Data exchange protocol in Dutch archaeology

Wansleeben, M.; Laan, W.; Visser, R.; Kalaycı, T.; Lambers, K.; Klinkenberg, V.

Kalaycı, T., K. Lambers and V. Klinkenberg. 2023. "Digital Archaeology: Promises and Impasses". *Analecta Praehistorica Leidensia* 51. Leiden: Sidestone Press, pp. 33-46. DOI: 10.59641/f48820ir.

Data Exchange Protocol in Dutch Archaeology

Milco Wansleeben, Walter Laan and Ronald Visser

1 Introduction

Around 15 years ago, a problem in Dutch archaeology started to become more and more apparent. Owing to the new Malta legislation, the number of archaeological fieldwork projects being conducted by commercial archaeology units in the Netherlands rose precipitously. The units were obliged to deposit their finds and documentation at the designated archaeological depot in the province or municipality within two years after finishing a project. Many of these units operated nationally and were confronted with a wide variety of rules and regulations connected to this deposit process. For each and every depot, they were required to provide different documents. Attempts in the early 2000s to promote the use of a single data management system at all depots had failed (Boogert 2006, 26; Taayke 2003). An exploratory study (Sueur *et al.* 2008) suggested the introduction of a single standardised XML data exchange format for the intermediate collection of data about the items delivered by any archaeology unit to any depot. This delivery slip ('*pakbon*'), also known as the SIKB0102 protocol, was introduced in around 2011 but was not widely adopted for a long time following its introduction. Only a few early adopters among the commercial units had prepared their excavation management software for the export of an XML exchange document.

From a relatively recent report, "Verder graven in depots" (Erfgoedinspectie 2018, 28):

"Wat betreft de pakbon geldt hetzelfde als bij de provinciale depots: er worden vrijwel geen pakbonnen aangeleverd. En er is slechts één gemeentelijk depot dat de pakbon kan inlezen. Veel gemeentelijke depotbeheerders zien geen meerwaarde in de pakbon voor het eigen depot. Degenen die sympathiek staan tegenover het oorspronkelijke idee van één standaard voor aanlevering van vondsten en documentatie, vinden dat de uitwerking daarvan in de huidige vorm niet voldoet. Enkele depots stellen desalniettemin het leveren van een pakbon verplicht, maar hebben daarbij ervaren dat slechts weinig bedrijven een pakbon (kunnen) leveren.".

"For the municipalities, the situation with the digital exchange protocol is more or less the same as with the provinces: these are hardly ever used. Only one municipality depot is capable of importing the SIKB0102 document. Many managers from the municipality depots don't feel the digital delivery slip offers additional value for their organization. Some embrace the original idea of having one standard document for the delivery of finds and documentation, but consider the implementation in its current form inadequate. Although a few depots have made the protocol compulsory, experience has shown that only a few commercial units are able to provide it."

Milco Wansleeben
Faculty of Archaeology
Leiden University
PO Box 9514
2300 RA Leiden
The Netherlands
m.wansleeben@arch.leidenuniv.nl

Walter Laan
Archol bv
Einsteinweg 2
2333 CC Leiden
The Netherlands
wlaan70@gmail.com

Ronald Visser
Saxion University of Applied Sciences
Handelskade 75
7417 DH Deventer
The Netherlands
r.m.visser@saxion.nl

This situation changed quickly after the province of Noord-Brabant (*Provinciaal Depot Bodemvondsten*) took the initiative to develop software that could read the SIKB0102 protocol, in a collaborative effort with a number of provincial archaeological depots (Meffert 2018). From that moment onwards, the supply and demand for this digital data exchange increased, and depots made its use compulsory for new projects. Unfortunately, many commercial units have since struggled with the implementation, encountering many practical obstacles and troublesome discussions with the provincial depots. The technical implementation of the SIKB0102 protocol turns out to be so complicated that many 'average' archaeologists are dropping out. This illustrates a huge technological gap between the 'whiz kids' in the SIKB0102 technical committee and the archaeologists performing the actual work at archaeological research projects and at the depots. Many archaeologists consider the exchange format to be a black box, are unable to understand the content and cannot solve the problems that arise in the process of depositing finds and documentation.

Archol BV is one of the commercial archaeological units in the Netherlands that was an early adopter of this data exchange protocol. Over the years, several versions of the protocol have been implemented in its excavation registration system, which is based on Microsoft Access. Other Dutch archaeological organizations regularly ask Archol for assistance in the use and implementation of the SIKB0102 protocol. Based on these experiences, this article tries to answer a few questions:

- Were there any mistakes made during the introduction of SIKB0102?
- What is the exact nature of the problem(s)?
- What are possible solutions?
- Which components are still needed in order to make the *pakbon* a success?

2 KNA and SIKB0102

Dutch archaeology has seen some major changes since the Malta treaty was introduced (1992) and implemented (2006) in the Netherlands through what has been referred to since 2015 as the *'Erfgoedwet'* (integral cultural heritage law). Today, many (semi-)commercial, governmental, and public organizations are involved in heritage management. These organizations are involved in a wide range of tasks, from providing advice to municipalities, undertaking process management, heritage mapping, surveying, conducting excavations, carrying out material analysis, and engaging in knowledge dissemination. In order to maintain minimal quality requirements, a set of protocols has been agreed upon for each of the steps in the heritage trajectory. These requirements are laid out in the quality standard for Dutch archaeology (KNA: *Kwaliteitsnorm Nederlandse Archeologie*). These protocols specify tasks, documents, and personnel for archaeological research both on land and under water. In addition, organizations that perform excavations or any other type of research that destroys the soil archive must be certified and registered. These protocols are maintained by the Dutch archaeological community and are updated every few years (version 4.1 is currently in force). A foundation for quality assurance in soil management (SIKB: *Stichting Infrastructuur Kwaliteitsborging Bodembeheer*) organizes the workflow and maintains and publishes these standards and protocols on its website (www.sikb.nl/archeologie).

