3D ARCHAEOLOGY IN EASY

Generating an object-based archaeological 3D dataset for the digital archive of DANS

Universiteit Leiden
The Netherlands

Julius Pilzecker

3D archaeology in EASY

Generating an object-based archaeological 3D dataset for the digital archive of DANS

Julius Pilzecker s1520202

Master Thesis Archaeological Science - 1084VTSY

Dr. Lambers

University of Leiden, Faculty of Archaeology

Leiden 1 July 2020: Final version

Table of Contents

Acknowledgements

First of all, I want to thank my thesis supervisor Dr. Karsten Lambers for guiding me with this thesis. With his helpful and critical perspective, he aided me to generate the thesis to its current state and greatly encouraged me throughout the writing procedure.

I also want to thank Vasiliki Lagari, Carol van Driel and the National Museum of Antiquities (Rijksmuseum van Oudheden).

I want to express my gratitude to Vasiliki for her insights on 3D modeling, but especially for allowing me to perform my research on the many exquisitely made 3D models she has generated for her research. I want to thank Carol and the National Museum of Antiquities for providing me with the CC0 access license that enables me to use and share the 3D models however I want.

On the personal side I want to give my gratitude to my housemates. They supported me throughout the thesis writing procedure, which unfortunately partly coincided with the COVID-19 pandemic. The pandemic was a strange and difficult period that initially demoralized my motivation greatly. However, my housemates provided daily support and a welcoming study environment that restored much of the missing motivation.

Finally, I want to acknowledge the investments and love my family have given me, not only during the writing of the thesis, but also for the pursuit of my academic goals. Their sincere and committed support during my ups and downs have made me the person I have become today. And that is something I am truly grateful for.

Chapter 1: Introduction

1.1 General introduction

The last decades have brought advancements in 3-dimensional (3D) modeling, offering researchers the possibility to generate large and complex 3D models with relative ease in a relatively short timespan. 3D models are used for multiple purposes, such as visualization, animation, inspection, navigation, or object identification (Remondino and El-Hakim 2006, 269). Within archaeology, 3D modeling has progressively been applied in archaeological excavations where recording techniques such as photogrammetry, laser scanning and structured light scanning have proven valuable tools (Niven *et al.* 2009, 2018-2019). These techniques provide documentation of the archaeological object with great accuracy and with few physical barriers or restrictions (Betts *et al.* 2011, 756; Niven and Richards 2017, 175-176; Tsirliganis *et al.* 2002, 766). Archaeology has a unique connection to collecting and recording data, because in the field, archaeologists generally utilize invasive techniques to obtain their data and destroy the possibility to repeat the process (Kansa *et al*. 2019, 41). Archaeologists work with fragile and delicate objects, such as bones and ceramics. Repetitive touching and moving of these archaeological objects in a laboratory or museum can cause damage and destroy important details on these objects. Generating 3D models of these objects secures a reserve copy of the original and provides digital access for studying the object from remote and distant locations, and if preserved correctly, for the long-term.

Within archaeology, 3D models are frequently made on the artefact or object scale and are consequently termed object-based 3D models. Alternatively, 3D models can be generated at the local or regional scale (Lambers and Remondino 2008, 27-29). However, those scales are not taken into consideration here, because covering all the different scales would be beyond the scope of this thesis. Another way of determining the different scales for 3D models in archaeology are the six entities of scale (Grimaud and Cassen 2019, 4-5). The six entities are: Geographical area (5km), Topography (1 km), surroundings (100m), Tomb structure, internal structure and slab. Although the terminology is slightly different, the entities and scales partially overlap. For example, the object scale and internal structure slab are both targeted as artefacts and the regional scale and the geography both target the surrounding hinterlands. For clarity and consistency throughout the thesis, the object scale is used. Mainly because it is used more in literature, but also because it suits the digital acquisition techniques better.

Before 3D recording techniques were used, two-dimensional (2D) representations were applied on archaeological artefacts as part of the documentation process. 2D representations of objects are dependent on the interpretation of the person writing or making the recording and are an imperfect replacement for representing an object if the original is unavailable. 2D representations create an inaccurate depiction of curves, angles and overall shape because of the lack of an additional dimension (Errikson 2017, 94; Kuzminsky and Gardiner 2012, 2746). 3D models for archaeological objects record the geometry of the object by measuring the X, Y and Z coordinates for each point of the object. Determining how many points a 3D model has, defines one of the key quality aspects of the model (Grimaud and Cassen 2019, 6).

Cultural heritage researchers progressively utilize 3D visualization applications for visualizing archaeological objects. Resulting in a variety of versatile 3D recording applications that develops an increasingly complex set of research data and documentation process of this data (Niven and Richards 2017, 175; Pfarr-Harfst 2016, 33; Revello Lami 2016, 422; Tsirliganis *et al.* 2002, 766). The research data and documentation are essential to preserve for verification purposes and reusability for subsequent research. A core subset of research documentation is metadata, which is data about the generated data. Metadata assists other users in understanding the data by providing information about the model. For example, metadata can address information such as title, size, subject, provenance, access license, general context of the model or context of a project. Retaining this data and metadata has many advantages for structure, access, and use (Niven and Richards 2017, 177). This is especially true for situations where future access to the original archaeological subjects is limited or impossible. Which is an occurrence that happens occasionally in archaeology with fragile materials that are at risk of bad maintenance, environmental damage, erosion or because of ethical reasons (McPherron *et al.* 2009, 19-20; Yannis and Philip 2016, 28).

Preserving digital data for the long-term is not a simple task to accomplish. Questions like 'What does the bitstream mean?' and 'what is the interpretation of this meaning in the future?' rise when considering future understanding of digital data (Horik 2005, 14). Developments in software for 3D models and 3D file formats still progress relatively quickly, suggesting short lifespans, also known as longevity, of current 3D models. The short longevity of 3D file formats means that in a short period of time, 3D models with a specific file format are no longer accessible. Although this sounds very extreme, this accessibility can easily be avoided if 3D models are exported into more advanced and state-of-the-art 3D files. This brings however the potential consequence of data loss. Converting from one 3D file format to another requires many calculations of 3D software. These calculations will be different each time and potentially alter the 3D points of the 3D object. Another option for extending the longevity of 3D formats is to present the 3D data in a human readable bit stream (ASCII), which might require more storage, but provides more support for accessibility for future use. Both these options are very useful to extend the longevity of 3D models, but to preserve these models, they also need to be digitally stored.

Digital archives can maintain digitally produced data to guarantee preservation of data for the long-term, meaning for ten or more years. To preserve this digital data in its most perfect state, they constructed preservation strategies and require a highly specific digital archiving structure. That means that if someone is to deposit data in this digital archive, they must adhere to certain requests.

Within the Netherlands the Electronic Archiving System (EASY) is a certified digital archiving system maintained by Data Archiving and Networked Services (DANS). EASY sustains the E-depot of Dutch archaeology (EDNA) (www.dans.knaw.nl; www.coretrustseal.org). As of this date[1], more than 140,000 datasets have been deposited within EASY, of which 84,000 are archaeology specific datasets. The strategy and structure of EASY enables datasets to be preserved for twenty or more years, if the deposited data adheres to DANS specified technical supporting specifications. In turn DANS ensures reusability and sustained access for deposited datasets. DANS also acknowledges and complies to the FAIR data principles to a certain extend (Tsoupra *et al.* 2018, 19). FAIR principles are a set of guidelines to make data Findable, Accessible, Interoperable and Reusable (Wilkinson *et al.* 2016, 1). The FAIR principles will be further described in chapter 3.1, along with other digital archival standards and specifications of EASY.

1.2 Research problem

EASYs design offers a structured archiving procedure for archaeological data and related considerations such as personal data, file formats and discipline specific requirements. For 3D data however, this structured procedure is more complicated and not standardized yet. 3D data is still relatively new and 3D data have many different purposes. An optimal 3D archiving format or a standardized 3D format does not exist yet because of this (M'Dhari *et al.* 2019, 49-50). A protocol for archiving object-based 3D models and their accompanying documentation has not been formulated yet and this absence of standardization and guidelines is associated with loss of information (Kuroczyński 2016, 150; Pfarr-Harfst 2016, 37; Pletinckx, 2012, 230). Research in object-based archaeological 3D models is oriented towards new technological improvements or specialized research questions, but research in maintenance and longevity of the models and its related documentation is generally avoided (Pfarr-Harfst 2016, 38). The creation of a (universal) standardization of 3D documentation will be an indispensable feature to ensure scholarly quality of 3D modeling in the future and mitigates the loss of knowledge. Not only will it provide stability and cohesion for the digital archive and its documents, standardization provides continuity and clarity for the users as well.

1.3 Aims and objectives

The main target of this thesis is providing a constructive and clear procedure of what an archaeological researcher has to do with object-based archaeological 3D models and their

[1] The numbers are derived from www.easy.dans.knaw.nl in June 2020.

accompanying data(set) to meet the requirements of a digital archive, the requirements of the users of the data and lastly, legal and institutional requirements. Users want to preserve the data as accurate and unaltered as possible and want to be time efficient and do not want to spend too much effort in gathering data. Users also want to be able to preview 3D models before downloading large 3D models. Digital archives want their depositors to engage in efforts to extend the longevity of their dataset and provide documentation for transparent and easy use. The digital archive must consider what to expect from users within a reasonable timeframe. Digital archives attempt to obtain the highest quality of data and want extensive and thorough documentation for each dataset but construct highly complicated structures that most users do not have time for to fully understand. On top of that, they must abide to regulations by their institution and should consider legal and accessibility regulations. It requires action from both the users and the digital archives to make concessions for reciprocate means, both will improve if a sound and clarified framework is given. This concept must fit in what is contemporary and what is technically possible in archaeology. Free and Open-Source Software (FOSS) are used during the workflow and results in this thesis for the purpose of reliability, transparency, accessibility, and finally, because the applications are obtainable without charge. Affordability is an element that is heavily preferred within archaeology and capital and wealth are not something archaeologists are well known for (Diara 2019, 6). FOSS have become a research environment in archaeology, because the transparency greatly improves the collaborative cycle of research (Ducke 2012, 577).

The focus of this thesis is on 3D models of pottery, a category of objects that is fragile and vulnerable to damage, but very frequently studied in the field of archaeology. This topic is also chosen to demarcate the process within this thesis. Although this specific case study is used, the outcome of this thesis has the potential to also be applicable to other materials and similar archaeological cases as well. In this thesis I will not indulge in generating object-based 3D models, but rather on how to alter and prepare existing 3D models to be deposit-ready for digital archives from a pragmatic perspective. This viewpoint is chosen because many digital archaeologists are not trained and unfamiliar with data preparation for preservation, let alone preservation of 3D data (Dell'Unto 2018, 54-56). The perspective considers all the stakeholders within the process of preserving 3D data and is therefore primarily relevant for researchers, students, and data managers in the academic field of archaeology and data preservation. The main research question of this thesis is formulated as follows:

"Which requirements are essential for digitally preserving a dataset of object-based archaeological 3D models for the long-term in digital archives?"

To answer the research question, the following sub-questions are formulated:

"Do the considerations EASY expect from users help 3D datasets in becoming FAIR?"

"Which of the DANS preferred 3D file formats fits the purposes of object-based archaeological 3D models best?"

"Which tools are useful for preparing 3D datasets and what benefits do FOSS provide in this preparation process?

The sub-research questions are relevant to address for providing the framework and scope of the main research question. Limiting the scope to only object-based archaeological 3D models and its preservation requirements, allows the possibility of an extensive and detailed workflow that excludes interference of different preservation, 3D modeling or archaeological complications. The research is performed according to the following structure.

1.4 Research Structure

After this introduction is finished, the material of the case study is addressed (see Chapter 2). The topic of this case study are 3D generated models from ceramic material from Syria that was acquired by Leiden University between 1972 and 1982 during an emergency excavation. The material is derived from the site on top of the mountain (*Jebel) Aruda*. The site is named after the mountain. The collection is owned by and stored at the faculty of archaeology of Leiden University and the 3D models are generated by Vasiliki Lagari, a MSc student at that faculty. The last part of chapter 2 shortly discusses why 3D modeling and archiving of this data is particularly relevant. Chapter 3 addresses the principles and standards of digital archives and has significant value for understanding how digital archives are designed and why EASY has strict guidelines for datasets. This chapter can be perceived as understanding the digital archive perspective.

The next chapter (chapter 4) is focused on 3D acquisition techniques, features of 3D models at the object scale and 3D file formats. The purposes that these object scale 3D models compose for archaeologists is also addressed. The purpose indicates what archaeologists require of models; thus, this chapter recognizes which aspects are important from the user perspective. The other elements of this chapter provide a more technical perspective of what 3D formats can offer to the user.

Chapter 5 starts with the description of the used software and hardware, followed by a constructed workflow of preparing preservation qualified datasets of object-based 3D models. After the conceptualization of the workflow, it is practically implemented on the archaeological 3D models of the case study of chapter 2. The 3D models of the case study are converted to all the recommended 3D formats in EASY. Ethical considerations, data structuring, file naming and documentation are also included in the workflow.

The results of this implementation are presented in chapter 6. This chapter critically evaluates the different 3D formats from the user and digital archive perspective and considers the current technological framework as a limiting factor. The findings of these results have methodological and practical implications. In chapter 7, the discussion commences of the challenges and complications encountered during the making of the workflow and results. In the discussion, the advantages and disadvantages of EASY are addressed and suggestions of improvements are presented. Strengths and weaknesses of adhering to digital archives for 3D models, using FOSS in the workflow and (im)perfections of 3D formats are also discussed.

Finally, chapter 8 presents the conclusions of this thesis and answers the research and sub-question stated in chapter 1.3. Chapter 8 also offers suggestions for future prospects of preserving object-based archaeological 3D models.

Chapter 2: Case Study: Ceramics of Jebel Aruda

This chapter presents the case study. The chapter starts with the archaeological background and begins with an overview of the Uruk period, followed by a specification on Uruk pottery and the Uruk site of Jebel Aruda. Next, the focus is on the pottery of Jebel Aruda, the 3D data acquisition of this pottery and finishes with the motivation for selecting this case study for this research topic.

2.1 Archaeological background

2.1.1 The Uruk period

The Uruk period is a cultural period in Early Bronze Age Mesopotamia. The Uruk period ranges from 4000 to 3100 BC and is named after the site where the distinctive plain pottery was first recognized (Crawford 1991, 27). Mesopotamia covers the area between and around the Tigris and Euphrates river system and roughly overlaps with Iraq, Kuwait and the western parts of Syria and Turkey. The Uruk period was preceded by the Ubaid period, dated from 5800 to 4200 BC, and was succeeded by the Jemdat Nasr period, which started in 3100 and ended about 2800 BC (Crawford 2015, 18-39).

The Uruk period is characterized by a rapid expansion in settlements and the appearance of settlements large enough to be determined as cities (Crawford 1991, 27; Lawrence 2012, 24-25). Uruk settlements spread along the Euphrates and the Tigris and moved towards the east throughout the Uruk period. This expansion to the east is regularly dubbed "the Uruk Expansion" and resulted in many of the sites visible in figure 2.1 (Akkermans and Schwartz 2003, 181). This expansion led to more complex administrative systems and a more stratified society, as well as long distance trade and the emergence of warfare (Lawrence 2012, 25). Many important technological innovations were conceived in the Uruk period, such as sophisticated casting processes, the use of the fast wheel in pottery, the use of cylinder seals and the first pictographic writing on clay tablets (Akkermans and Schwartz 2003, 183; Crawford 1991, 28; Lawrence 2012, 24).

The motivation for all these developments in Mesopotamia is debatable, but irrigation, the backbone of the region's agriculture, is pointed as one of the main reasons (Crawford 1991, 19). Irrigation created large grain surpluses that were used to provide food to specialists, like craftsmen, bureaucrats, and rulers. The surpluses disregarded the need for food production by these specialists themselves. Besides irrigation, the large expanse of the alluvial plains of southern Mesopotamia arguably provided the right environmental conditions to support populations for an urban and complex society (Akkermans and Schwartz 2003, 184).

