, Verhart, L (eds.), De Steentijd van de Nederland. Archaeologie, 11/12. Meppel: Stichting Archeologie. pp. 261–282.

Richards, J, Tudhope, D and **Vlachidis, A.** 2015. 'Text Mining in Archaeology: Extracting Information from Archaeological Reports'. In: Barcelo, J A and Bogdanovic, I (eds.), *Mathematics and Archaeology*. Boca Raton: CRC Press. pp. 240–254. DOI: https://doi.org/10.1201/b18530-15

Rijksdienst Cultureel Erfgoed. 2023. *Erfgoedmonitor 2023 – Feiten en cijfers over ons erfgoed in 2023*. Amersfoort: RCE.

Stokkel, PJA. 2017. 'Sporen en structuren'. In: Stokkel, P J A and Bulten, E E B (eds.), *De Wateringse Binnentuinen, Gemeente Den Haag: Een vlaardingennederzetting in het Wateringse Veld*. Haagse Oudheidkundige Publicaties 20. Den Haag: Afdeling Archeologie & Natuur- en Milieueducatie Dienst Stadsbeheer. pp. 53–100.

't Hart, R, Norde, EHLD and **Tuinstra, JWD.** 2019. *Prehistorische bewoning langs een geul in het plangebied Hekkeslag in Ewijk, gemeente Beuningen; archeologisch onderzoek: een opgraving.* RAAP-Rapport 3679. Weesp: RAAP.

Teubner, B and **Tuijn, W.** 2010. 'Vlaardingen-cultuur in Wijchen en omgeving'. *Westerheem*, Special 2010: Vlaardingen-cultuur (2): 180–185.

Tjong Kim Sang, EF. 2002. 'Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition'. *arXiv*. DOI: https://doi.org/10.3115/1118853.1118877

Van Beek, BL. 1990. *Steentijd te Vlaardingen, Leidschendam en Voorschoten. De vondstverspreiding in laat-neolithische nederzettingen in het hollandse kustgebied Universiteit van Amsterdam*. Unpublished thesis (PhD), Universiteit van Amsterdam.

Van den Broeke, PW. 1993. 'A crowded peat area; observations in Vlaardingen-West and the Iron Age habitation of southern Midden-Delfland'. *Analecta Praehistorica Leidensia,* 26: 59–82.

Van den Dikkenberg, L. 2024. 'Vlaardingen culture or Stein group? From object biographies to communities of practice'. In: Verbaas, A, Langejans, G, Little, A and Chan B (eds.), *Artefact Biographies from Mesolithic and Neolithic Europe and Beyond. Papers in honour of Professor Annelou van Gijn.* Analecta Praehistorica Leidensia, 52. pp. 63–72.

Van Gijn, AL. 2010. Het gebruik van vuursteen in de Vlaardingen-tijd, Westerheem, Special 2010: Vlaardingen-cultuur(2): 83–89.

Van Gijn, AL and **Bakker, JA.** 2005. 'Megalith builders and sturgeon fishers. Middle Neolithic B: Funnel Beaker culture and Vlaardingen group'. In: Louwe Kooijmans, LP, Van den Broeke, PW, Fokkens, H and Van Gijn, AL (eds.), *The Prehistory of the Netherlands*. Amsterdam: Amsterdam University Press. pp. 281–306.

Van Iterson Scholten, FR. 1988. 'Inventory and Protection of the Archaeological Monument of Zandwerven'. *Helinium* 28(2): 232–243.

Van Kampen, JCG. 2013. 'Bewoning in het Neolithicum'. In: Van Kampen, JCG and Van den Brink, V (eds.), *Archeologisch onderzoek op de Habraken te Veldhoven. Twee unike nederzettingen uit het Laat Neolithicum en de Midden Bronstijd en een erf uit de Volle Middeleeuwen.* Zuidnederlandse Archeologische Rapporten 52. Amsterdam: VUhbs archeologie. pp. 39–61.

Van Noort, G. 1998. 'De Neolitische boeren van de Vlaardingen-cultuur op Texel'. *Uitgave van de Historische Vereniging Texel*, 48: 3–7. DOI: https://doi.org/10.2307/939154

Van Regteren Altena, JF. 1958. 'Nieuwe opgravingen van de Neolithische nederzetting te Zandwerven gemeente Zandbroek'. *West Frieslands Oud en Nieuw* 25: 144–159.