The protocol 4004 (Excavations on Land) also includes the required steps for depositing the finds, samples, and documentation at the end of the archaeological process (OS17). Within the Netherlands, there is an agreement that archaeologists will publish their results within two years after finishing the fieldwork, although this is not always possible, especially when it comes to large-scale excavations. During the pre-publication period, the archaeological objects and (digital and analogue) documentation will remain stored at the excavating organization. After the results are published, the objects and documents are to be transferred to the designated depot at a province or municipality. The province or municipality is assigned as the legal owner of the archaeological discoveries by the *Erfgoedwet*. To make this deposit process as smooth as possible, a number of inventory lists must be presented to the depot beforehand. These include a list of boxes containing the finds or samples, a list of field drawings, a list of digital files (including the digital photos), and a list of binders and reports. With these lists, the depot knows what they will receive and can check whether the documentation is properly done. In the past, individual depots had their own sets of requirements for these inventory lists, but this has now been harmonised in this digital delivery slip ('*pakbon*'). The *pakbon* (as part of OS17) is a digital document with a fixed, prescribed structure and standardised content. Additional information has been added to this delivery slip over the years, including the description

The Forum on Information Standards in Heritage (FISH) describes a standard and vocabulary (MIDAS Heritage) for a uniform description of the historic environment (www.heritage-standards.org.uk/midas-heritage/). In 2004, a FISH Interoperability Toolkit was developed by Oxford-ArchDigital that serves as an aid for efficient data exchange and archiving via a MIDAS XML document. In 2012 this was adapted by the Archaeology Data Service. The toolkit included tools for mapping data schemes and vocabularies to the Midas Heritage standard and an XML validator (www.heritage-standards.org.uk/fish-interoperability-toolkit-archived/). From 2015 onwards, this XML toolkit is no longer supported, owing to the lack of implementation.

Figure 1: The FISH interoperability toolkit is unfortunately no longer supported.

of non-physical items such as archaeological features, structures, and persons involved. This document has grown into more of a general-purpose data exchange document, making it very versatile and ready for the exchange of any archaeological information between any archaeological organizations. It is implemented as an XML document and referred to as the SIKB0102 protocol.

The original possibility for data exchange between archaeological organizations in the Netherlands was put forward in 2008 in an exploratory study in the Vestigia Rapport V513 (Sueur *et al.* 2008; Verhagen *et al.* 2011). This was initiated by the SIKB and inspired by the SIKB0101 protocol that already existed for soil technical engineers and the FISH protocol (figure 1) that existed in the UK for archaeologists (Gilman and Newman 2007).

The SIKB0102 protocol was introduced in the KNA in 2011 (SIKB 2011). Formally, it became a compulsory format in Dutch archaeology for new projects from that point onwards (www.forumstandaardisatie.nl/open-standaarden/verplicht). However, only a few archaeological organizations immediately implemented it (voluntarily) in their day-to-day practice. Many were hesitant for a long time, for the reasons discussed below.

3 SIKB0102 XML file format

The data exchange format (protocol SIKB0102) has been carefully designed over a long period of time,

```
<sikb0102>
    <project>
        <project_name>
        <start_date>
        <project_identifier>
        <research_type>
        <description>
        <authority>
            <person_identifier>
        <contractor>
            <person_identifier>
        <designated_depot>
```

Figure 2: XML-like snippet with selected and translated elements (tags) of the start of the exchange format, without namespace.

```
<easymetadata>
    <titles>
        <title>
            <alternative_title>
    <subjects>
        <subject>
    <dates>
        <date>
        <date_created>
        <date_available>
        <date_submitted>
    <coverage>
        <spatial>
            <point>
                <x>
                <y>
        <temporal>
```

Figure 3: XML-like snippet with simplified (adjusted) version of some of the tags of the DANS EASY dataset metadata, without namespaces.

following multiple design sessions with a group of mainly ICT-trained 'computer' archaeologists. The initial structure (Sueur *et al.* 2008) is still visible in the first part of the XML document, the project description (figure 2). Here, information about the project and actors involved is stored in a hierarchical structure that is almost readable by humans.

```
<archis3>
    <events>
        <event>
            <event_identifier>
            <event_start_date>
            <event_methodology_code>
            <event_methodology_label>
            <find_locations>
                <find_location>
                    <municipality>
                    <toponym>
                    <x-coordinate>
                    <y-coordinate>
                    <finds>
                        <find>
                            <amount>
                            <material>
                            <type>
                            <start_date>
                            <end_date>
                    <features>
                        <feature>
                            <amount>
                            <type>
```