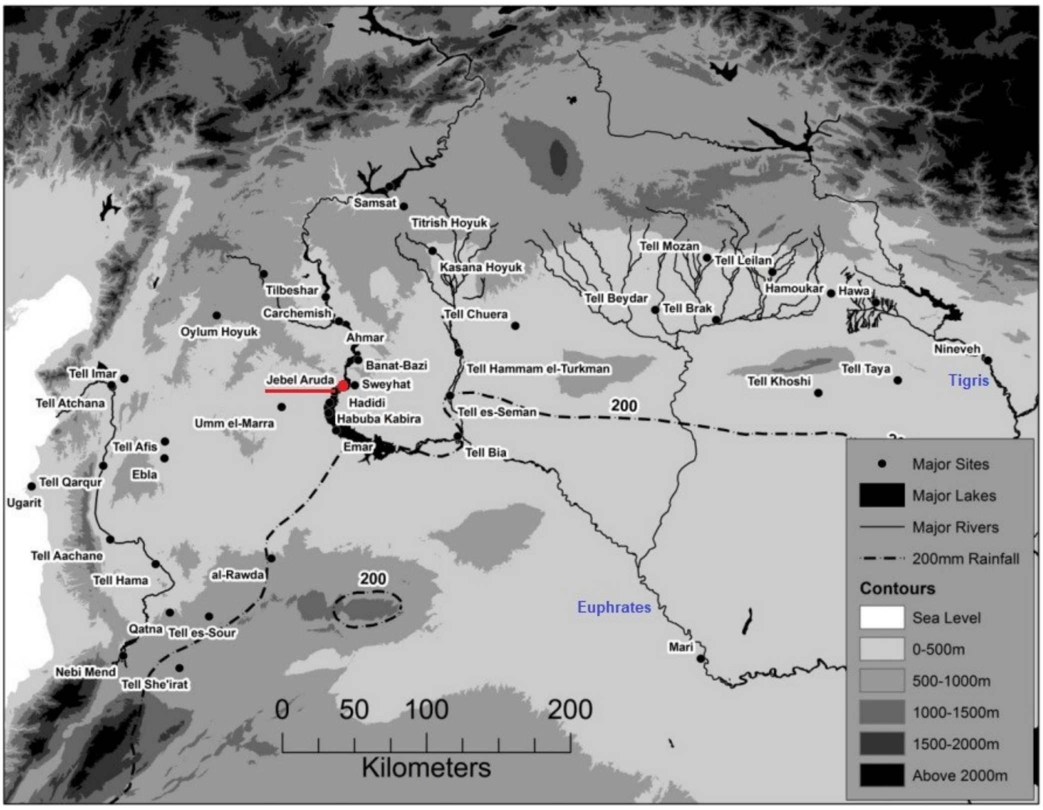


FIGURE 2.1: MAP OF SITES IN THE NORTHERN FERTILE CRESCENT BETWEEN 4000 AND 3000 BC. JEBEL ARUDA IS UNDERLINED IN RED (LAWRENCE 2012, 26).

The Northern part of Mesopotamia seems to have been less urbanized than Southern Mesopotamia, although this might be caused by the irregular archaeological survey evidence (Crawford 1991, 116). However, both the northern and southern part of Mesopotamia started in the fourth millennium BC with a huge increase in settlement growth (Algaze 2008, 117). Differences in settlement development occurred from approximately 3500 BC, when the growth rate halted in North Mesopotamia.

The architecture of Uruk was lavishly decorated and immense compared to the previous Ubaid period (Akkermans and Schwartz 2003, 183). Large public buildings required huge amounts of labor and specialized craftsmen and were constructed with a tripartite plan and clay cone or multi-colored mosaics (Akkermans and Schwartz 2003, 191). However, it is unclear if these public buildings served as elite residences or for secular purposes.

2.1.2 Uruk Pottery

Unlike the lavishly decorated public buildings in the Uruk period, the Uruk pottery can be defined as relentlessly plain and undecorated (Akkermans and Schwartz 2003, 184). Especially when compared to the richly decorated pottery assemblages of the Ubaid period (Potts 2009, 4). This contrast is probably caused by of the transition to ceramic mass production in the uruk period (Sanjurjo and Fenollos 2012, 265). The first and most prominent Uruk pottery are beveled rim bowls (henceforth called BRB), figure 2.2 and 2.3.

FIGURE 2.2: AN EXAMPLE OF A BEVELLED RIM BOWL (HTTPS://RMO.NL).

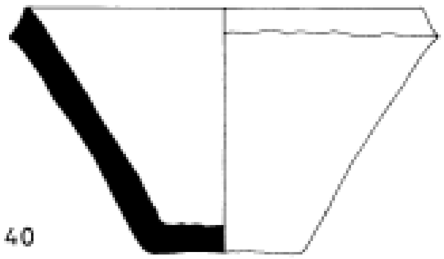


FIGURE 2.3: THE SHAPE OF A BEVELED RIM BOWL (OATES 1985, 185).

The production started with clay being roughly pushed into various sized molds. The surplus clay was removed around the mouth of the bowl by cutting it off, creating a beveled rim. No pottery wheel was used in the making of these bowls and no molds have been recovered, suggesting that the mold was made of wood or other perishable materials. BRBs have a porous texture and are sometimes described as badly fired pottery, because the clay is only lightly fired (Millard 1988, 51). They are normally 10 cm in height and 18 cm in diameter. However, their size is not always consistent and the carrying capacity can vary between 0.4 and 0.95 liter (Beale 1978, 290). This bowl type was generally found in stacked and large quantities in small and large Uruk settlements and were common throughout the whole Uruk period, as visible in figure 2.4. They were also produced locally and made with local clay (Crawford 2015, 32). Although BRBs were widespread, the function of these pots is still unclear. Multiple hypotheses are available, ranging from the measuring of rations of grain for workers and offering containers to bread molds and utilization of salt commerce (Buccellati 1990, 24; Crawford 1991, 180; Potts 2009, 4; Sanjurjo and Fenollos 2012, 265). A multi-purpose functionality is the most likely, considering the various locations of the bowls in excavations.

Another, less researched pottery type that can be connected to the Uruk period is the flowerpot (henceforth called FP), displayed in figure 2.5 and 2.6. FPs are considered crude chaff-tempered bowls and are relatively like BRBs in ware but differ in shape (Rothman 2002, 55). The sides flare out and have a string-cut base. They vary in their production style as they are wheel-made and are generally around 16 cm in height and 16 cm in diameter (Akkermans and Schwartz 2003, 193; Oates 1985, 183). The decline of BRBs in 3200-3100 overlaps with the increase in flowerpots, see figure 2.4. However, flowerpots are relatively unknown because both ‘flowerpot’ and ‘conical cup/bowl’ have been used to describe these vessels (Fielden 1981, 158). The use of flowerpots is like the use of BRBs debatable, hypotheses vary from containers for a baby funeral to a mixing bowl for bitumen (asphalt). Future research between these pots can be useful to determine if these flowerpots were also used for making bread after the decline of BRBs (Goulder 2010, 359).

BC (cal)	Period	S Mesopotamia	N Mesopotamia	Iran
(4400) 4200–3900	**Late Ubaid**	Coarse flint-scraped bowls	Wide flowerpots (slow wheel), coarse flint-scraped bowls	Coarse flint-scraped bowls
3900–3600	**Early Uruk**	**BRBs appear *c.* 3900**	Wide flowerpots cease; Coba bowls, other non-bevelled coarse bowls e.g. Arslantepe; **a few BRBs – Brak, Nineveh, Amuq**	Coarse flint-scraped bowls
3600–3400	**Mid Uruk**	**BRBs common** Conical cups	**BRBs common by 3600** Conical cups #	**BRBs common at some sites by 3600** Conical cups #
3400–3200	**Late Uruk**			
3200–3100	**Jemdet Nasr**	**BRBs decline** Tall flower-pots increase; conical cups (including solid-footed goblets) increase	**BRBs continue** Tall flower-pots, conical cups increase	**BRBs decline** Tall flower-pots increase; conical cups (including solid-footed goblets) increase
3100–2900	**ED1**	Solid-footed goblets	**BRBs cease by 3000 except a few in Syro-Anatolia**	Solid-footed goblets

Contemporary with BRBs at e.g. Godin Tepe, Şarafabad, Hacınebi

FIGURE 2.4: APPROXIMATE TIMELINE OF THE SPREAD OF BEVELED RIM BOWL (GOULDER 2010, 352).

FIGURE 2.5: AN EXAMPLE OF A FLOWERPOT (HTTPS://WWW.RMO.NL).

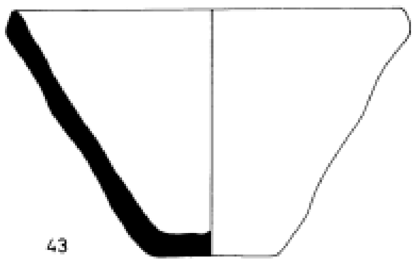


FIGURE 2.6: THE SHAPE OF A FLOWERPOT (OATES 1985, 185).

2.1.3 Jebel Aruda

Jebel Aruda was an Uruk colonial enclave overlooking the Euphrates, located in modern day Syria, as depicted in figure 2.7. The site is named after the 60 meters high mountain on which it is located (Akkermans and Schwartz 2003, 194). The Uruk enclave was the first and only occupation on this mountain and covered between three to four hectares, which is relatively small compared to contemporaneous Uruk sites (Algaze 2008, 70). The start of the settlement is traced back to the Late Uruk period, 3400-3200 BC, and was abandoned a century after its occupation (Bakker *et al*. 1999, 782). The cause of this abandonment can be related to a violent and thorough conflagration (Driel 2002, 191-192). During its occupation, Jebel Aruda may have represented an associated administrative quarter for the nearby settlement Habub Kabira-Süd, but the site might also have had religious functions (Akkermans and Schwartz 2003, 196).

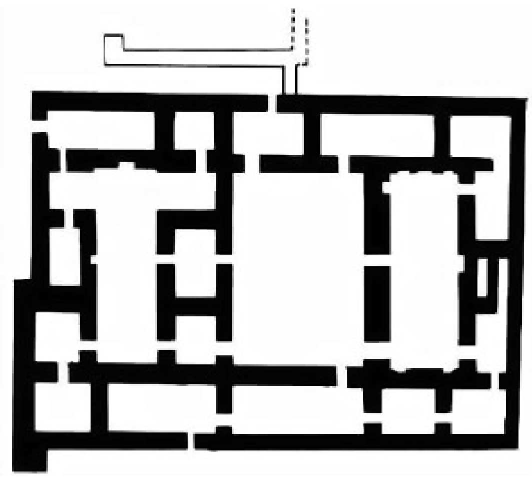


FIGURE 2.7: EXCAVATED AREAS OF JEBEL ARUDA (ALGAZE 2008, 71)

FIGURE 2.8: AN EXAMPLE OF A HOUSE PLAN FROM JEBEL ARUDA (CRAWFORD 1991, 138).

The terrain of the site is very resilient, and it seemed that the inhabitants were not able to remove the rocks and stone or did not care enough to remove them (Driel 1977, 43). This is visible in the orientation of structures, as they follow the orientation of the terrain instead of a structured pattern, as visible in figure 2.7. The site is dominated by two tripartite temples with niched and supported facades. Surrounding these temples are well-constructed, typical Uruk style tripartite houses for the important residents and visitors of the site. Other houses on the site are rectangular houses built around an open court or large room, as depicted in figure 2.7 and 2.8 (Crawford 1991, 138). Jebel Aruda is located near many other Uruk hubs in the river valley of the Euphrates, such as Habuba Kabira North, Habuba Kabira South, Mureybet, Sheikh Hassan, Hadidi, and Tell el-Hajj (Akkermans and Schwartz 2003, 196). The excavation of the site of Jebel Aruda by Leiden University was performed between 1974 and 1982 (Driel 2002, 191). The site was excavated as an emergency excavation, because in the 1960s the Syrian government wanted to build a dam in the Euphrates to provide energy for northern Syria. The archaeologists of Leiden University started the excavation with trial tranches and later expanded to full excavations of buildings (Driel 1977, 45).

2.2 Pottery background

The site of Jebel Aruda is established and two different Uruk pottery have been addressed. The next part narrows the topic down to the Uruk pottery found in Jebel Aruda and how they were digitized in 3D.

2.2.1 Pottery of Jebel Aruda

The pottery assemblage of *in situ* household inventories on Jebel Aruda are displayed in appendix A. They demonstrate the differentiation of the basic Late Uruk Pottery spectrum (Driel 2002, 194). There is notable evidence for individual preference of pottery, with specific pottery types being limited to one house or to a particular room. The most eccentric vessels of the collection, and an indication of pottery preferences, are three specially formed hedgehog vessels found in one of the southern houses, visible in appendix A (Driel 2002, 195).

The distribution of pottery, also visible in appendix A, indicates that pottery in the northern part of the site is generally absent, which is supported by distributions of the torpedo-shaped vessels and 'rolly-bins'. BRBs and flowerpots earlier described in chapter 2.1.2 are both found in Jebel Aruda as well. At the end of the Jebel Aruda project, all the complete and unbroken finds, including ceramics, were brought to the National Museum of Aleppo. The excavators were allowed and able to take the broken artefacts and sherds to the Netherlands, where the sherds have been glued and taped back together (Van Driel 1977, 46). Most of the glued and taped artefacts are preserved in National Museum of Antiquities (Rijksmuseum van Oudheden), although a selection is stored in the depot of the Faculty of Archaeology of Leiden University.

2.2.2 3D scanning of the pottery

Nine BRBs and twenty-four FPs are currently preserved in total within the depot of the Faculty of Archaeology of Leiden University. Vasiliki Lagari, a fellow Digital Archaeology student, has generated 3D models of each of these 33 pots using photogrammetry. Photogrammetry generates

3D models by overlapping multiple images of an object from different locations and angles through measurement and interpretation methods, further addressed in chapter 4.2.2 (Luhmann *et al.* 2014, 2; Robson *et al.* 2012, 92). Photogrammetry offers a high level of accuracy and is timewise a quick method. Table 2.1 displays the technical specifications of the 3D data acquisition. For her research, Lagari uses these 3D models in her research to understand how these models contribute to the field and if they help advance the archaeological study of ceramics. Lagari also addresses the debate of the function of BRBs and flowerpots.

Of the 33 3D models, 30 are used for this thesis. The three models that are not used, were incorrectly transferred or did not incorporate the correct additional files of MTL or JPG (see *file format used* in table 2.1 or chapter 4.4.6 for more information on additional files of OBJ). Not all the pottery of the dataset contains the same number of faces (what a face is, is addressed in chapter 4.3.1). The standard size is 2,500,000 faces, but some of the models have a face count of 2,000,000 or 1,000,000 faces, which were created by accident. These lesser face count models are also smaller in file size and therefore require less storage space. The procedure for these smaller sized 3D models is, apart from a few naming modifications, not altered in the workflow. These files are still incorporated in the thesis, because they are valuable in the results chapter for time analysis and storage space comparisons.

Scanning method:	Photogrammetry
Scanning software:	Agisoft Metashape Professional 1.5.5 (64 bit)
Dense Point Cloud and Mesh generated:	High Quality
Reprojection error	<0.5
Number of cameras:	The number of cameras varied from 120 to 160 (parameters: size, height). Photo textures were generated in Agisoft Metashape (4096 x 1).
Model units:	1 unit = 1 m
Illumination source	The objects were lit using 6300 K led lamps (cool white).
File formats used	the 3D meshes were exported to an ASCII OBJ file. In addition to the OBJ files are necessary MTL files and the textures. The textures are stored in JPG.
Comments:	The objects were generated with 2,500,000 faces and then decimated to 200,000 faces to

	be easily manageable in Blender software for the purposes of the research of Lagari).
Individual processing procedure	Alignment was done by the automatic align feature of Metashape. Different models were created for the bottom and top sides of each vessel. Next, masks were created from each model. Lastly, the masks were combined to one chunk to create the complete final model. The mesh was exported from Agisoft Metashape as an ASCII OBJ file.

TABLE 2.1: TECHNICAL SPECIFICATIONS OF THE 3D DATA ACQUISITION OF THE FPS AND BRBS OF THE CASE STUDY.

2.3 Relevance of 3D modeling and digitally archiving

There is a plurality of reasons to preserve these specific 3D models. First, preserving models online helps the models in being more accessible for archaeological researchers specialized in Uruk and Mesopotamian pottery. As became evident earlier in in this chapter, more research on Uruk pottery is still required to generate new insights on the debate of functionality of both BRBs and flowerpots. Secondly, if the 3D models are preserved correctly, other end-users, whether student, researcher or member of the general public can access the models with ease and without having to request access to visit the deposit of the National Museum of Antiquities or the Faculty of Archaeology of Leiden University. The high number of faces of the models (2,500,000, as described in table 2.1) provide a visual quality that is to a certain extend comparable to the actual object. 3D models also provide a much better overall quality compared to images, are much more adaptable when used in 3D modeling applications and allow extremely accurate measurements.