Van Regteren Altena, JF and **Bakker, JA.** 1961. 'De neolithische woonplaats Zandwerven (N.H.)'. In: Glasbergen, W and Groenman-van Waateringe, W (eds.),

In het voetspoor van van A.E. van Giffen. Groningen: J.B. Wolters. pp. 33–40.

Van Regteren Altena, JF, Bakker, JA, Clason, AT, Glasbergen, W, Groenman-van Waateringe, W and **Pons, LJ.** 1962a. 'The Vlaardingen culture I'. *Helinium* 2: 3–35.

Van Regteren Altena, JF, Bakker, JA, Clason, AT, Glasbergen, W, Groenman-van Waateringe, W and **Pons, LJ.** 1962b. 'The Vlaardingen culture II'. *Helinium* 2: 97–103.

Van Regteren Altena, JF, Bakker, JA, Clason, AT, Glasbergen, W, Groenman-van Waateringe, W and **Pons, LJ.** 1962c. 'The Vlaardingen culture III'. *Helinium* 2: 215–243.

Van Zoolingen, RJ. 2021. 'Sporen en Structuren'. In: Van Zoolingen, RJ and Bulten, EEB (eds.), *Steynhof (Anne Frankpark) Gemeente Den Haag Een Vlaardingennederzetting in het Wateringse Veld*. Haagse Oudheidkundige Publicaties 23. Den Haag: Afdeling Archeologie & Natuur- en Milieueducatie Dienst Stadsbeheer. pp. 45–82.

Van Zoolingen, RJ and **Bulten EEB.** 2021. 'Synthese'. In: Van Zoolingen, RJ and Bulten, EEB (eds.), *Steynhof (Anne Frankpark) Gemeente Den Haag Een Vlaardingennederzetting in het Wateringse Veld*. Haagse Oudheidkundige Publicaties 23. Den Haag: Afdeling Archeologie & Natuur- en Milieueducatie Dienst Stadsbeheer. pp. 291–309.

Van Zoolingen, RJ and **Rieffe, C.** 2023. *Noordweg 76 Gemeente Den Haag Sporen uit het neolithicum, de ijzertijd en Romeinse tijd in het Wateringse Veld*. Haagse Archeologische Rapportage 2301. Den Haag: Afdeling Archeologie & Natuur- en Milieueducatie Dienst Stadsbeheer.

Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, Kaiser, Ł and **Polosukhin, I.** 2017. 'Attention is all you need'. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

Verhart, LBM. 1992. 'Settling or trekking? The late Neolithic house plans of Haamstede-Brabers and their counterparts'. *Oudheidkundige Mededelingen uit het Rijksmuseum van Oudheden te Leiden* 72, 73–99.

Verhart, LBM. 2010. 'De zuidelijke tegenhangers van de Vlaardingen-cultuur: een schamel spoor van scherven'. *Westerheem*, Special 2010: Vlaardingen-cultuur (2): 207–222.

Verhart, LBM and **De Ridder, T.** 2010. 'Een waaier van Vlaardingen'. *Westerheem*, Special 2010: Vlaardingen-cultuur (2): 4–11.

Vlachidis, A and **Tudhope, D.** 2012. 'A pilot investigation of information extraction in the semantic annotation of archaeological reports'. *International Journal of Metadata, Semantics and Ontologies* 7(3): 222–222. DOI: https://doi.org/10.1504/IJMSO.2012.050183

TO CITE THIS ARTICLE:
Van den Dikkenberg, L and Brandsen, A. 2025. Using Text Mining to Search for Neolithic Vlaardingen Culture Sites in the Rhine-Meuse-Scheldt Delta. *Journal of Computer Applications in Archaeology*, 8(1): 110–124. DOI: https://doi.org/10.5334/jcaa.205

Submitted: 10 February 2025 **Accepted:** 13 February 2025 **Published:** 24 March 2025

COPYRIGHT:
© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Journal of Computer Applications in Archaeology is a peer-reviewed open access journal published by Ubiquity Press.