Figure 4: XML-like snippet with selected (translated) tags from the Archis 3 project event data, as stored in the national Archis 3 database at the RCE.

```
<sikb0102 id="3307ce94-7a05-4185-9f0f-2de69f3f88a1" version="4.1.0">
    <project id="ae628cde-9bd7-4c4b-9289-7138f0c0bb43">
    <organization id="449ad8b2-b1dd-41c9-983e-39e3d22b077f">
    <person id="0ab1d5f5-fa03-4571-b6f4-406deb2222bc">
    <person id="e9abc724-9805-449d-b657-2d2ca7643f7f">
    <person id="bc3dc3ec-ef03-424f-89a1-0f732d9799fd">
    <file id="04785ae0-7d5f-43b9-9bc4-d55bb9aa5617">
    <file id="f17b5cce-d2de-4614-887c-a66dde17ae89">
    <file id="89912ec1-6722-44a1-82fd-fd3154ced52b">
    <find id="e77038a6-2285-422d-8340-ccebb037db17">
        <source_identifier>
        <context_UUID>
        <box_UUID>
    <find id="4aa5f7f5-7919-4c4d-9300-afa62a8f8bdd">
    <find id="b59c92e9-0924-464a-bfb9-0b2591fa0528">
    <feature id="8f0da90c-7aba-4a7d-a8e0-b1881cbce351">
    <feature id="7752c3d8-a565-4d5b-a5b6-551c628bcd9f">
    <feature id="12de5b57-8752-4705-aafd-9949a306249b">
    <photo id="c222573c-7e3d-4e92-bca1-98c14d17a247">
    <photo id="74d8d0ea-4662-458d-83f9-677bbb93ec79">
    <photo id="1a7f30c0-44c5-4eee-9f08-6db3482b317c">
```

Figure 5: XML-like snippet with selected (translated) parts of the SIKB0102 protocol to illustrate the object-oriented characteristics.

This part of the document can be read by archaeologists once opened in a web browser, an advanced text editor (like Notepad++), or a special XML editor. It is easy to find information, identify errors, and make corrections.

Many (archaeological) organizations that use XML for data exchange use this hierarchical (nested) property of XML, in order to make the often-complex structure of the information manageable. Two examples from Dutch archaeology are presented below: figure 3 is the metadata export from a dataset in the DANS data archive (EASY), and figure 4 contains a project description from Archis 3 at the Cultural Heritage Agency (RCE).

Both archaeological examples show a hierarchy that represents the underlying data model. Archis, for instance, stores research events that can result in 0, 1, or many find locations, which can yield (any number of) finds and/or archaeological features. This is a hierarchy that reflects the 1-to-many relationship between tables, which is familiar from relational database designs. It is immediately clear to which find location, discovered in which event, a find belongs. Most web browsers even allow end users to expand or hide some of the (nested) branches of the tree to obtain an overview more quickly in the often extensive XML documents.

The simple hierarchical model shown in the XML examples above could have been a template for the SIKB0102 format: starting with a single (main) branch for the project, and branching out to several major entities such as find boxes, documentation, and find locations. A branch for the find locations could have included further (sub-)branches for samples, finds, and features. A branch for the documentation could have included (sub-)branches with attributes about (field) drawings, photos, reports, and files. How such a hierarchy would actually be implemented in the XML does depend on arbitrary decisions, given the recognised fact that archaeological data does not follow a single simple hierarchy (Boasson and Visser 2017). For instance, finds are not only related to find locations; they are also stored in boxes and illustrated on (field) drawings and photos. Features are documented through photos, where information about the same features is part of cross-section drawings and associated with surveying (X, Y, Z) data. The various entities documented during an archaeological project have a large number of potential relationships. The information is therefore generally stored in a relational database that enables us to dynamically create links between the various parts of our data whenever needed by using a query. A static XML structure would hold only one of the possible representations.

The design team of the SIKB0102 protocol therefore wanted to combine a more relational database- and an object-oriented approach. They used a technical structure that is very similar to the way Linked Open Data (LOD) is modelled. The XML document contains, for the most part, a (very) long list of all possible information objects, in more or less alphabetical order. It simply lists all of the find boxes, along with their properties, followed by flat lists of all the finds, corings, drawings, photos, people, organizations, features, structures, and so on. Each information object is uniquely identified with a UUID (Universally Unique IDentifer). This identifier serves as the key for creating relationships between the information objects. A find will have its own UUID and properties for an (indirect) link to the discovery location (context_UUID) and to the box in which it is now stored (box_UUID) (figure 5). This object-oriented data model does provide more flexibility in coping with archaeological data from a wide variety of excavations (Boasson and Visser 2017). One of the first to adapt this in archaeology was the Swedish Intrasis software (www.intrasis.com) from the Swedish National Heritage Board, a general-purpose excavation documentation system. The same principles were adapted in the Netherlands in, for example, the ArcheoLINK software from the private company QLC BV (nowadays part of TijdLab) and the ODILE software from the archaeological company RAAP BV.

One of the downsides of this approach is that the (main) hierarchy (relationships) is not explicitly included in the structure of the XML document. This means that to the average archaeologist, this document is unreadable and is just an arbitrary flat list of entities that are interconnected in an inextricable knot by their arbitrary (UUID) identifiers, which archaeologists no longer recognize as the find number printed on the find label or the filename of the digital photo.

4 Practical problems

Is this object-oriented (Linked Open Data-like) model a (big) problem? Some archaeologists feel it isn't, as the data exchange document is not intended for humans to read. Rather, it is a protocol for facilitating communication between a computer and a computer. Commercial units should invest in an additional piece of software that creates the proper XML document, and the depots should likewise invest in the import

software. However, there are a few arguments that complicate the matter, as explained below.