My motivation for choosing these 3D models as the case study is that these models provide a perfect example of high-quality archaeological 3D data. The high quality of the models is not only based on the faces, but also on the correct use of illumination sources during generation and the extensive processing description provided by Lagari. The data acquisition has been performed with a very recent version of Agisoft Metashape (v1.5.5). The effort and time that has been put into the generation of the 3D data, should not be underestimated. Although hardware is getting better in generating high quality 3D models, creating a model like these can, depending on the quality, still take up hours (Olsen *et al.* 2013, 252). Making 3D models also requires budget for equipment as well as costs for licenses and staff, thus reusing these models can be financially helpful (Berchum and Grootveld 2016, 77).

In summary, this chapter has introduced the case study of two pottery types found in the Uruk site Jebel Aruda. The pottery types are beveled rim bowls (BRB) and flowerpots (FP). The Uruk period is considered one of the important and innovative periods in the history of Mesopotamia. Archaeological research is still performed on both this cultural period and on these bowls and pots. The function of the beveled rim bowls and flowerpots are still debatable, thus different approaches can provide new insights. A total of 33 pots and bowls of these types are stored in the Faculty of Archaeology at Leiden University and are generated in 3D models using photogrammetry by Vasiliki Lagari. Only 30 of the pots and bowls will be used for this thesis. Reasons for digitally preserving these particular 3D models are to generate new insights for in the debate of functionality of the pottery and to allow ease of accessibility to other students and researchers around the world. The high quality ensures that useful and highly accurate measurement methods can be performed on the 3D models.

Now both the material and the relevance for digitally preserving these 3D models are explained, it is time to explain how digital preservation works and how archaeological 3D data has to be altered or described to generate data that is digitally accessible and preserved for the long-term.

Chapter 3: Standards and Principles of digital archiving

This chapter first addresses digital preservation and how digital archives are composed. This composition is divided in the description of preservation techniques and archiving infrastructure. The online archiving system EASY will be addressed in specific and digital preservation policies of this digital archive are presented. In the second part, documentation of datasets that are stored within these digital archives is introduced and will consist of the different levels of metadata. The third part introduces the FAIR principles and how these principles are relevant for digital archiving, 3D modeling and archaeology.

3.1 Digital preservation

Nowadays, much of the scientific production has become digital and is produced digitally (El Idrissi 2019, 1). The digitalization of information has led to an increase of data accessibility, but also cause new challenges. Data is accessible now, but will it be accessible in 10 or 20 years? And will this data still contain its original information and value? Is loss of data preventable?

Digital preservation aims to conserve the digital data and should guarantee that data remains accessible, is stored safely and is understandable in the future. For scientific purposes, (digital) data should be preserved indefinitely to allow other researchers to perform further experiments and studies on the data. Digital archives are digital locations where data can be stored for long periods of time. Their fundamental aim is to ensure that digital data deposited in these digital archives are safeguarded (https://guides.archaeologydataservice.ac.uk). Digital archives thrive because of two elements: correct data preservation and dataset documentation. Dataset documentation involves information about how the data is collected, which standards are used and how they are managed.

Digital archives have constructed preservation strategies to ensure correct preservation and long-term access. Storing data digitally is very difficult and requires many considerations from the hardware, software, and file formats. Therefore, the preservation strategies of digital archives will be discussed first. Digital preservation strategies have to be implemented in a digital archive and require a digital archiving architecture to function properly. The Open Archival Information System (OAIS) is a much-used reference model for digital archive structures and will be addressed after the preservation strategies. After that, the digital archive 'Electronic Archiving SYstem' (EASY) maintained by Data Archiving and Networked Services (DANS) will be addressed. EASY is, as mentioned in the introduction, a digital archiving system that sustains the E-depot of Dutch archaeology (EDNA) and requests specific actions of data depositors because of its strategy and structure.

3.1.1 Preservation strategies in digital archives

Preservation strategies are properly deliberated methods of documentation for preservation of digital content (Shimray and Ramaiah 2018, 47). They address long-term archiving, data retainment and data file formats. The purposes of these strategies are to look for long-term solutions for preserving documents and to be able to view documents with the same frame of reference as the writer, translator, or viewer of the original document (Rothenberg 1999, 3-6). It also deals with the ability to handle contemporary and future datasets in a uniform way. Archiving strategies can use a combination of dependencies, including hardware, software, or file formats. However, they can also rely on the active adoption, type and complexity of the digital information itself (Lee *et al.* 2002, 103).

Appendix B displays six preservation strategies and includes short descriptions of the functionality of each strategy and (dis)advantages of each strategy. The strategies are:

- Migration
- Technology preservation
- Emulation
- Encapsulation
- Standardization
- Obsolescence-prevention

Selecting preservation strategies must be done carefully by examining their advantages and disadvantages, the appropriateness, their cost effectiveness, and metadata creation (Shimray and Ramaiah 2018, 47). So far, no preservation strategy has got a clear edge for an overall preservation strategy. A specific strategy might be appropriate for preserving one file format but can be irrelevant for other file formats. Therefore, a combination of strategies should be considered to eliminate some of the disadvantages (El Idrissi 2019, 5; Lee *et al.* 2002, 103). For example, Encapsulation thrives on its independence of computer platforms, but carries risks if the encapsulated data is stored in incomplete file formats. Standardization can prevent the incomplete file format if an openly available, stable, and universally accepted file format is used. Combining strategies can therefore remove the disadvantages of one of the preservation strategies. However, this combination does not imply that disadvantages are not present anymore, because standardization still requires many investments of institutions, commercial companies, and other stakeholders to maintain the standardized formats (El Idrissi 2019, 5).

3.1.2 Infrastructure of digital archives

The Open Archival Information System (OAIS) model was mentioned as a reference model for digital archives. The OAIS is a reference model for long-term information preservation and making this information accessible for a designated community (CCSDS 2012, 1-1). The groundworks of the OAIS are published by the Consultative Committee for Space Data systems (CCSDS) of NASA in 2002 (https://guides.archaeologydataservice.ac.uk). The model was originally designed for space systems, but its genericity made it also useful for several digital archiving systems (El Idrissi 2019, 5). It is deemed a reference model, because the model operates in a high level of abstraction and is therefore considered as a conceptual framework (CCSDS 2012, 1-12; Lee 2010, 4024). Further specific developments such as the implementation of a chain of discipline specific standards are still required and need to be applied for the OAIS to be functionally applicable in digital archives (www.guides.archaeologydataservice.ac.uk).

The activities researchers or data depositors can perform before depositing data in accordance with the OAIS model are called *pre-ingest* actions, meaning actions before data ingestion into the digital archive. Pre-ingest actions are the main focus of this thesis. The OAIS indicates specific requests of documentation and file formats that should be considered before ingesting, when preserving data for the long term. These requests exist because of the influence of certain functional entities of the OAIS. Figure 3.1 displays the main functional entities of the OAIS. The entities are explained in order of the procedure within the reference model.

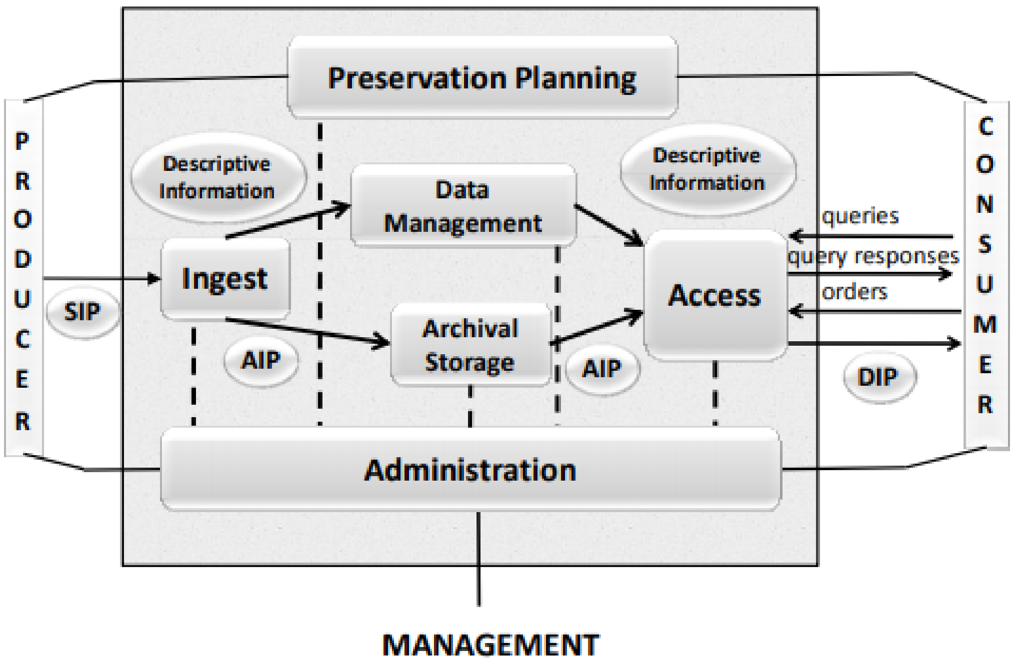


FIGURE 3.1: THE FUNCTIONAL ENTITIES OF THE OAIS MODEL (CCSDS 2012, 4-1).

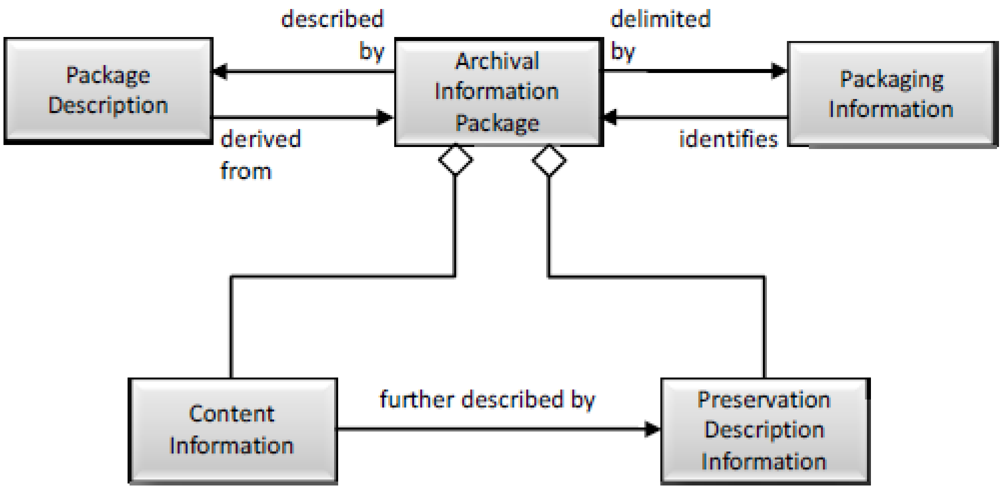


FIGURE 3.2: THE CONCEPT FRAMEWORK OF AN ARCHIVAL INFORMATION PACKAGE (AIP) (CCSDS 2012, 4-37).

The data depositor deposits their dataset in a digital archive where it is called a *Submission Information Package* (SIP), which is a deposit of digital data with the addition of metadata documentation. This addition is deemed necessary for reuse and long-term preservation purposes. The SIP is a basis for the development of *Archival Information Packages* (AIP) and *Dissemination Information Packages* (DIP), both displayed in figure 3.1. These packages are the encapsulated forms of the original documents. The SIP is entered or *ingested* in the digital archive and where necessary, different versions are created of the SIP. One version of the SIP is created for preservation, the AIP, and another one for dissemination, the DIP. An example of how the model works:
An ingested Microsoft Word document will be converted to an Extensible Markup Language (XML) based format, such as the TXT format as an AIP, for long term preservation, and to PDF as a DIP, for dissemination.

The SIP data must be preserved for the creation of the AIP in a suitable preservation format or needs clear migration to a suitable preservation format. The OAIS preserves SIPs and secures the quality through the ingest and coordinates updates to archival storage and data management (Lee 2010, 4025). Archival storage guarantees permanent storage and periodic refreshing of the media, as well as regular error checking (CCSDS 2012, 4-2).

An AIP consists of two types of information, the Content Information (CI) and the Preservation Description Information (PDI), as is visible in figure 3.2. The Content Information depicts the set of information of which the main objective is the preservation. Meaning that it contains the sequences of bits, as well as the representational information for making the interpretation of the data meaningful. In datasets, this representational information of data is the file format and is thus incredibly important to appoint correctly (Lee 2010, 4027).

The PDI ascribes information deemed necessary for acceptable preservation of the Content Information. For 3D data, The PDI are in a couple of 3D file formats described and are located at the beginning of the document in a header. The PDI characterizes information such as the provenance, context, reference and fixity of the Content Information (CCSDS 2012, 2-6; El Idrissi 2019, 5; Lee *et al.* 2002, 98; Waugh *et al.* 2000, 180).

The outline of the OAIS model for preserving object-based 3D models is that adequate and correct documentation and accurate usage of 3D file format are essential elements to provide for long term preservation. Documentation inadequacy is the largest obstacle for reuse of data in the future (www.guides.archaeologydataservice.ac.uk).
Preferred formats for digital archives are formats that are suitable for preservation and dissemination, which would result in storing a dataset only once throughout the OAIS model (CCSDS 2012, 4-29). However, these preferred formats require a simple human readable format, meaning an ASCII or XML format. These formats require larger file sizes, because the original binary string must be converted to this human readable format. Consumers prefer binary files format, mainly because the file size is significantly smaller, which is useful for storage and transfer (www.guides.archaeologydataservice.ac.uk).

3.1.3 Archiving policy of EASY

The Electronic Archiving SYstem (EASY) maintained by Data Archiving and Networked Services (DANS) is a Dutch digital academic repository that assumes responsibility for long-term preservation of research data and accessibility of digital objects (https://dans.knaw.nl/). DANS offers three services, DataverseNL, NARCIS and EASY. NARCIS is the Dutch research portal for scientific information and research data and DataverseNL functions as a repository for data during and after research for the short term.

EASY is the core service of DANS that provides reuse and long-term archiving of research data. The minimum retention period of raw research data is ten years for data to be considered retained for the long-term (https://dans.knaw.nl). However, the earliest data EASY retains is from 1964 and DANS indicates that data in general will still be accessible in EASY after the minimum retention period is over. The data stored within EASY is very heterogeneous of data types, file formats, sizes, and usage. The purposes of the generated data and the processes of generation are also diverse. The variety is large, because of the broad international community DANS designates its services to. For archaeological data, EASY functions as the E-depot for Dutch Archaeology (EDNA). The EDNA is targeted at sustainable archiving and accessibility of archaeological research data. This archaeological research data within EASY should be considered in its broadest sense, both commercial and scholarly research data are appropriate and present within EASY.

To accommodate the researchers and depositors of EASY, DANS has set up some precautions. DANS has evaluated file formats of many different data types that have a high chance of remaining usable in the far future, resulting in a list of preferred and acceptable formats for each data type. DANS evaluates on a regular basis if these preferred and acceptable formats are still relevant or if they are danger of becoming obsolete. Based on these evaluations, the format list changes over time. When file formats are no longer accepted or preferred, they are migrated to a successor format. These actions are also performed if the integrity or security of the dataset are compromised.
The implementation strategy of EASY is structured around the central functional concepts of the OAIS reference model (https://dans.knaw.nl). The ingested data is retained in its original version in a directory that closely resembles the SIP. When ingestion is correctly performed and approved, the data will be published and stored as an AIP. Which is added to the permanent storage facility of DANS. This facility is monitored and refreshes and migrates media when necessary. If data conversion of file formats is required for preservation or access purposes, the data will be converted, but the original file will be maintained as well.

3.2 Dataset documentation

As mentioned earlier in this chapter, digital archives thrive because of two elements: correct data preservation and dataset documentation preservation. A widely accepted approach of the documentation of results is using metadata, which is added to the dataset by the researcher or depositor before ingest (Münster *et al.* 2016, 17). Metadata is structured information that explains, describes, locates, and helps retrievability of information resources (Doyle *et al.* 2009, 165). There are many different types of metadata that heavily depend on the discipline and data acquisition technique, there is no 'one size fit all' metadata schema. Metadata is considered an essential part of long-term digital preservation. A digital archive cannot be perceived as functionable without the implementation of correct metadata (DANS 2011, 6).

For future use, metadata can address many descriptive details: how the digital data in a dataset are comprised, when, by whom, if it has been modified and if the content is trustworthy. It can contain technical details about the acquisition technique and what the required software for rendering are. Metadata can also function as an administrative resource that functions as an overview of all the data within a dataset. The amount of metadata added depends on the researcher or depositor. A specific form of metadata is *paradata*. Paradata is contextual information referring to archaeological (3D) data creation and analysis context (Kansa *et al.* 2019, 45). For 3D modeling it involves information about the collection and modeling process of 3D data and can function as a quality control audit for the 3D data (Corns *et al*. 2015, 38-39).

From a digital archive point of view, richer metadata indicates a better dataset. However, from a practical point of view, much information about a dataset can be presented, but not everything will be essential and relevant for future users. Gathering metadata can also be a task that requires a lot of time and effort. For that reason, the next part mainly addresses the minimal requirements for metadata standards indicated in EASY guidelines. A few small additions are addressed as well, which although not required, can increase the value of the information while doing very little additional effort.

Documentation is distinctive on multiple levels when considering preservation of object-based archaeological 3D models. For the creation of a dataset in EASY, there is project level metadata, file level metadata and a codebook. Each level of metadata is specified according to DANS guidelines.

3.2.1 Project level metadata

Metadata in EASY is implemented by the user (i.e. researcher, project data manager, depositor etc.) during deposition of the dataset and contains general information about the research project. It is based on the Dublin Core (DC) metadata standard and presents a structured and substantiated overview of the project. DC metadata is a standard for representing content on the Internet in a formal, shared, accessible and broad applicable language. DC metadata consist of seventeen elements, six of them are obligatory in EASY: Title, Creator, Description, Date (created), Rights and Audience (https://dans.knaw.nl). The full list of DC elements with description is given in appendix C. Archaeology specific elements are location, subject and time period. Project level metadata is also associated with rightsholders, access rights and licenses, as is visible in figure 3.3. Project level metadata is not located within the dataset, but in the 'description' component of EASY that is always accessible for everybody. The project level metadata does not adhere to the licenses and is accessible even if the license and access rights indicate a very restricted or no access. Most datasets in EASY are licensed under Creative Commons licenses (www.dans.knaw.nl). A list of the Creative Commons licenses and what they guarantee to the creator is provided in table 3.1 and is derived from Pejšová and Vaska (2014, 6-7). EASY metadata is represented as a language that is both readable for both humans and computers by its XML language (Tsoupra *et al.* 2018, 9-10).

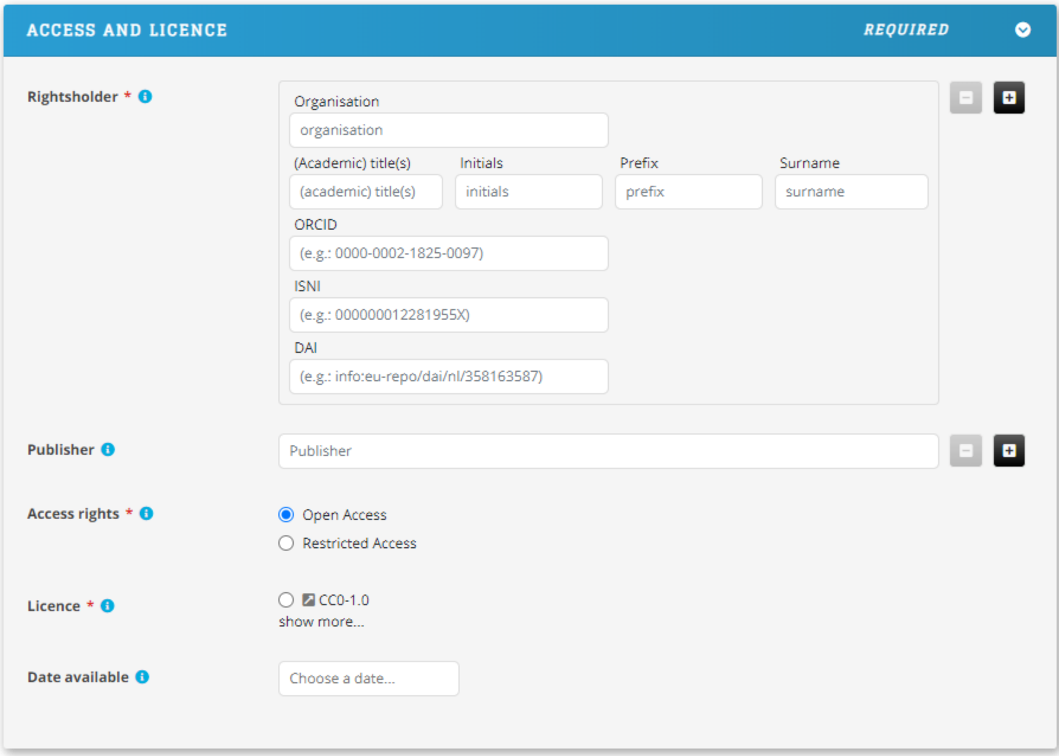


FIGURE 3.3: ACCESS AND LICENSE SECTION OF THE PROJECT LEVEL METADATA IN EASY (WWW.EASY.DANS.KNAW.NL).

CC0	Means no rights reserved. It provides the opportunity to opt out any copyright and protection of databases.
CC BY	Means attribution. Allows others to distribute and build further upon the work of the creator, as long as the original work is given credit.
CC BY SA	Means attribution or ShareAlike. Allows others to build further upon the work of the creator, as long as credit is given to the original work and the same terms are used for the license of the new creation.
CC BY ND	Means attribution and no derivatives. Allows for redistribution, as long as the work is credited to the original and is unaltered.
CC BY NC	Means attribution and non-commercial. Allows others to build upon the original work, but only for non-commercial purposes. The new work must also acknowledge the creator.
CC BY NC SA	Means attribution, Non-commercial and ShareAlike. Allows others to build upon the original work, but only for non-commercial purposes. The new work has to acknowledge and credit the creator and the license of the new creation has to be identical to the original.
CC BY NC ND	Means attribution, non-commercial and no derivatives. Puts the most restrictions on the work of all the licenses. Others are allowed to only download

	and share the original work. But only if credit is given and the original is not changed and used for commercial purposes.

TABLE 3.1: LIST OF CREATIVE COMMONS LICENSES AND WHAT THEY RESTRICT FOR OTHER USERS (PEJŠOVÁ AND VASKA 2014, 6-7).

3.2.2 File level metadata

Metadata on the file level addresses technical and content descriptions of each file separately (DANS 2011, 2). It is generally stored in a database or spreadsheet in the form of a list where each file has multiple characteristics described. The file level mostly addresses description metadata and is an enhancement of each file. Lists of these files are very useful for clarifying content specifications and giving an overview all the files. Specifications that are necessary to insert in a file list are 'File_name', 'File_content', 'Software' and 'Othmat_codebook', visible in figure 3.4. A complete list of specifications, including non-necessary specifications, is visible in part two of appendix C.

File list are stored in an XML format and therefore incorporate specific description constraints. Particular punctuation marks that cannot be used in XML files and should be avoided are ampersand (&), smaller than (<), larger than (>), quotes ("), percentages (%) and umlauts (Ä). Metadata file lists can partially be generated automatically. The automatic extraction of information depends mainly on the file naming. It is therefore important to name files based on a specific order that presents information in a useful way. For example, files can be named after the site of the archaeological artefact or the acquisition technique in combination with the unique number or name each artefact has been given.

file_name	file_content	software	othmat_codebook	notes
META.rtf	Kanttekeningen bij de dataset	WordPad		
V20_BORB.DBF	profielgegevens boringen vindplaats 20	dBase III	V20_boringen.doc	
V20_BORP.DBF	kopgegevens boringen vindplaats 20	dBase III	V20_boringen.doc; META.rtf	
V20_boringen.doc	metadata bij boringen vindplaats 20	MS-Word 2003		
Archeologisch_rapport_V20.pdf	Rapportage onderzoek vindplaats 20	Adobe Acrobat 5.0		
V20_bp.TAB	kaartlaag: boorpunten vindplaats 20	Mapinfo		
V20_BPN.TAB	kaartlaag: boornummers vindplaats 20	Mapinfo		
V20_kd.TAB	kaartlaag: kader vindplaats 20	Mapinfo		
V20_top.TAB	kaartlaag: topografie gebied vindplaats 20	Mapinfo		
AW	Specialistentabel: Aardewerkdeterminatie		REF_AWBK; REF_AWBS	tabel in DB_2.2.0.mdb
FOTOBEST	Lijst van alle fotobestanden met fotonummers			tabel in DB_2.2.0.mdb
FOTOLST	Fotolijst			tabel in DB_2.2.0.mdb
PROJECT	Lijst met belangrijkste projectgegevens			tabel in DB_2.2.0.mdb
PUTVLAK	Lijst met alle gebruikte put-vlak combinaties			tabel in DB_2.2.0.mdb
REF_AWBK	Referentietabel aardewerk baksels Middeleeuwen			tabel in DB_2.2.0.mdb
REF_AWBS	Referentietabel aardewerk Romeinse baksels			tabel in DB_2.2.0.mdb
RIMG0001.jpg	Veldfoto		FOTOBEST; FOTOLST	FOTOBEST is tabel met koppeling bestandsnaam aan fotonummer in FOTOLST
RIMG0002.jpg	Veldfoto		FOTOBEST; FOTOLST	FOTOBEST is tabel met koppeling bestandsnaam aan fotonummer in FOTOLST
RIMG0003.jpg	Veldfoto		FOTOBEST; FOTOLST	FOTOBEST is tabel met koppeling bestandsnaam aan fotonummer in FOTOLST
V20_ARCHIS.doc	Archis-gegevens vindplaats 20	MS-Word 2003		
V20_ra.dwg	profiel boorraai A vindplaats 20	AutoCAD		
V20AI01.wmf	voorkomen archeologische indicatoren vindplaats 20		META.txt	
V20HK01.wmf	voorkomen houtskool vindplaats 20		META.txt	
V20ZD01.wmf	zanddieptekaart vindplaats 20		META.txt	
V20_VONDSTEN.doc	beschrijving vondsten vindplaats 20	MS-Word 2003	META.txt	

FIGURE 3.4: A DUTCH EXAMPLE OF A FILE LIST (DANS 2011, 8). THE ESSENTIAL COLUMNS: FILE_NAME, FILE_CONTENT, SOFTWARE, AND OTHMAT_CODEBOOK ARE DISPLAYED IN THE TOP ROW.

3.2.3 The codebook

Variables and codes that are project specific are stored in a *codebook*. The codebook displays used abbreviations and what these abbreviations mean. Codebooks also display which parameters are used during data acquisition techniques and which problems occurred them, or in other words the paradata. The codebook file is created for future users of the dataset to determine and evaluate the digital files. The structure of a codebook is not obliged to be produced in accordance with specific guidelines but is rather structured according to the structure of the research project. The highly specific 3D heritage metadata schema *CARARE* can be of use to address and display specific 3D object related information (https://pro.carare.eu). CARARE is compliant to EASY, is extremely extensive and utilizes many of the DC metadata concepts (Tsoupra *et al*. 2018, 8).

If a dataset contains multiple data files, it is also possible to generate a codebook for each file separately. However, this is only useful if these data files have different acquisition techniques and are made for different purposes. Generally, each group of files with similar traits have one codebook. In Dutch commercial archaeology, the concept of a codebook is generally already implemented by the use of a PvE (programma van Eisen, in English *Brief*). In the file list metadata, it is important to refer to the codebook if each file in the othmat_codebook header. Only then are future users of the dataset able to easily ascertain where a list of abbreviations of the dataset is and what each abbreviation means. Codebooks are preferably stored in a preservation friendly and sustainable ASCII format in an XML structure.

3.3 FAIR Data Principles

This chapter first addresses the background of the FAIR principles, followed by an extended display of each principle separately. The target of the FAIR data principles is to bring clarity around the goals and urgencies of good data management and stewardship (Wilkinson *et al.* 2016, 1). Good data management is an essential element for knowledge discovery and innovation, for integration of data and knowledge and, after the data is published, reuse by the community. FAIR principles are a set of guidelines that make data Findable, Accessible, Interoperable and Reusable. The principles are simple guideposts that help inform researchers and those who publish and preserve scholarly data. Besides human research, the data principles are also facilitating computational applications for data analysis and data retrieval or 'computer stakeholders'. Computer stakeholders are demanding more attention as their relevance grows and their knowledge production improves. The principles are described in short in figure 3.5.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

FIGURE 3.5: THE LIST OF FAIR PRINCIPLES (WILKINSON *ET AL.* 2016, 4).

The concept of the principles is conceived during the 2014 Lorentz workshop[2] in Leiden called 'Jointly Designing a Data Fairport'. During this workshop it became evident that a minimal set of community-agreed guiding principles and practices are useful for human and computational stakeholders. With these guiding principles, human and computational stakeholders can more easily access, integrate, discover, cite, and reuse generated data of contemporary data-intensive science (Wilkinson *et al.* 2016, 3). Their strength is in simplicity and flexibility, and therefore provides common ground for developments of data and metadata standards and between shared agendas (Boeckhout *et al.* 2018, 935).

The principles are not constructed to be a standard or specification. The principles act as a guide for assisting data publishers and stewards in evaluating specific choices of implementation and for making digital data Findable, Accessible, Interoperable and Reusable. Their influence within European research funds is strong. Even though they are not constructed to be a standard, they are slowly converting into an essential element of policies in research and research data management plans (Boeckhout *et al.* 2018, 931).

3.3.1 Findable

To make data findable, a globally unique and persistent identifier needs to be added to the (meta)data (Wilkinson *et al.* 2016, 4). The findability principle not only focusses on researchers to make data easily findable, but digital archives are also expected to participate. They participate by assigning a globally unique and persistent identifier to the (meta)data. The findable principle

[2] This specific Lorentz workshop was organized by a collaboration between Barend Mons and the Lorentz center, a Dutch Technology center for scientific workshops in all disciplines.

ensure that data should be identified, described, and registered or indexed in a clear and unequivocal manner and in a searchable resource (Boeckhout 2018, 932). The Findable principle exists of four components, as depicted in figure 3.5.

F1 describes the assignment of a globally unique and persistent identifier to (meta)data. The unique identifier is according to the official website of the FAIR principles arguably the most important principle. Mainly because it lays the groundwork for the other aspects of the principles (www.go-fair.org). A persistent and globally unique identifier discards ambiguity in hyperlinks by assigning a unique identifier to every element of (meta)data and datasets. The identifiers stipulate two conditions:

1. The identifier is to be globally unique. It is possible to obtain globally unique identifiers by contacting a registry service that utilizes algorithms which guarantees the uniqueness of newly created identifiers.
2. The identifier is to be persistent. Persistency is important as Internet links tend to expire or become invalid over time, because it takes time and money to maintain online links. Registry services can provide resolvability in the future, to a certain extent.

These identifiers are generally generated by digital data repositories and consist of a unique Internet link. An identifier does not only help (other) people to understand and find the data, it also provides use for citation purposes and helps computers interpret data for relevant information (www.go-fair.org). Researchers are responsible for putting this unique identifier clearly and explicitly in the metadata, which allows repositories and archives to register, index and harvest the data and the metadata (Wilkinson *et al.* 2016, 4). An example of repository harvesting is the Open Archival Initiative Protocol for Metadata Harvesting (OAI-PMH) and is developed to gather metadata between and from digital repositories. EASY is part of the OAI-PMH.

Examples of websites and identifier systems that provide globally unique and persistent identifiers for digital archives are:

- Digital Object Identifier (DOI) is a persistent identifier with a wide use in professional, governmental, and academic information. Each dataset in EASY has a unique DOI assigned during deposition. DOIs are resolved at http://www.doi.org.
- URN:NBN is a persistent identifier that functions on both national and international level and is specifically built for national libraries. For example, for the Netherlands URN:NBN:NL. Although URNs are less common, DANS is involved in the URN identifier project and assimilates this PID in datasets in EASY as well.

F2 is focused on rich metadata descriptions of data. Metadata should be generous and extensive and should include descriptive information regarding context, quality, state of the data, and characteristics of the data. A good example of technical metadata is table 2.1 in chapter 2.3.2. Elaborate metadata allows computers to do automatic sorting and routine searches, allowing researchers to prioritize their work and workflow. Rich metadata can be perceived as a separate or different approach of finding data without having the identifier of the data (www.go-fair.org). F3 addresses the inclusion, linkage, and explicit mentioning of the unique identifier within the metadata. The mentioning of the identifier also helps with enriching the metadata and interlacing the data and its related metadata.