In many cases, there is a need for personal communication between employees of the commercial unit and the depot about the content of the XML document. This could arise, for example, in the case used in a request for additional information (*e.g.,* this object is used in the publication, but not included in the *pakbon*), an explanation of why some samples are no longer stored in one of the boxes, or a request to correct a few (minor) mistakes. These discussions regularly become laborious and troublesome. The depots have problems defining the exact issue (*e.g.,* there is "something" wrong with the samples, since the import tool is rejecting the *pakbon*), and the commercial unit has trouble pinpointing the records that must be added or changed in order to solve the problem. The XML does not provide a simple and clear basis for the discussion (*e.g.,* "here at line 34567, a value is used that is not allowed"). As a result, both sides send many messages back and forth to explain things in more detail, until frustration arises on both sides: "You made these changes, but now something else has become incorrect…"

The conversion to an object-oriented model from a relational database is not easy, nor is it familiar to many Dutch archaeologists. It requires highly specialised knowledge about:

- technical issues, such as conventions used in UML, XML, XSD, namespaces and the use of UUIDs. The conversion is not a matter of a few simple queries. Rearranging the data tables often also requires some programming skills (*e.g.,* SQL, VBA).
- the actual archaeological content and how the commercial units' tables and fields correspond to the SIKB0102 entities. To comply with the SIKB thesauri (enumerations), a detailed mapping (concordance) with the prescribed fields and values is required.

The required investments (in terms of both time and money) are substantial, and many Dutch archaeological organizations struggle with that.

- Many commercial units are relatively small companies, with material specialists at one end of the scale, since they often operate as single-handed (sub)contractors. They sometimes even lack the knowledge to create proper databases and work in a less formal manner with small datasets in spreadsheets. Some smaller units have already joined forces to invest collaboratively in a database system with a SIKB0102 export.
- Even some larger archaeological organizations, such as the RCE and the University of Groningen, have recently asked Archol BV for assistance in implementing the SIKB0102 protocol.
- Provincial depots also struggle to make the required investment individually. On the initiative of the provincial depot in Noord-Brabant, a shared system called the 'Archeodepot' was created (www.archeodepot.nl) (figure 6). Eight provinces and DANS work together in this initiative to provide a single counter for commercial units to upload their SIBK0102 documents prior to depositing their finds and (digital) documentation. Together, they have the necessary critical mass and have succeeded in providing a working system.
- Municipality depots are smaller than the provincial depots and are currently very reluctant to implement the SIKB0102 protocol at all. They are waiting for a (financial) opportunity to create the tools for themselves or to join the Archeodepot.
- Only a few software developers in the Netherlands produce an archaeological (excavation) documentation tool that can create the exchange document. These packages can be bought off the shelf by commercial units, but often at a relatively high cost or with the need for major changes to their daily archaeological practices.

It appears that the expertise required to implement the SIB0102 protocol in its current form is simply too high for the average archaeologist. They must therefore leave this to a few digital experts and start viewing the SIKB0102 document as a black box that is difficult to create and to read. Some organizations prefer not to use the protocol at all ("we just ignore it, until we are forced to use it") or rely on the occasional (almost) free tools that are made available.

After paying a yearly financial contribution, the SIKB provides access to a validation tool. This enables organizations to check the structure and content of the XML document before sending it to the Archeodepot. This validator prevents many mistakes such as the use of the wrong enumeration version, incorrect tag order, or missing UUIDs. However, the tool also creates (technical) error messages that the average archaeologist finds difficult to understand.

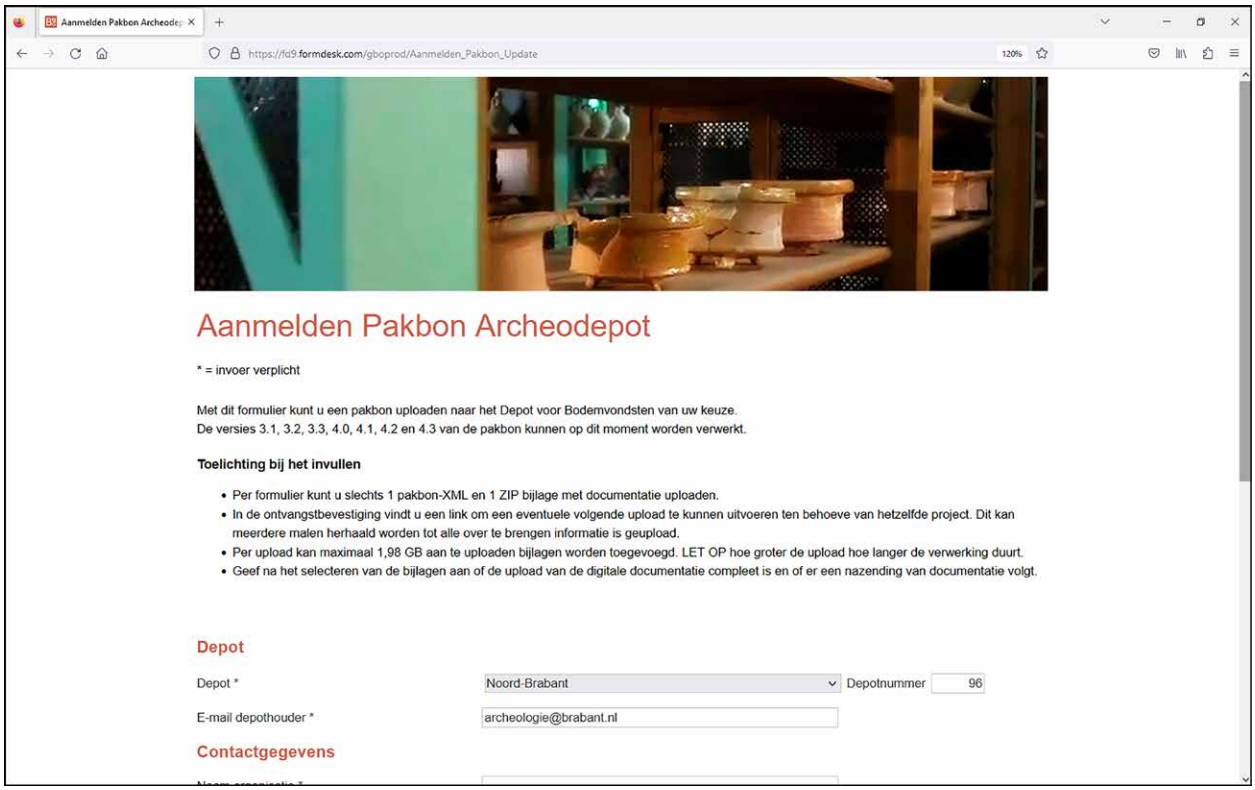

Figure 6: The splash screen of the national counter for depositing the SIKB0102 XML exchange documents. This system validates the documentation before it is forwarded to the designated province.