A dataset, digital repository, or service can be hidden for search algorithms if it is not indexed in a searchable resource such as a (big) search-engine, which is the focus of F4. An example of a search engine almost everybody uses is Google. Google automatically indexes web pages and performs this action also on scholarly data. However, indexing scholarly data should be carried out with effort and care, to obtain the optimal distribution and findability of the data(set).

3.3.2 Accessible

The second principle is accessibility and addresses the retrievability of datasets through a clearly defined and preferably automated access protocol (Boeckhout 2018, 932). The protocol should be free, open, and universally implementable. Where necessary, the protocol also has to address authentication and authorization procedures for access of the data and metadata (Wilkinson *et al.* 2016, 4). The metadata of a dataset should always be accessible regardless of the availability or context of underlying data (Berchum and Grootveld 2016, 77; Boeckhout 2018, 932).
A1 is focused on the retrievability of (meta)data by their identifier in combination of using a standardized communication protocol. Clicking a link on a website, regardless for what purpose, results in the computer executing a protocol called a transmission Control Protocol (TCP). The TCP should, according to the accessible principle, be mediated without communication methods or specialized tools. Which means that a clear definition and description of the people that can access the data and how this access is acquired, needs to be defined. An example of a TCP is HyperText Transfer Protocol (HTTP) and is used for many hyperlinks to and from Internet websites.
A1.1 describes that for data reuse purposes, a TCP should be open(-sourced), free (of costs) and thus facilitating data retrieval if implemented on a global scale. A1.1 does not implicate that data can be obtained without identification of the consumer or without costs but indicates that the exact conditions for the accessibility of the data should be provided (www.go-fair.org). A1.2 focusses on this concept as well by allowing digital archives to authenticate owners and setting user-specific access rights.

A2 addresses the accessibility of metadata and states that anyone with access to the Internet should be able to access the metadata of a dataset. However, that concept does not suffice for the data(set) itself. Data can contain privacy sensitive information. Thus, for these cases it is perfectly reasonable to only provide an email, telephone number or (skype)name of the contact person. Which provides the possibility for future users to discuss if access to the data is still possible.

3.3.3 Interoperable

I1 describes that researchers should be able to obtain and interpret the data of another researcher, preferably in a readable format without the need of a translator. This concept especially applies to computers, as computer systems need to know the data exchange formats of other computer systems to be able to read the transferred data (www.go-fair.org). A combination of both human and computer readable data provides the best potential of data longevity. Digital archives therefore requests data to confine to such exchangeable file formats.

Interoperability should be performed through the utilization of standardized and commonly used vocabularies, jargon, and metadata schemes, as described in I2. The vocabulary and jargon of any research should be documented thoroughly and must be easily findable within the dataset for consumers of the dataset. An example of a metadata standard that is used within many digital archives, including EASY, is the DC metadata schema, earlier described in chapter 3.2.1 (Tsoupra *et al.* 2018, 6). Within archaeological datasets, codebooks and file lists are used for describing content of the datasets. These documentation methods are stored within datasets but should utilize the appropriate data exchange formats as well.

The interoperable principle in I3 addresses the cross-referencing between different datasets. Not only is an indication of the association between relating datasets valuable, the intent of the reference is even more meaningful. The intent of referencing and citing could be, for example, to build upon previous datasets, or to complete data. The citing should be performed properly and include the globally unique and persistent identifier.

3.3.4 Reusable

R1 describes that reuse of data is facilitated, if many labels are attributed to the data and metadata. The reusability principle partially overlaps with the findability principle. However, the difference is situated within the decision of finding data useful in a specific context. R1.1 focusses on the necessity of licensing of data for reusability purposes. Reuse depends on legal acquisition and the conditions for acquiring legal data should be clear and applicable for both researchers and machines. Research data is generally licensed under Creative Commons (CC), which are described earlier in chapter 3.3.1. The CC licenses are a license description offering authors, researchers, artists, and teachers the freedom to flexibly deal with copyright. With CC-licenses the owner maintains the rights but allows certain permissions to let other researchers or artists to use and/or share your work. The cc license of the work should be clearly presented within the metadata.

Metadata must provide information for findability but should also provide context of provenance and how generation of the data occurred. The creator of the (meta)data should not anticipate the intent of the consumer, thus metadata should be provided in its most extensive and richest form, as described in R1.2. Even if information seems insignificant or trivial. R1.3 indicates that data also must meet domain-relevant and domain-specific standards. For an archaeological context, the purpose for collection, the limitations, the documentation, and the version of the data are essential elements of metadata.

The outline of this chapter confines itself mainly to the preferences and concepts of digital archives. The chapter started with explaining what digital preservation is and how conserving digital data guarantees accessibility, safe storage, and understandability of data in the future. Digital preservation is however very complex and depends on preservation strategies and infrastructure to preserve data correctly. These strategies and infrastructure demand specific requirements of data deposits to minimize the potentiality of data loss. EASY is a digital academic archive of DANS that stores, among other things, archaeological research data. The policies of EASY generate certain specific requests of deposited data that is related to file formats, but also to documentation. Documentation in EASY is mainly performed by generating information about data or metadata. Documentation is divided in three levels: the project level metadata, file level metadata and the codebook. The file level and codebook are implemented within the dataset and the project level metadata is implemented after the dataset is complete, and takes place during the ingest of the data in EASY.

Documentation and digital preservation have been made easier by the FAIR principles. The FAIR principles are guidelines to make data Findable, Accessible, Interoperable and Reusable. Although they are not constructed to be a standard, they are slowly converting into an essential element of policies in research and research data management plans. They are produced to facilitate the generation of human and computer readable data, which provides the best potential of data longevity.

Now the requests of the perspective of digital archives are clarified, the user and technical perspectives can be assessed. These perspectives are more focused on 3D modeling and how they can be used in research.

Chapter 4: Object-based 3D models in archaeology

This chapter first addresses the user perspective of object-based 3D models in archaeology, by assessing the purposes of 3D models. Next, the chapter becomes more technical and discusses three very commonly used 3D data acquisition techniques. After these techniques, the focus will be towards 3D modeling features. 3D modeling features are explained to understand why storing 3D models is difficult and why many 3D file formats currently exist. The last part of this chapter presents six 3D file formats that are used frequently in many 3D modeling applications and are recommended in EASY by DANS.

4.1 Fit for purpose

Object-based 3D models are used for a plurality of reasons in archaeological and cultural heritage contexts. First, the purposes and use of 3D models in museums are addressed, followed by purposes within archaeological research. The archaeological research part of this chapter is focused on analysis of ceramic 3D models in specific. The reason for generating a 3D model is solely dependent on what the user of the 3D model wants to do with it. Therefore, the fit for purpose is considered as the user perspective.

4.1.1 Object-based 3D modeling purposes in museums

In museums object-based 3D models are mainly used for interactivity, immersion, augmented reality, virtual reality, and heritage preservation. Interactivity is generally accomplished using tangible experiences of archaeological material. Tangibility is a useful aid to explain the backstory of objects and helps to bridge the gap between academia and society (Revello Lami *et al.* 2018, 72). Interactivity in museums is generally actualized by using 3D visualizations on touchscreens, which can be played with to help museum visitors understand the geometry of vessels. Another option is to 3D print replicas of objects. It allows manipulation of replicas if the original object is fragile and can also be used to visually reconstruct additional parts that were missing or are potentially historically and archaeologically accurate (Stenborg 2018, 112). Movability and interaction with archaeological objects enthuse the imagination of museum visitors and engages and educate audiences to comment and critically think of the past (Revello Lami *et al.* 77). 3D models for museum visitors need to portray 3D models on touchscreens effortlessly, indicating that the 3D models need to be smooth and cannot be of very high quality. High quality models slow down the interaction and maneuverability of 3D model in 3D applications considerably. However, the largest problem of museums is affordability. Smaller sized or decimated 3D models are cheaper to generate and store and are therefore advantageous for 3D visualization for museums (Stenborg 2018, 112). Object-based 3D models are also used as components to immerse visitors in the environment of the provenance or period of museum objects (Petersson and Larsson 2018, 72). Immersion of and within objects can be done by adding visual layers upon the existing environment using projectors and enhancing the reality (augmented reality), as displayed in figure 4.1. Immersion is also used in museums by digitally constructing the complete environment (virtual reality). Augmented reality can also be used to visually reconstruct additional non-existent parts but require highly accurate 3D models to correctly portray the additional parts.

FIGURE 4.1: DISPLAY OF A 3D MODEL IN THE MUSEUM OF HISTORY AND INDUSTRY IN SEATTLE (HTTPS://NEWS.ARTNET.COM).

In the online digital collections of museums, 3D models are used to help visitors to learn more about archaeological materials and topics, before or after a museum visit (Petersson and Larsson 2018, 72). The online digital museum collection is no substitution for an actual physical visit; thus, the objects need to be appealing and catch the attention of visitors. Although the feature quality is important, object-based 3D models should be small, for similar reasons as described for 3D visualizations on touchscreens.

The last purpose of 3D models for museums is to preserve the contemporary heritage, which is also very valuable for archaeological research. 3D models can function as a digital back-up that functions as a time capsule of a snapshot of the original object. 3D models are separate entities than the sometimes-fragile original object and cannot be harmed, lost, or destroyed in a similar fashion as the original object. 3D models do however have their own limitations, which will become evident when file formats are discussed later this chapter. To preserve the most accurate depiction of the original object, these back-ups are preferred to be highly accurate, accessible, findable and have exceptional longevity.

4.1.2 Object-based 3D modeling purposes for archaeological research

In archaeological analyses, object-based 3D models are used to display archaeological restoration, to investigate the meaning of the production and to understand technical knowledge embedded in materials (Landes *et al.* 2019, 409; Revello Lami *et al*. 2018, 63).

The main target for archaeological restoration is to generate a realistic representation or close approximation of what exists now or of what may have existed once (Rua and Alvito 2011). This target is performed by digitizing what is left of artefacts and digitally adding the leftovers based on estimations and comparisons. These leftovers can be focused on realism or represent uncertainty using color scales, as displayed in figure 4.2 (Landes *et al.* 2019, 410). For objects using color scales, the quality of the model is less important and only depends on the accuracy of the volumetric measurements. However, restorations based on realism require accurate and very high-quality 3D models that can be integrated in virtual reality environments.

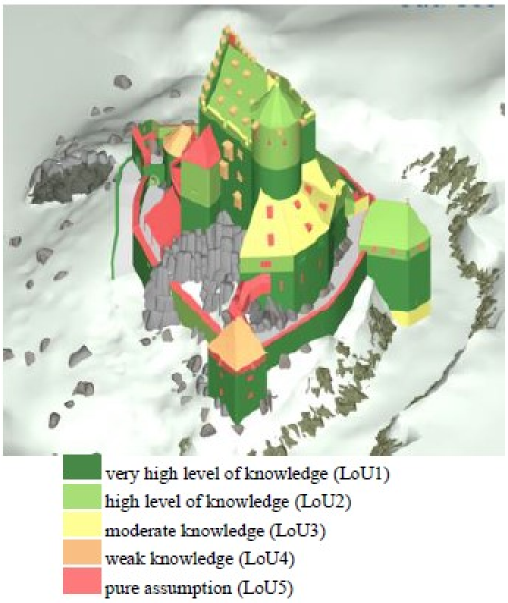


FIGURE 4.2: EXAMPLE OF AN UNCERTAINTY VISUALIZATION, OF THE RUINS OF THE CASTLE OF KAGENFELS IN FRANCE USING COLORS (LANDES ET AL. 2019, 414).

3D imaging techniques and 3D analytical tools are used to investigate the meaning of the production and technical knowledge embedded in ceramics (Revello Lami *et al*. 2018, 63). The tools and techniques assist in understanding the complexity of social relationships manifested in the manufacturing and use of ceramics. And are used to obtain more quality and quantity of archaeological information that is imbedded in ceramic vessels. Within ceramic analysis, the integration of 3D scanning technologies demonstrates that standard procedures of macrotrace analysis enable to record the surface topographies of vessels more systematically and in greater detail than analyzing the object with the naked eye. Using a combination of contours, decimations algorithms, smoothing and curvature analysis, surface features are more easily visible to detect (Revello Lami *et al*. 2018, 64). With the help of 3D comparison software, 3D data reinforces manually derived conclusions. Future archaeological ceramic analyses can potentially use 3D technologies to automate the recognition of decoration and features. To realize these analytical purposes, high-resolution, high-accuracy, photorealism, completeness of data and user interactivity are top priorities and serve as a trust marker for future data reuse (Remondino and El Hakim 2006, 274).

Digitally preserving 3D models makes the data more interactive and accessible, which is very useful if the models are used for future verification (Stenborg 2018, 113). Through online platforms, collected 3D assemblages can be compared to other assemblage and easily browsed by specialists and non-specialists. This helps researchers to feel more connected and involved in the reconstruction of stories of the objects. However, when comparing 3D models or searching for specific character traits of a 3D digitized ceramic vessel, smaller and decimated 3D files are more useful for easier and faster compliance. Smaller files are quicker to load and are therefore more easily accessible and suitable for interactive means.

4.2 3D Data acquisition in archaeology

In its broadest sense, the classification of 3D object reconstruction and measurement techniques can be divided into *contact* and *non-contact* methods (Remondino and El-Hakim 2003, 269-270). *Contact* methods use tools such as rulers and calipers to measure the object. *Non-contact* methods use laser scanners and imaging techniques like X-Ray and photogrammetry to model and measure 3D objects.

As mentioned before, this thesis is focused on object-based archaeological 3D models. Implicating that the generated 3D models are based on actual existing archaeological objects. The non-contact methods that will be described in this chapter are specifically for modeling objects based in reality, also known as *reality-based modeling* (Ikeuchi and Sato 2001, 182-183). The other option of 3D modeling is computer graphic recreation of artificial world models, which allows reconstruction of extinct and destroyed objects or of things that have never existed. A good example of a graphic recreation are uncertainty visualizations, described in 4.1.2 and figure 4.2.

Three-dimensional models in archaeology are created to address specific research questions and depend on several characteristics (Dell'Unto 2018, 56):

- Which instrument is used for acquisition?
- Which acquisition process is performed by the operator?
- Which type of post processing is applied to produce the final visualization?

These characteristics must meet certain quality standards, be documented following specific guidelines, and be linked to similar and related projects to be relevant for research purposes (Kansa and Kansa 2013, 2). The increasing production of 3D models and their associated digital libraries in the archaeological discipline create a focus on accuracy, resolution, efficiency, consistency, speed, cost and reliability of acquiring 3D models (Boehler and Marbs 2002, 10; Geng 2011, 148; Poux *et al.* 2017, 96). Generating 3D models requires a careful evaluation of the advantages and disadvantages between multiple 3D data acquisition techniques (Remondino and El-Hakim 2006, 269). 3D data acquisition starts with the 3D recording method, which is dependent on technical aspects such as the complexity of the type, size, shape, diversity, the typology of the recorded material(s) and lastly, the morphological complexity (Level of Detail or LoD) (Dell'Unto 2018, 59; Poux *et al.* 2017, 96). Other factors related to budget, time (management), user experience and availability limit the instrument and technique of choice as well (Geng 2011, 148, Poux *et al.* 2017, 96). It is also relevant for each 3D scanning technique to consider if the measurements are performed onsite (outside) or indoors, as angle of incidence of light and the surroundings influences the outcome 3D models (Rieke-Zapp and Royo 2017, 247). The last aspect that should be considered before starting to model 3D objects is the purpose of creation, or the earlier described user perspective. If the purpose is not taken into consideration, the generated 3D models have barely any or no value for future generations.

For the data of this thesis, only photogrammetry is used to generate the models. However, this is not the only commonly used method of generating 3D models in archaeology. Two other popular methods, the laser scanner and the structured light scanner are addressed in this chapter as well. These three methods are all non-contact methods and have the advantage that the created digital point clouds do not disturb the sometimes-fragile surfaces of archaeological objects (Hess 2017, 199; Rieke-Zapp and Royo 2017, 247).

Only photogrammetric generated 3D models will be subjected to the workflow of this thesis, but the other techniques have a similar procedure of storing 3D data. These procedures are similar because the generated point clouds can be stored using the same geometry features and are stored in equivalent 3D file formats.

4.2.1 Laser scanning

Laser scanning has been defined as a fast-performing method and provides automatic data acquisition of high-density 3D coordinate points (Grussenmeyer *et al.* 2018, 305). Laser-scanning of archaeological objects should not be confused with airborne or terrestrial laser scanners used in landscape surveys (Mara and Portl 2013, 25). Laser scanning is an active sensing system, creates its own electromagnetic radiation and use the principle of Time-Of-Flight (TOF). The method is an active remote sensing system as it creates its own electromagnetic radiation. The 3D points of an objects surface are collected in a systematic pattern, in near real time (Boehler and Marbs 2002, 10). The term laser scanner comprises an array of instruments on different levels of accuracy and precision and with different principles in mind (Barber and Mills 2018, 3). Laser scan data can be acquired via different methods. Originally the laser scanning was developed for airborne applications, but has now expanded to be used via tripods, vehicles and for close-range applications, such as handheld or backpack systems (Luhmann 2013, 240). The close-range applications are especially useful for free movement around the object and for scanning objects at difficult to reach locations such as archaeological sites (Hess 2017, 200; Historic England 2018, 3).

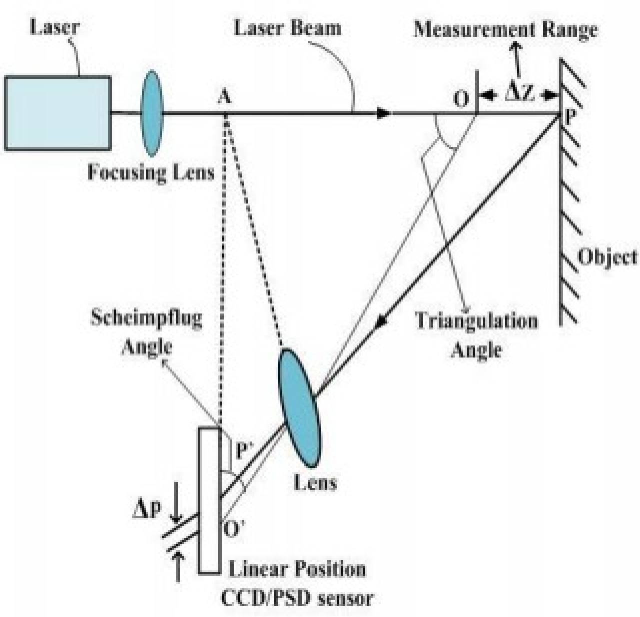


FIGURE 4.3: AN EXAMPLE OF THE CONCEPT OF A LASER SCANNER (ABOALI *ET AL*. 2017, 2414).

Scanning of medium and small sized objects (15 to 600 cm) is calculated through a *triangulation* system (Hess 2017, 200). Acquisition through triangulation is depicted in figure 4.3 and is conducted by using a laser that generates a beam toward the object's surface, which then reflects off this surface at a specific angle (described in figure 4.3 by *P*). This beam is captured via a lens on a linear positioned detector or sensor and has a depth relatively high accuracy ranging from mm to cm (Aboali *et al.* 2017, 2414). The degree of angle is based on the distance between the object and the laser. If the object is closer, the distance changes and the laser is portrayed on a different area on the detector and measures the distance.

Advantages of laser scanning are the ease of implementation and the independence on illumination sources (Jeciç and Drvar 2003, 244). These advantages helped this technique to become one of the most prevalent technologies on the global market of 3D recording methods (Aboali *et al.* 2017, 2414).

A disadvantage of laser scanners is that an additional camera is required to accurately acquire color and grayscale information of an object's surface (Mara and Portl 2013, 26). Although the surface color information could be captured by laser scanners, it is practically much more useful to use images to capture the object (Barber and Mills 2018, 10; Hess 2017, 199). The camera makes color photos of the object that are wrapped around the laser scanned point cloud. These wrapped photos form a texture, which is further explained in chapter 4.3.2.

4.2.2 Image-based modeling and Photogrammetry

There are multiple methods of 3D model generation using images, the most popular method is Photogrammetry (Guery *et al.* 2017, 229). Photogrammetry utilizes image measurement and interpretation methods to derive the shape of an object by overlapping multiple images of an object from different locations and angles making an *image network* (Luhmann *et al.* 2014, 2; Robson *et al.* 2012, 92). Photogrammetry is categorized as a passive remote sensing system because photography only records reflected light from surfaces illuminated by a constant natural or artificial lighting and is not dependent on its own self transmitted light (Barber and Mills 2018, 52). This method is based on imagery, which means that the size, shape, and type of the object do not matter if enough images of the object are made. More images do however require significantly more processing time.

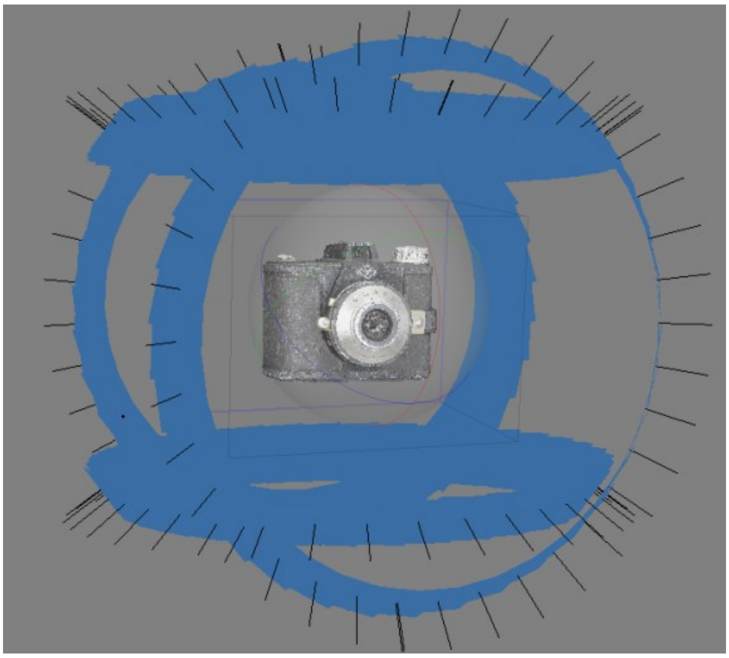


FIGURE 4.4: AN EXAMPLE OF A POINT CLOUD GENERATED IN AGISOFT PHOTOSCAN (CURRENLTY KNOWN AS AGISOFT METASHAPE). THE BLUE AREAS WITH THE BLACK LINES REPRESENT THE LOCATION AND POISITION OF THE CAMERA, WHEN THE IMAGE WAS TAKEN (WWW.3DSCANEXPERT.COM).

A relatively new automated technique for photogrammetry is the use of Structure from Motion (SfM) and Multi-View Stereo (MVS) (Chandler 2016, 2). SfM integrates computer vision algorithms to efficiently and automatically match images if parts of the images overlap. SfM photogrammetry relies on the combination of matching distinctive features on images, the artificial position of the camera and the interior optical parameters to produce an accurate 3D surface measurement, as depicted by the blue rectangles in figure 4.4 (Guery *et al.* 2018, 230). The blue rectangles indicate the artificial position of the camera and combined with the lines that pop out of the rectangles, the interior optical parameters are measured. The position of the camera and the interior parameters are essential for photogrammetry because it contributes to automated image features and area matching procedures. The measurements of scale are dependent on adding scalebars to one of the used images, or the model can rely on known lengths or separations between camera pairs (Guery *et al.* 2018, 230; Robson *et al.* 2012, 93). Without scale information the size of the object in the model is impossible to establish precisely.

A disadvantage of photogrammetry is the dependence on a constant natural or artificial lighting source. If absent, the 3D model will not include a homogeneous color texture. The accuracy of the model is dependent on the camera quality, the environment of the object, the distance between the images and the object and the soft- and hardware utilized. Using photogrammetry outside is therefore more difficult than inside.

4.2.3 Structured Light Scanning

SLS is a technique that is utilized to create a 3D representation which can be moved, rotated, and magnified in a virtual environment (Errickson 2017, 94). During the scanning procedure the technique projects a known number of light patterns, that vary from dots to stripes, onto an object (Rieke-Zapp and Royo 2017, 247; Jeciç and Drvar 2003, 238). The pattern is typically projected in a regular and periodic process and its reflectance, distorted by the object shape, is captured via one or more cameras. The captured data is recovered by dedicated software using projection geometries or triangulation. SLS can produce very dense and accurate point clouds and can therefore create a seamless high-resolution 3D image of a scene (McPherron 2009, 23). SLS generates its own structured light patterns and is therefore an active remote sensing system, as is visible in figure 4.5. It also relies on the contrast of light patterns and should therefore be performed indoors or in shaded areas as these areas have very little natural or ambient lighting. Natural or ambient lighting should be avoided as much as possible. The scanning of reflective or transparent surfaces, such as shiny metal, glass, bones, marble, or teeth, generate scanning problems for any optical scanning system (Rieke-Zapp and Royo 2017, 247). A solution for scanning such surfaces is using a whitening spray, although this is inadvisable for archaeological objects, because of the permanent damage it causes (Fink 2017, 228).

The covered area per image is dependent on the size of the object, the brightness of the projector and on the (feasible) distance between the object and the camera. The scanning distance and area of coverage depends on its relation to depth resolution, as larger area coverage produces lower resolutions and accuracy. Spatial resolution is based on the samples per unit distance, which means that each pixel on the image contains a certain number of samples. If the camera is moved closer to the object, the pixels can capture better depth detail, which results in better accuracy. The solution for large or complex objects is to capture multiple scans from different angles and merge these scans into a single 3D model.

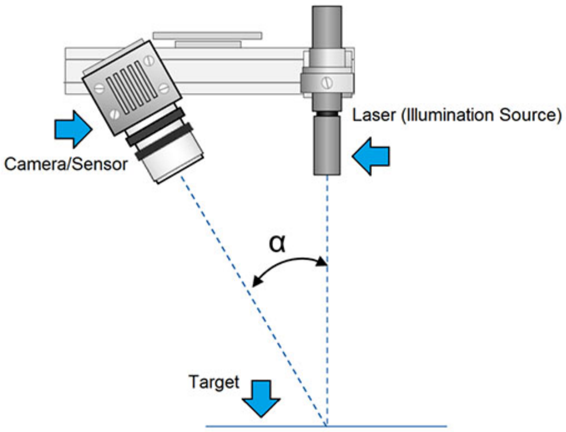


FIGURE 4.5: CONCEPTUAL PLAN OF STRUCTURED-LIGHT SCANNER. THE LASER IS THE ILLUMINATION SOURCE, MEANING THAT THIS IS AN ACTIVE REMOTE SYSTEM (DERIVED FROM WWW.3DNATIVES.COM)

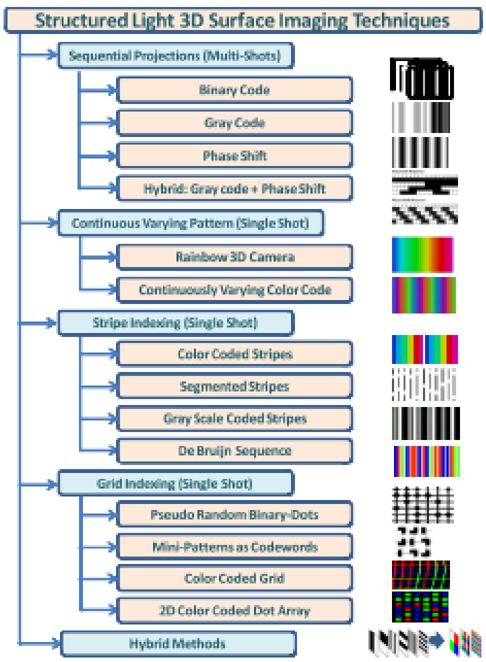


FIGURE 4.6: THE VARYING METHODS OF CAPTURING STRUCTURED LIGHT (GENG 2011, 133).

SLS is meant for short-range 3D scanning. Scanning from greater distances will heavily limit the quality and accuracy, because the generated light will not reflect sufficiently from the scanned object. Scanning large objects also generates huge amounts of redundant data, which negatively affects the scanning duration, post processing and the need for storage space (Jeciç and Drvar 2003, 243).

The projector of SLS works with light of the visible spectrum or near-infrared by using a monochrome or color capturing camera. The monochrome cameras capture more light per pixel, and produce less pixel noise, resulting in high quality images. Color cameras are advantageous if both color and shape need to be captured simultaneously. There are multiple ranges of capturing structured light imaging, as described in figure 4.6. The main advantages of SLS are the high accuracy that can range from µm to cm per pixel and the high data acquisition rate (Aboali *et al*. 2017, 2413).

The disadvantages of this technique are accredited with missing data in correspondence to occlusions and shadows, the medium quality of fine details such as edges and steps, the problems with scanning transparent or reflective surfaces and the high software complexity (Aboali *et al.* 2017, 2417; Jecić and Drvar 2003, 239; Rieke-Zapp and Royo 2017, 250).

4.3 Modeling 3D features

3D data consists of four feature categories: geometry, appearance, scenery, and animation (McHenry and Bajcsy 2008, 1-3; www.loc.gov). Each category is explained from a technical perspective. The features of 3D models need to be defined to get an accurate understanding why an abundance of 3D file formats currently exists. 3D modeling features have significant value for deciding which file format to use and therefore, for the fit for purpose of the 3D model generation. However, these features are not the sole reason for choosing 3D file formats, preservation and storage are concepts to consider as well. Preservation considerations of file formats have been established in chapter 3. Chapter 4.4 will discuss 3D file formats. First, the four 3D modeling features.

4.3.1 Geometry

The geometry of a 3D model is generally stored through a combination of 3D points, called *vertices* (singular is vertex), figure 4.7. Vertices are described by their position on a(n) X-, Y- and Z-axis in a Cartesian coordinate system. These positions are generally linked to an arbitrary system, but can, if its relevant for the purpose of the 3D model, be supported within real-world coordinate systems (www.guides.archaeologydataservice.ac.uk). Links or lines between two vertices are called *edges*.

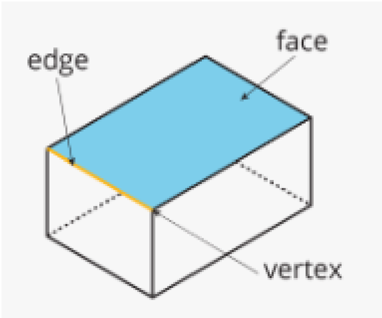


FIGURE 4.7: THE VERTICES ARE THE POINTS THAT REPRESENT THE CORNERS, THE EDGES ARE THE BEAMS BETWEEN CORNERS AND THE BLUE AREA REPRESENTS THE AREA OF A FACE (HTTPS://WWW.PNGITEM.COM).

If multiple vertices are connected to form an enclosed area, a polygon or *face* is created. A face requires at least 3 vertices, although more vertices are possible. Faces from 3 vertices is the most common type in 3D modeling (McHenry and Bajcsy 2008, 1-3). The concept of creating triangles from vertices is called triangulation. Faces and polygons are technically not the same thing, as a triangle face or quad face has three and respectively four edges. A polygon can exist of multiple faces as it covers more edges. Generally, faces and polygons are terms often intermingled because an individual polygon is equal to a face. For consistency, both polygons and faces will indicate the same concept in this thesis.

A model made of only vertices is called a point cloud. A model with a combination of vertices and edges a wire-framed model and if faces are added, it is called a mesh. Applications that utilize 3D graphics can render point clouds into meshes quickly but require a lot of triangulation for situations of scaling and extreme precision (www.al3dp.com). This has the side effect of demanding more storage space. A 3D polygonal mesh does not describe the exact shape of an object, but only approximates a smooth geometry that comes close to the geometry of the actual object. Therefore, this technique is called an *approximate mesh*. These type of meshes are mostly used within object-based 3D modeling, because approximate meshes are generated by the previously mentioned 3D acquisition techniques of chapter 4.2.

Sometimes approximate meshes do not provide enough geometrical accuracy for a mesh, for example with ultrafine resolutions. To create a smooth surface, *Non-Uniform Rational B-Splines* (NURBS) are used, figure 4.8. NURBS are so accurate that the term *precise geometry* is used. NURBS are a form of modeling that works with relatively few vertices and uses curved surfaces through the utilization of a small number of weighted control points and knots, which define a set of parameters (Rabin 2009, 669). With the help of these knots and control points, computers can calculate the surface through interpolation, as visible in figure 4.8. If the parameters of the dimensions of the model change, e.g. for scaling, there is no loss of detail with using NURBS, because the surfaces are based on mathematically described curves and these calculations scale coherently. Regardless of the scale of the model, the representation of the model remains the same (www.guides.archaeologydataservice.ac.uk). NURBS are therefore also described as *parametric representations*.