Under the same conditions, the SIKB also provides a reporting tool that converts a valid XML exchange document into an Excel spreadsheet. This visualizes the XML document's content in a number of separate worksheets that reflect the groups of items in the SIKB0102 model. However, this tool is far from complete. Some municipality depots use this tool to view the content, but this has led to complaints that something is missing in the XML document, although in fact the rendering is incomplete. The reporting tool does not come close enough to the way archaeologists are used to working with their (excavation) data.

Altogether, this situation is arguably a dangerous deadlock that could threaten the implementation of the SIKB0102 protocol. The collaborative initiative taken by the eight provincial depots has probably prevented the Dutch SIKB0102 from suffering the same fate as the UK's FISH Toolkit. Still, many improvements are needed to make the SIKB0102 a real success. Many more actors in Dutch archaeology, such as material specialists and municipality depots, need to board this train in the near future.

5 Possible solutions

If the SIKB0102 document had a simple (albeit arbitrary) hierarchical structure, some of the discussions would probably have been easier. Any web browser would have been sufficient to interactively render the XML into a readable document for humans, providing a solution with a very low threshold. However, making fundamental modifications to the XML structure now seems out of the question. We have already passed a point of no return, given that too many large organizations have already implemented the document. And in any case, this discussion might not be beneficial at all, as the object-oriented data model does have the flexibility to store and share the complexity of the archaeological data (Boasson and Visser 2017). We must therefore consider other options. Here we suggest some possible improvements.

The documentation provided by the SIKB about the protocol is sufficient (UML, XSD, example files) from a technician's point of view. For the average archaeologist, who is usually not really interested in an Entity-Relation model of the database, let alone a UML

The SIKB0102 XML document is a simple text file. That means that in principle it can be opened in any text processing software, from a web browser, a simple Windows Notepad or Wordpad, through an advanced text editor such as NotePad++, to more specialised XML editors like XMLSpy or XmlNotepad. Some software is better suited for reading large XML files and rendering the XML elements in a user-friendly way than others.

An XML element always consists of a start tag <> and an end tag </>, like one of the first elements in the SIKB0102 document that stores the name of the project:

```
<sikb:projectnaam>Oegstgeest – Nieuw Rhijngeest-Zuid</sikb:projectnaam>
```

The starting tag can have additional attributes, for instance for identification, and tags can be grouped (nested) within other tags, as in the following example with a few attributes of a drawing in the SIKB0102 structure.

```
<sikb:tekening bronId=”OEGR12_T051” sikb:id=”a2012515-e902-4360-9e6c-85e8a53f1aa6”>
    <sikb:naam>Detail Coupe 445</sikb:naam>
    <sikb:tekeningtype>DET</sikb:tekeningtype>
    <sikb:tekeningmateriaal>houtvrij mm papier</sikb:tekeningmateriaal>
</sikb:tekening>
```

Sometimes the content between the start and end tag is free text; sometimes it is restricted to one of the prescribed values (enumerations). The allowed values for the type of drawing can be found at codes.sikb.nl, where DET stands for a “detail drawing”.

To find the box with this special find from the early Medieval excavation at Oegstgeest in the *pakbon*, follow the following steps:

- Open the XML file in your browser (or text editor of choice)
- Use edit – find (Ctrl+F) or equivalent to search for: “zilveren schaal” (silver bowl)

```
<sikb:vondst bronId=”OEGR12_V002.001” sikb:id=”575ba625-69a4-4149-8b70-34cb3b9ff585”>
    <sikb:naam>zilveren schaal</sikb:naam>
    <sikb:veldvondstId>906d115d-a7b3-44a3-b4e2-10e696030e16</sikb:veldvondstId>
    <sikb:aantal>1</sikb:aantal>
    <sikb:materiaalcategorie>MAG</sikb:materiaalcategorie>
    <sikb:beginperiode>MEVB</sikb:beginperiode>
    <sikb:eindperiode>MEVB</sikb:eindperiode>
    <sikb:exposabel>true</sikb:exposabel>
    <sikb:verpakkingseenheidId>c9f95c57-2c29-41d9-8dde-d44804ba6d49</sikb:verpakkingseenheidId>
</sikb:vondst>
```

The original find number of the excavator is documented in the attribute bronId as part of the sikb:vondst tag, while the context of the find is recorded in the sikb:veldvondstId tag and the storage location is stated in the sikb:verpakkingseenheidId tag, both with their own UUID.