Interoperability between precise meshes and approximate meshes often leads to loss of data. The calculations of precise meshes are to be converted from a parametric representation to a wire-frame model using triangulation and leads to information loss of the surface structure. The created mesh might also require an enormous number of polygons if the conversion quality is set to (ultra) high. This process is considered as preservation unfavorable, because a very high number of faces requires lots of storage space.

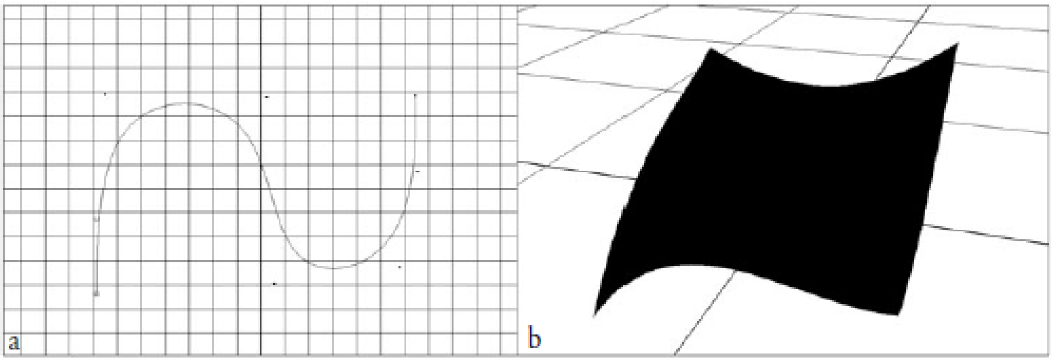


FIGURE 4.8: A DISPLAYS THE MATHEMATICAL CALCULATIONS TO GENERATE 3D SHAPES. B IS THE VISUAL 3D DISPLAY OF THESE CALCULATIONS (RABIN 2009, 669).

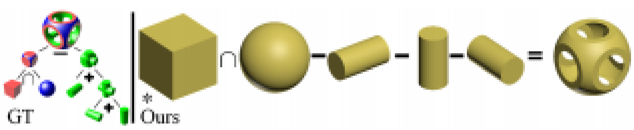


FIGURE 4.9: AN EXAMPLE OF A CSG CONSTRUCTION (WU *ET AL.* 2018, 229).

The last option of developing geometry is through Constructive Solid Geometry (CSG) and can be perceived as a bottom-up approach. Instead of capturing reality-based object and calculating the corresponding digital parameters, models using CSG are created from scratch without using shape indicators of other sources such as images. CSG operates in minimization and utilizes cubes, spheres, cylinders, and other simple geometric shapes (also known as simple geometric solids or Boolean operators) to create a 3D model. These models are constructed by adding, intersecting, or removing these solids together to create other geometrics shapes, as depicted in figure 4.9.

Converting a model from CSG to an approximate geometry can be useful for storage space. However, converting approximate meshes (back) to CSG is very difficult, because CSG functions on minimization and (larger) polygonal meshes are considered too complex for simplistic objects (Buchele and Crawford 1991, 1063; Du *et al.* 2018, 2-3; Shapiro and Vossler 1991, 4).

4.3.2 Appearance

Besides geometrical features, the appearance of 3D models, also known as surface properties, are stored in 3D file formats as well. The surface properties are used for attaching color to meshes and for making objects look more realistic. Color and realism in 3D models are achieved by applying actual derivatives of 2D images onto the mesh. Under the appearance feature we can ascribe the concepts of material type, texture, color and bump maps.

The most simplistic method for applying appearance to models is by associating each vertex within a point cloud or mesh to a set of attributes that indicate a color (RGB, HSV or YUV) value or a certain intensity value (McHenry and Bajcsy 2008, 2-3).

The most common method for applying appearance is by wrapping a texture, which is a flat image, onto the surface of a 3D model by giving each vertex a corresponding point on a related 2D image (Rabin 2009, 688-689). These corresponding points are called a *texture element* or *texel* and represent the pixels on a texture map. Texture mapping, as this technique is called, can be used to project real colors on a 3D model by using images (www.guides.archaeologydataservice.ac.uk). The coordinates of each texel, also known as UV [3] coordinates, are connected to relating XYZ coordinates of vertices of the mesh. The connection can change between different vertices, for example to fit the texture map more accurately on the mesh and allows changes in resolution and location of the texels on the mesh (Rabin 2009, 691).

The image texture(s) connected to the 3D model might display topological features such as glyphs, cracks or bumps that give realism to the model (Corns *et al.* 2013, 22). If these features would be implemented as geometrical features, it would require large amounts of extra polygons. Using images for such details is much more efficient and storage friendly. Another method for recreating realism is by adding a bump-, normal- or transparency map to a model, which generates displacements such as bumps and wrinkles on the objects surface (Max & Becker 1994, 18). These maps adjust the reaction of the light upon the illuminated surface of a 3D model. The bump map does not interact with the geometry and thus avoids unnecessary additions of polygons. See figure 4.10 for a bump map and how it reflects the light.

[3] The letters U and V are coordinate letters similar to the coordinate letters XYZ, used for 3D models. For image textures only two letters are used, because they denote the 2D axes.

A surface of a 3D model can also have a specular component that indicates the intensity and color of a mirror reflection of a nearby illumination source, as well as to other nearby surfaces (McHenry and Bajcsy 2008, 2-3). Surfaces can be transparent or semi transparent and are able to distort light that passes through them. This distortion is depicted by the *index of refraction* but is dependent on the material type of the model. In archaeological context these specular components are barely used, because most archaeological materials like ceramics, bones or stones are not transparent. 3D models are able to reflect light from a light source through bump maps and refractions, but during rendering these models can also utilize *shaders*. Shaders are a set of instructions that give the impression of realistic surfaces of materials, such as smooth, rough, radiating, or metallic objects. These complex materials are referred to as *Physically based rendering* (BPR) models.

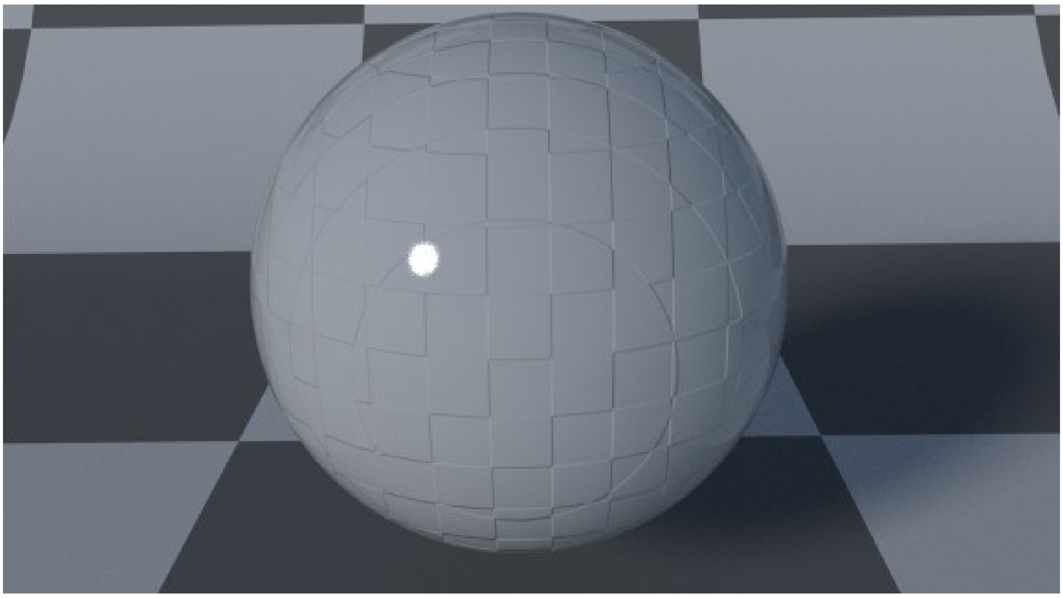


FIGURE 4.10: THE BUMP MAP REACTS AS IF A LIGHT SOURCE SHINES UPON THE BALL (DOCS.BLENDER.ORG).

4.3.3 Scene

he scene of a 3D model is referred to as the setup and environment surroundings of the model, the scale and positioning of the model in contrast to this environment, light sources and other 3D models (McHenry and Bajcsy 2008, 3). To capture a 3D model for a modeling visualization, it is necessary to capture the model in a viewport and from a specific angle through a camera. A viewport is like a stage or location and defines the depth, width, and height of the model in comparison to the surrounding area. For the camera, the position and viewing direction are important to specify what part of a 3D model needs to be viewed and visualized, figure 4.11. Without rendering illumination sources, the camera would only capture darkness and thus at least one or multiple illumination sources are needed to be added to view the object correctly (www.guides.archaeologydataservice.ac.uk). Besides adding a light source, the position, shape, and brightness of each scenery object need to be configured as well. Although it is possible to determine these settings manually, most 3D applications provide automatic pre-sets for convenience and simplicity. If multiple objects exist within a model and the objects are connected to one another in some way, they can be grouped together into one or multiple groups. The position of each object in relation to other objects needs to be stored and can be described through transformations such as scaling, rotating, shifting, or reconfiguring of objects.

McHenry and Bajcsy indicate that scenery attributes are not a necessity for some 3D formats (2008, 3). They state that a format can abstain scenery by transforming parts of a model in the correct positions before saving or exporting a file. The camera and lighting sources can be ignored as well if it fits the purpose of the model and if the user has complete freedom over the visual aspects of the model. Although the absence of scenery can fit some purposes of 3D modeling, it is for animations and modern graphic visualizations of textures of the highest importance that the lighting and correct angle of view is portrayed to showcase the most realistic and accurate results (Rabin 2009, 720). The scenery features of object-based archaeological 3D models are very purpose dependent and can be considered important if the models are positioned in a visualization of the excavation or replication of the site. Light sources can provide new 'light' on details of archaeological objects that are not visible without illumination.

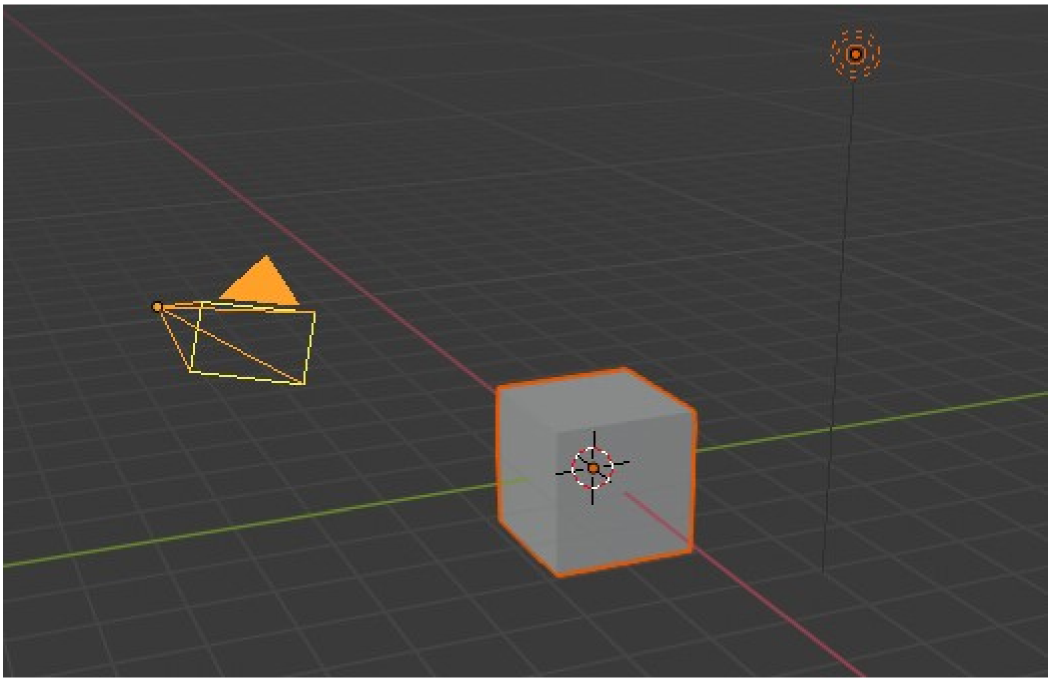


FIGURE 4.11: SCENERY ELEMENTS OF THE START SCREEN IN BLENDER. THE LEFT OBJECT REPRESENTS THE CAMERA AND ITS ANGLE. THE RIGHT CIRCULAR OBJECT REPRESENTS THE LIGHT SOURCE. THE CUBE IN THE CENTER REPRESENTS A 3D OBJECT (BLENDER 2.80 SCREENSHOT).

4.3.4 Animation

Animations of a 3D model means that parts of, or the complete 3D model, can be moved, rotated, or scaled. Two basic and essential elements for the illusion of movement within animations are *space* and *timing* (Rabin 2009, 729). In animations not only the object, but also the surroundings and the camera are interactive. Indicating that more details of a 3D model can be displayed if the purpose of the model requires it. Animations should not be confused with interactivity of a model. Animations are a set of instructions that manipulate the object or parts of the object in a pre-set pattern (www.guides.archaeologydataservice.ac.uk). Interactivity is generally focused on visualizing the complete 3D model by a user, which means it does not follow a pre-set pattern.

Animations generally require much more storage compared to static models and need high processing powers for rendering. Because of this, animations prefer approximate meshes over NURBS or CSG models. An alternative for exporting animations in 3D formats is to export the model as a digital video format. However, exporting animations for data preservation in this format relies on a completely different workflow and does not support the flexibility that 3D models generally have (www.guides.archaeologydataservice.ac.uk). Therefore, digital video formats are not considered in this thesis.

4.4 3D File formats

The primary purpose of 3D file formats is to store information about 3D models. The storage can be described in a human readable (ASCII) form or in computer code (binary). They both describe the same 3D properties, but ASCII tends to require more storage because the human readable letters require binary strings to describe each letter. The stored information in 3D file formats describe one or more of the 3D modeling features described in chapter 4.2.

3D file formats also have other specifications. Sometimes, 3D file formats are specifically made for a 3D modeling application. The BLEND format, only usable in Blender, is a perfect example of a format for specific 3D software, which are called *proprietary* file formats. These 3D formats are sometimes highly specialized and might be used very frequently or hardly ever, depending on the use of the application. On the other hand, there are also formats that are operative in multiple 3D modeling applications and intend to solve interoperability of 3D formats. An example is the OBJ format, that can be imported in Blender, 3DS MAX, Cinema4D and many other 3D modeling software. Formats like OBJ are called *data exchange*, *neutral* or *open-source* file formats.

Besides the above-mentioned specifications, each format has specific properties that influence how the 3D data and modeling features are stored as well. The next part of the chapter will shortly delve into 3D file formats that are recommended in EASY. Table 4.1 displays which 3D modeling features each of the six 3D file formats of EASY are able store. DANS currently recommends preferred and non-preferred file formats in EASY. Preferred formats in EASY offer the best long-term guarantee for usability, sustainability, and accessibility (https://dans.knaw.nl). Non-preferred formats do not imply not-accepted formats. It rather implies that the file formats are moderately to reasonably accessible, robust and usable in the long-term.

Besides these preferred and non-preferred 3D formats in EASY, many other 3D formats exist in the 3D computer graphic scene (http://3doc.i3dconverter.com; McHenry and Bajcsy 2008, 3-7). However, other 3D formats are not able to store approximate meshes, do not offer interoperability, are open-source or provide potential for long-term usability and preservation.

		OBJ	PLY	X3D	DAE	FBX	BLEND
Geometry	Approximate mesh						
	Precise mesh						
	CSG						
Appearance	Color						
	Material						
	Texture						
Scene	Camera						
	Lights						
	Relative Positioning						
Animation	Animation						

TABLE 4.1: 3D MODELING FEATURES OF THE PREFERRED AND NON-PREFERRED FORMATS OF EASY. BASED ON MCHENRY AND BAJCSY (2008, 18) AND WWW.ALL3DP.COM.

4.4.1 The OBJ format

The OBJ file format is first used in 1990 and is a neutral ASCII-based format generally referred to as ‘Wavefront OBJ’ (https://www.loc.gov). The format is still hugely popular and is used in many different 3D modeling applications, but mainly 3D graphics and 3D printing. Table 4.1 demonstrates that OBJ files can store approximate meshes, precise meshes and all the appearance features. However, no scenery information or animation can be stored. The textures of OBJ files are stored as images in a separate JPG file that must be linked to the OBJ file using a *Material*

Template Library (MTL) file. Thus, a total of 3 files (OBJ, JPG and MTL) are required to correctly open one textured 3D model of the OBJ format.