- Copy the UUID from the verpakkingseenheidId tag
- Perform a new search throughout the entire document with that UUID

This will provide you with the packaging unit (usually the plastic find bag) and its attributes.

```
<sikb:verpakkingseenheid bronId=”OEGR12_V002” sikb:id=”c9f95c57-2c29-41d9-8dde-d44804ba6d49”>
    <sikb:naam>OEGR12_V002</sikb:naam>
    <sikb:doosId>8f0da90c-7aba-4a7d-a8e0-b1881cbce351</sikb:doosId>
</sikb:verpakkingseenheid>
```

- Use the UUID from the sikb:doosId tag once more to find the storage box

```
<sikb:doos bronId=”OEGR12_D002” sikb:id=”8f0da90c-7aba-4a7d-a8e0-b1881cbce351”>
    <sikb:naam>OEGR12_D002</sikb:naam>
    <sikb:breekbaar>true</sikb:breekbaar>
</sikb:doos>
```

As can be seen from the above, a small series of simple text find requests can bring up the box of this special find, in this case OEGR12_D002, without the need for special software.

(Opposite page and above) Figure 7: Mock-up of a part of a manual that can help non-digital specialists to use and identify items in the SIKB0102 XML document.

document, this doesn’t help much. Our first suggestion is therefore, that a clear manual is produced covering the possibilities and limitations of the protocol in simple words: What is actually in the XML, and where can I find it? This would be a manual with a strong visual component, breaking down the protocol into entities with which archaeologists are familiar, such as boxes-and-finds or features-and-photos (figure 7).

The SIKB reporting tool is not ideal, nor is it free for everyone. Both conditions could be improved if

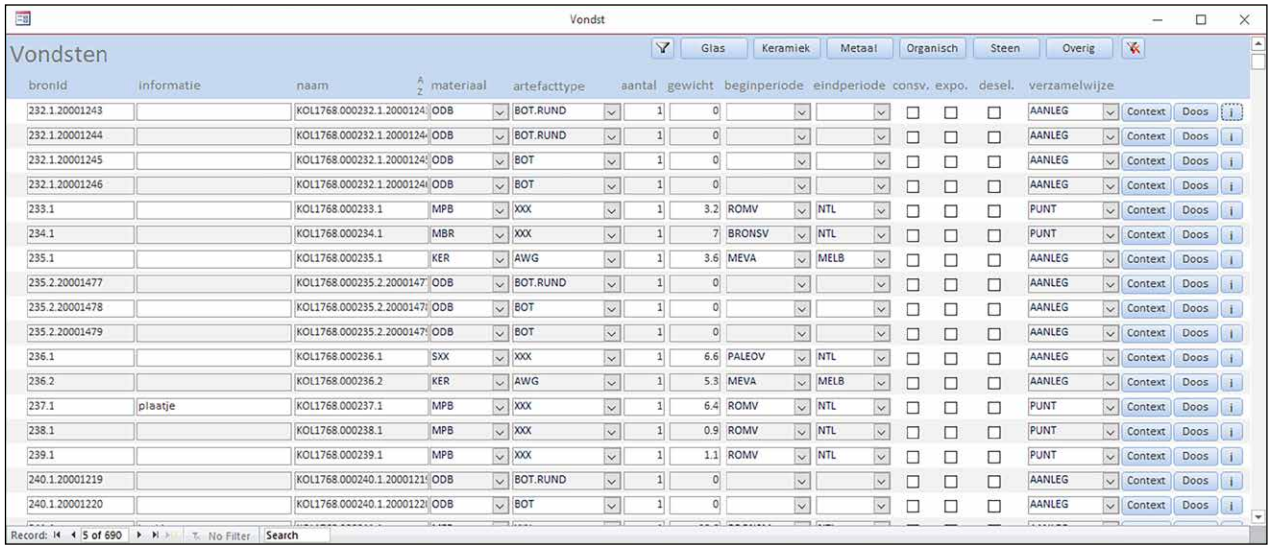

Figure 8: A prototype database (proof of concept) has been created with Microsoft Access. This database imports a pakbon and renders it in several ways through hierarchical forms, in this case starting with the finds.

the SIKB is willing to put in the required effort (in the near future). As a second suggestion, a more full-scale viewer should be developed, one that uses a database instead of a spreadsheet as a basis. Such a database would allow for a user interface that visualizes the entities in a way that is very close to archaeologists' day-to-day practice (figure 8). A database would allow for multiple relationships to be presented (boxes – finds, finds – context, context – photos, photos – files). Such a standardised viewer would probably be very helpful for guiding and streamlining the discussions between employees of the depositor and depot. Any problems could likely be identified (interactively) and solved much more quickly if both parties were using the exact same visualization.

Such a tool should be open-source, cross-platform, freely available to all, and maintained for future releases of the SIKB0102 protocol and database software used. Of course, some (financial) arrangements would need to be established to achieve this goal.

Dutch archaeology contains many small, single-handed companies that carry out small archaeological projects with a limited number of finds, photos and data tables. A small commercial unit might for instance be commissioned to check the archaeological potential of a small area in a private garden with two or three corings. These small-scale projects occur frequently and often have a short turnaround time, minimum dataset and standardised report. It would probably be very effective to have a tool that directly stores the available data in a small database that exactly mimics the entities in the SIKB0102 protocol. Such a free, open-source data entry tool would probably also serve to educate a much wider group of archaeologists about the content of the XML. It would become simply a matter of "learning by doing".

A few simple additions could potentially take the SIKB0102 out of the realm of technicians and bring it into the world of many Dutch archaeologists. Where, in our view, priority should be given to the creation of an open-source toolkit to visualize the content of an XML exchange document (option 2).

6 Additional bonus: SIKB0102 as an archiving format

An XML document is basically a simple text file, with a simple markup structure. Each element has a starting tag (*e.g.,* <project>) and ending tag (</project>). Elements can be nested to as many levels as required, as long as the document remains well formed. The content of an XML document is (should be) documented in an XSD (XML Schema Definition Language) document, which prescribes the tags in great detail, for example: which tag names, in which order, compulsory or not, multiple values allowed or not, with a predefined value (enumeration) or free text. These characteristics make XML a very promising candidate for the long-term archiving of

DANS DATASET DESCRIPTION

The SIKB maintains data exchange protocols for both soil researchers and archaeologists. The protocol for Archaeology (SIKB0102) has been in place since 2011 (*) and has seen many minor and major revisions. The SIKB makes only the most recent versions of the protocol available on its website. The SIKB's view is that the exchange documents are used only once, *i.e.,* at the moment an excavation is finished and the finds and documentation are deposited at one of the provincial or municipality depots.

However, a data exchange document ('*pakbon*') brings together data about many aspects of an excavation in a highly standardised manner. Although this is probably not as detailed and rich as the original databases and (GIS) files from the excavator, it is and will remain a valuable resource for future archaeologists. The XML documents are stored in the repositories as separate files. In order to understand and reuse such a document, the technical description of the specific version of the protocol should be available at any time. This dataset brings together that metadata, for as many different versions of the *pakbon* as (still) possible.

(*): SIKB 2011, Wijzigingsblad KNA 3.2 Landbodems, Stichting Infrastructuur Kwaliteitsborging Bodembeheer, Gouda.

Figure 9: Dataset description at the DANS EASY repository of the (historic) documentation of the SIKB0102 protocol (doi.org/10.17026/dans-zbn-be94).

digital information. DANS already experimented with this idea in a pilot project called MIXED in 2007 (Van Horik and Roorda 2011). Once digital information is converted to a well-documented XML file, it does not need much attention to be readable and understandable in, say, 20 years' time.

Some archaeological units have been experimenting with and using the SIKB0102 protocol for quite some time now, and their number is increasing rapidly as a result of provincial repositories requesting the protocol. The SIKB0102 specifications are updated yearly. This entails not only minor alterations in response to (for instance) changes in the municipalities, new commercial units, or improvements to the protocol, but also more fundamental changes to the structure of the document. Branches have been renamed, replaced, added, and removed from the protocol, *e.g.,* when specific information is no longer relevant (like many Archis 2 attributes). The SIKB publishes the protocol on its website, along with the XSD, enumerations and examples (www.sikb.nl/datastandaarden/richtlijnen/sikb0102). Any changes between versions are documented in a separate report.

The SIKB is convinced that the purpose of the data exchange document is limited to a single-use event: at the closing stages of a research project, at the moment a commercial unit sends the XML document to the designed depot in the province (or municipality). The depot imports the data, and the lifespan of the XML document is over. Preferably, both parties should be using the current, in-force version of the protocol. But because software updates can easily lag behind, a few older versions of the protocol are still supported and available on the SIKB website. Currently (as of May 2023) only the previous version (4.3) and the current version (4.4) are available. The data model (UML), XSD, enumerations (*domeintabellen*) and examples can be downloaded. The thesauri (enumerations) are also published on a separate website (codes.sikb.nl), for the current version of the protocol.

In Dutch archaeology, the XML document is accepted at the repositories of the provincial depots, as well as by DANS and Archis, among others. These repositories not only convert the content of the *pakbon* into their own management system, but also store the *pakbon* for the long run. As this XML document holds the majority of a project's produced archaeological data, it is a valuable resource in its own right. The data may not be as detailed and rich as the original data files from the commercial unit, but they hold the information in a highly standardised manner. Suppose that we could bring together all these XML documents produced over the past five years. This would allow for a very useful and smart search engine to cross-search harmonised archaeological projects on a national scale.

Synthesizing information at that scale of detail is impossible in any of the current repositories. Several examples of such a 'big data' approach, such as text mining (Brandsen 2022; Fischer *et al.* 2021) and data mining (Wilcke 2022), have recently produced promising results in Dutch archaeology.

However, one essential condition for using SIKB0102 as an archiving format is missing: the XSD and enumerations should be available for any possible version of the exchange format. Unfortunately, the SIKB currently fails to make these available. We feel that this is a crucial missed opportunity. We lack the UML and especially the XSD of the previous versions. A provisional dataset (figure 9) has recently been created at DANS (DOI:10.17026/dans-zbn-be94); this holds the documentation of as many previous versions of the protocol that could still be gathered from the SIKB website and local copies by the authors.

XML as an archiving format was also recently proposed by Brandsen (2022) in his Ph.D. thesis on text mining archaeological reports. Since the Malta legislation entered into force, these grey literature reports have been produced by commercial units in such large numbers that it has become very difficult to keep up with all new discoveries. A special grant scheme, the 'Oogst van Malta' (Malta harvest), also addresses this problem, enabling archaeologist to dedicate their time to special topics, in order to translate the information from these separate reports into a coherent archaeological narrative. One of the problems Brandsen encountered is the long-term archiving format used by the repositories. The PDF format is very difficult to re-engineer into text. Drawing distinctions between different elements of the publication – such as text body, captions, chapters, paragraphs, table content, and so on – is almost impossible. For text mining, understanding the structure of the document is paramount, as the flint arrowhead described in a chapter called 'Neolithic period' is implicitly dated. The best possible solution would be storing archaeological publications in a generic XML format that incorporates the document's hierarchical structure, thereby making explicit what we as human readers can deduce immediately. From an archiving standpoint, the XML format is also preferred to the PDF/A (Adobe-specific) format. Although this would be a major step forward, PDF will probably remain the de facto standard worldwide for some years. Still, we could consider asking the SIKB to add some of the content (*e.g.*, the summary) of the report to the data exchange format SIKB0102.

When archiving is designated as one of the functions of the SIKB0102 protocol, an additional incentive to use the protocol's spatial components is also needed, as this seems to be used very rarely at present. Enabling even a fuller use with, for instance, the location of trenches, finds and excavated features.

7 Conclusions

It goes almost without saying that Dutch archaeology is in need of a standardised exchange protocol, especially now that almost all archaeological partners collect archaeological documentation digitally under the Malta trajectory. The SIKB0102 protocol already fulfils a key role, although not all archaeologists are comfortable with the chosen format. It is important that all partners are convinced of the benefits and are provided with a series of low-threshold tools. While the *pakbon* was originally designed to improve the transition between the excavator and the legal owner of the finds and documentation, SIKB0102 has the potential to become a versatile, flexible and universal exchange and archiving format for a highly standardised set of archaeological data.

The data exchange protocol SIKB0102 was almost exclusively designed by a group of (heavily) ICT-trained archaeologists. The SIKB seems to have misjudged the gap in interests and skills between the 'average' archaeologist and the 'computer' archaeologist. The deposit process has not become easier for those who have to do this on a daily basis in accordance with the KNA regulations. Many archaeologists and personnel at the depots still consider the *pakbon* a complex and difficult black box and are still reluctant to introduce the protocol at all. The problems that occur in the handover cannot yet be solved easily. Both parties need to have a shared knowledge base and easily available tools in order to discuss problems and find solutions to discrepancies.

We feel that the most important next step is improving the availability of open-source additions to the "SIKB Interoperability Toolkit". Some simple additions would equip archaeologists with tools to really explore the data that they currently experience as being hidden inside the *pakbon*. Within a (geo) graphical user interface, they could view the data, see links between entities, and (most importantly) share a common dashboard via which different parties can solve problems. Only with tools like this can archaeologists experience the benefits of a protocol such as the SIKB0102. We would like to see the SIKB and/or RCE initiate such a development together with the Dutch archaeological community. Creating and maintaining

these open-source tools should be a joint effort from all actors in the Dutch archaeological field, because all of us would benefit. The tools should not only facilitate the deposit process today, but also create valuable standardised descriptions in XML that could serve as a resource of archaeological information for the future.

References

Boasson, W. and R. Visser. 2017. "SIKB0102: Synchronizing Excavation Data for Preservation and Re-Use." *Studies in Digital Heritage* 1, no. 2: 206-224. DOI: 10.14434/sdh.v1i2.23262.

Boogert, S.J. 2006. *Wie wat bewaart, die heeft wat: een onderzoek naar de overdracht van archeologische vondstcomplexen aan provinciale depots*. Den Haag: Erfgoedinspectie / Archeologie.

Brandsen, A. 2022. "*Digging in documents: using text mining to access the hidden knowledge in Dutch archaeological excavation reports*." PhD diss., Leiden University. https://hdl.handle.net/1887/3274287.

Erfgoedinspectie. 2018. *"Verder graven in depots: behoud en toegankelijkheid in gemeentelijke archeologische depots."* Den Haag: Erfgoedinspectie, Ministerie van Onderwijs, Cultuur en Wetenschap.

Fischer, A.D., H. van Londen, A.L. Blonk-van den Bercken, R.M. Visser, and J. Renes. 2021. *"Urban farming and ruralisation in the Netherlands (1250 up to the nineteenth century), unravelling farming practice and the use of (open) space by synthesising archaeological reports using text mining."* Amersfoort: Cultural Heritage Agency of the Netherlands (Nederlandse Archeologische Rapporten 68).

Gilman, P. and M. Newman. 2007. *"Data standards for the Historic Environment, part of Informing the Future of the Past, Guidlines for the Historic Environment Records, Archaeology Data Service."* WiKi, Electronic Document. Accessed June 12, 2021 https://archaeologydataservice.ac.uk/ifp/Wiki.jsp?page=SectionB.7.

Meffert, M. 2018. *"De pakbon voorbij: de impact van digitaal werken in de Archeologie."* Presentation at SIKB jaarcongres 2018, De verbindende kracht van data. Electronic Document. Accessed June 12, 2021. https://www.sikb.nl/doc/congres2018/presentaties/sessie1_1_PakbonVoorbij_MartinMeffert.pdf.

SIKB. 2011. *"Wijzigingsblad KNA 3.2 landbodems. Gouda: Stichting Infrastructuur Kwaliteitsborging Bodembeheer."* Electronic Document. Accessed June 12, 2021. https://www.sikb.nl/doc/KNA32/Wijzigingsblad_KNA_3.2_landbodems_12_december_2011.pdf.

Sueur, C., J.W.H.P. Verhagen, and M. Wansleeben. 2008. *"Advies archeologische uitwisselingsprotocollen SIKB."* Amersfoort: Vestigia BV (Vestigia-rapport V513).

Van Horik, R. and D. Roorda. 2011. "Migration to Intermediate XML for Electronic Data (MIXED): Repository of Durable File Format Conversions." *The International Journal of Digital Curation* 6, no. 2: 245-252. DOI: 10.2218/ijdc.v6i2.200.

Verhagen, J.W.H.P., M. Wansleeben, and C. Sueur. 2011. "Setting a Standard for the Exchange of Archaeological Data in the Netherlands." In *On the Road to Reconstructing the Past: Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 36th International Conference, Budapest, April 2-6, 2008*, edited by E. Jerem, F. Redö, and V. Szeverényi, 152-155. Budapest: Archaeolingua.

Wilcke, W-E. X. 2022. "*Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge*." PhD diss., Vrije Universiteit Amsterdam.