4.4.2 The PLY format

The PLY format originates from the Stanford Computer Graphics Laboratory and is first used in the 1990s as well. PLY is used to generates models from triangulation (approximate meshes) and is therefore also known as the Stanford Triangle Format. The main reason for its existence is to be simple and easily implementable, but also to be general enough to use for a large variety of models (Turk 1994, 1). A PLY file can only describe a singular 3D object, regardless of the size. The textures in PLY files can sometimes get lost, depending on the PLY support in the used 3D modeling application. The internal structure of PLY files starts with the allocation of the first three characters to describe 'PLY', visible in figure 4.12. This is a bootstrap, indicating a self-starting process. Without going in too much detail, bootstrap assures that software can recognize the uncompressed file structure based on these first characters. After the bootstrap the PDI is indicated in the header.

JA_3Dmodel_BRB_2_temp_100000f_PLY.ply

```
ply
format ascii 1.0
comment Created by Blender 2.80 (sub 75) - www.blender.org, source file: 'JA_3Dmodel_BI
element vertex 51845
property float x
property float y
property float z
property float nx
property float ny
property float nz
property float s
property float t
element face 99988
property list uchar uint vertex_indices
end_header
-0.103254 -0.035624 0.020639 -0.199652 -0.082827 -0.976348 0.762488 0.472840
-0.103596 -0.035468 0.020136 0.623218 -0.647511 0.438520 0.761821 0.472261
-0.103232 -0.035514 0.020380 -0.532579 -0.841548 -0.090060 0.762177 0.471610
-0.103602 -0.035752 0.020173 0.803552 0.489334 0.338847 0.762077 0.472667
-0.103305 -0.035890 0.020304 -0.112003 0.992523 -0.048006 0.761947 0.473277
-0.103136 -0.035712 0.020173 -0.809839 0.123722 0.573412 0.761121 0.469608
-0.038276 -0.082250 0.022338 0.693960 -0.239601 -0.678945 0.879020 0.618835
-0.038272 -0.082287 0.022041 0.643727 0.017090 0.765038 0.879235 0.618515
-0.038006 -0.082274 0.022081 -0.762139 -0.123783 0.635426 0.879660 0.618807
-0.038021 -0.082348 0.022349 -0.615162 0.330149 -0.715903 0.879385 0.619156
```

length : 5.826.218 lines : Ln : 1 Col : 1 Sel : 0 | 0 Unix (LF) UTF-8 INS

FIGURE 4.12: INTERNAL STRUCTURE OF A PLY FILE. THE FIRST THREE CHARACTERS ARE ALWAYS RESERVED TO DESCRIBE 'PLY'.

4.4.3 The X3D format

The X3D format is a succession of the VRML 3D file format and is released in 2001. X3D files inherit an Extensible Markup Language (XML), which is a structure defined by rules that is both human- and machine-readable (McHenry and Bajcsy 2008, 15). The X3D format is also neutral and is, unlike OBJ and PLY formats, able to store information of scenery and animations. The main purpose of X3D files is to support a virtual environment but can also be allocated in other 3D contexts.

4.4.4 The DAE format

The DAE format is an extension of COLLADA since 2004 and is mainly used in the videogame and film industry (https://all3dp.com). Although COLLADA is the official name of the file format, DAE is used throughout this thesis for import and export convenience. DAE incorporates an XML structure and requires the addition of a separate JPG file for texture mapping. The structure of DAE can link the texture to the mesh itself. Two files (DAE and JPG) are therefore required to correctly open one DAE file.

4.4.5 The FBX format

The Filmbox (FBX) is a proprietary format owned by Autodesk, a 3D software company, and is developed in 1996. FBX is mostly used for animations and is, just like DAE, used in the videogame and film industry. FBX is used as an interchange format for all 3D software of Autodesk[4], but other 3D software can use it as well. FBX files have the option to store textures for texture mapping within the FBX file itself, indicating that only the FBX file is required to view a 3D mesh. Each FBX file starts with 'Kaydara FBX Binary' as bootstrap.

4.4.6 The BLEND format

The BLEND format is a proprietary 3D file format made and produced by Blender. Although Blender is FOSS, they have released very little information on specifications of the 3D format. The BLEND format is not like other interchange formats, because it dumps the internal data structure from memory to disk. Although BLEND is highly specific to Blender, it is backward compatible with

[4] Autodesk owns many 3D software, such as AutoCAD, Fusion 360, Maya and 3DS Max. These are not FOSS and therefore not considered in this thesis.

previous versions of Blender. Each BLEND file start with allocating the first 12 characters for bootstrap purposes and contains information about the version and hardware structure as well.

In summary, this chapter addressed both the user and technical aspects of 3D modeling. The user aspects are represented by the fit for purpose for archaeological research and museums. The technical aspects attended commonly used 3D data acquisition techniques, 3D modeling features and 3D file formats. It became evident that 3D file formats are influenced by its human readable and or machine-readable structures, 3D modeling features, bootstraps, and connections to textures. 3D file formats can also be open or proprietary and heavily depend on 3D modeling applications.

Overall, the main limitations and preferences of the user and technical aspects are clarified and discussed. The next chapter implements these aspects, in combination with the digital archive perspective, in a workflow for the generation of a 3D dataset.

Chapter 5: Workflow

In this chapter the software, hardware, the workflow, and the implementation of the workflow are addressed. First, specifications of the software and hardware are presented. Next, the workflow of the generation of object-based archaeological 3D datasets is addressed. The three earlier defined aspects of the user, digital archive and technical requirements, influence the generation process and assemble the essential elements that are required to develop 3D datasets. Lastly, the implementation of the workflow is presented. The implementation is addressed on both a conceptual level and practical level.

5.1 Software and hardware specifications

5.1.1 Software specifications

Blender v2.80, released in 2019, was the main application used for the workflow of this thesis. Blender is an open-source software for high quality 3D modeling and animations. More detailed specifications of Blender can be found in appendix D. Blender has extensive options for import and export of 3D models as well as a wide variety of 3D file formats. For the workflow, the following add-ons were used in Blender: import-export: FBX format (V4.14.14), Import-export: Stanford PLY format (V1.0.0), Import-export: Wavefront OBJ format (V3.5.9), Import-export: Web3d X3D/VRML2 format (V2.2.1). These add-ons are obtainable within the official add-ons tab in Blender. Blender 2.80 is not the most current version of the application. However, the newest version 2.82, does not provide any updates that deviate the results of this thesis.

Notepad++ v7.8.8 (64 bit) is a FOSS text editor similar to Microsoft Notepad. Notepad++ is licensed under the GNU general public license and is used in the workflow for adjusting information that links the textures to the 3D model. The complete source code of Notepad++ is available on https://github.com.

DirLister v2 Beta 4 is a free obtainable and open-source software used for the creation of a metadata list for the file level metadata. The latest version is available via https://github.com and is licensed under the Mozilla Public License 2.0[5]. Dirlister generates lists of selected files and offers an array of information columns. The applications support a wide array of exporting, including HTML, CSV, and TXT.

[5]The Mozilla Public License 2.0 is a license belonging to the Open Source Initiative.

LibreOffice Calc V6.4.4.2 (64bit) is a spreadsheet FOSS alternative to Microsoft Excel. LibreOffice is similar to Microsoft Office, but FOSS and is available for the operating systems Windows, Mac OS X and Linux. The complete LibreOffice is freely obtainable under the Mozilla Public License 2.0. LibreOffice Calc is very extensive and is capable of exporting files to a wide variety of spreadsheet formats.

5.1.2 Hardware specifications

A personal laptop computer owned by the author of this thesis was used for developing and applying the workflow and for completing the thesis. The laptop was a Lenovo V110-15ISK with an Intel Core i5-6200U CPU and 8 GB RAM, running Windows 10 Home with a 64-bit Operating System. This system is considered mid-level.

5.2 Workflow of generating preservation acceptable 3D datasets

The workflow consists of four parts and will follow the structure of figure 5.1. Almost all parts are actions performed after a researcher is finished with his research and during the pre-ingest phase, meaning before depositing. Data within the dataset can still be altered before depositing, while during depositing, only information about the dataset can be given.

The first part addresses ethical guidelines. Ethical guidelines are important to address before altering any data in the dataset, because the dataset must comply to certain requirements before it is deposited on an academically supported digital archive. Ethical guidelines are not provided to change the outcome of the data, but function as a verification to conform the dataset to legal and institutional aspects. The guidelines in the workflow are considerations from the Faculty of Archaeology of Leiden University and DANS.

The second part establishes the file structuring and file naming of the dataset. Naming and structuring procedures are considered before and during the adaption of data, because adjusting 3D data after it is exported can contribute to many potential transferring and naming mistakes. The archival system and preservation strategy of EASY require filenames to adhere to certain standards, which facilitates future users for finding the data within the dataset as well.

The third part consists of modifying the 3D data within the dataset to conform to the requirements of the user and the digital archive, while also considering what is technically possible with 3D file formats. This means that the 3D data is stored in 3D file formats to be preserved with good longevity, while also incorporating 3D features and considerations to fit the purpose. This part has many little steps that need to be performed.

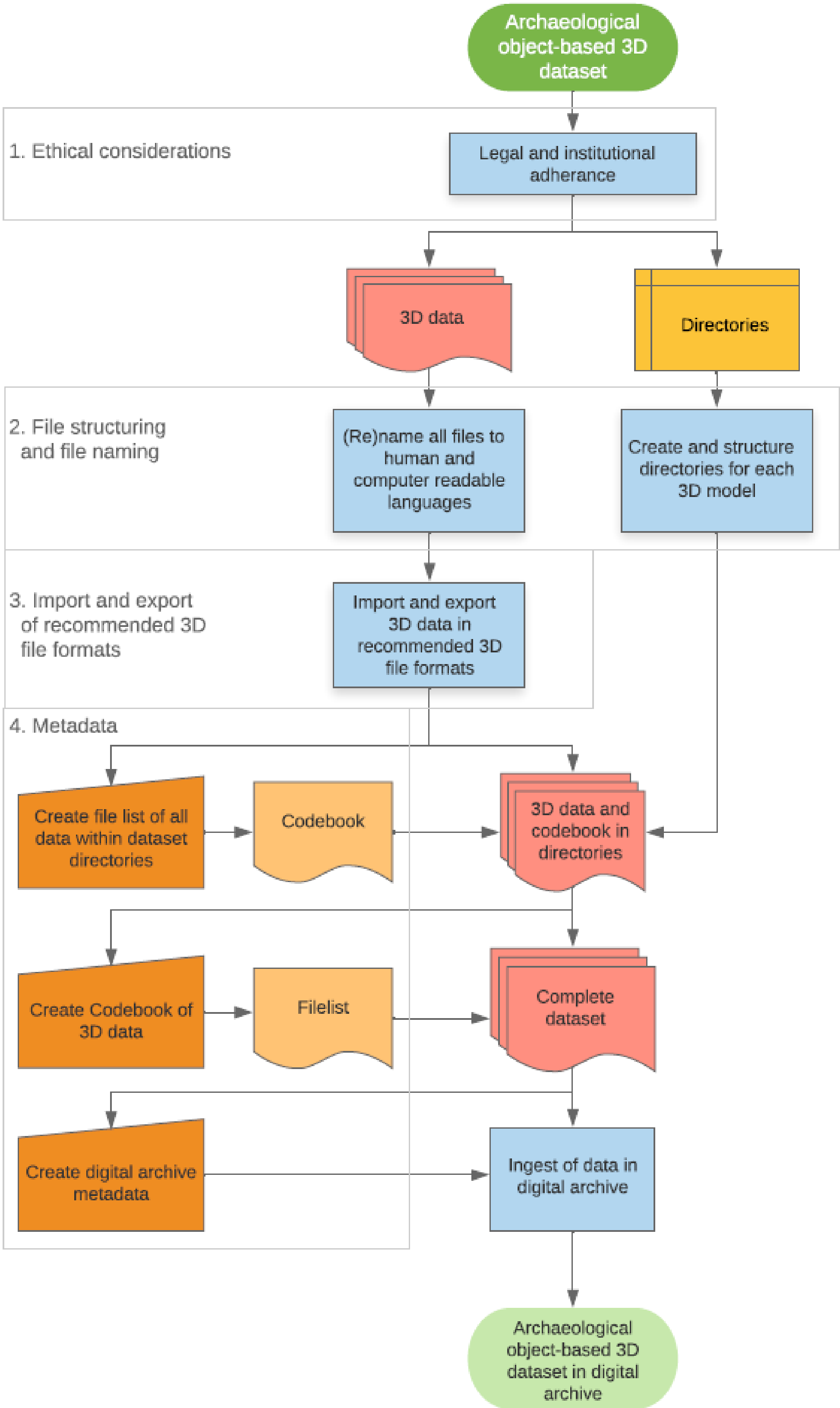


FIGURE 5.1: THE WORKFLOW OF A 3D DATASET AND THE PROCEDURES THAT MUST BE PERFORMED TO ADHERE TO THE REQUESTS OF THE DIGITAL ARCHIVE, LEGAL INSTITUTIONAL AND USER (MADE BY THE AUTHOR).

Therefore, a detailed in-depth workflow of part three is displayed in figure 5.4 of chapter 5.3.2. The 3D models will be stored in its original size and a decimated size. Part three consists of the archaeological object-based 3D models described in chapter 2.2, which are imported and exported to the six formats that are recommended in the EASY, described in chapter 4.4. The import and export procedure are performed in Blender 2.80 and uses the 3D file formats of BLEND, OBJ, DAE, FBX, PLY, X3D.

The fourth and last part of the workflow addresses three levels of metadata, described in chapter 3.2. The codebook and metadata file list are stored within the dataset and guide users of the dataset through the data. The codebook introduces the topic of the dataset and used abbreviations and the file list provides information on each data file. These two metadata levels are the last additions to the dataset, meaning that after these steps, the dataset is ready to be deposited in EASY. During deposition, the last metadata level is added in EASY. EASY metadata consists of information of the general project, but also functions to attract and inform potential future users about the dataset.

5.3 Implementation of the workflow

5.3.1 Ethical considerations

The first step that was to reflect on ethical perspectives and decisions of the research. The ethical guidelines are derived from GDPR (General Data Protection Regulation) of the European Union, with the specifications of The Faculty of Archaeology (FoA) of Leiden University (LU). The regulations DANS requests before depositing are also considered. The FoA of LU state that the responsibility of ethically acceptable research for undergraduate and graduate students (BA and MA/MSc) lies with the supervisor or instructor (www.organisatiegids.universiteitleiden.nl). After graduation, when a researcher is advanced in his/her research track, it is their own responsibility. Ethical considerations are to be performed within all the layers of archaeological research, but especially when dealing with sensitive, private, or personal information. The FoA of LU has created a checklist specifically for archaeological research that is to be followed to determine if research is considered ethically acceptable, or if the ethical commission is to be advised. DANS obligates similar as well as other regulations, because DANS is more concerned about personal data and making this data anonymous (www.dans.knaw.nl). The complete list of questions is visible in appendix E. The questions that are specifically relevant for archaeological 3D data are:

- Has similar previous research been denied by the Ethics Committee?
- Are you aware of similar research conducted by other researchers that has been cancelled by the Ethics Committee?

- Does the research involve working with people or human remains?
- Does the research involve collecting, processing, or analyzing personal data of (living or dead) people?
- Are personal data or samples exported to or imported from non-EU member states?

And for DANS:

- Is data acquired with informed consent?
- Is the personal data anonymized?
- If data contains personal data and it cannot be anonymized for research purposes, is a processing agreement with DANS conducted?

Typical ethical issues for archaeology are research of (3D models of) bones of people or 3D recreation of vessels that have much cultural value for local tribes (Decker and Ford 2017, 190). Another typical example is the collection survey obtained personal information of local people near a site.

Although very specific institutional and digital archive specifications are proposed, specifications of other institutions generally do not differ much to these considerations. Every institution and digital archive that deals with 3D data has similar or analogous requests.

5.3.2 File structuring and file naming

The second step involves generating a new directory called "DANS_data" in the location where data of the research is stored. This directory is the location where all the generated data for DANS will (temporarily) be stored before ingest. The underscore between DANS and data is important because spacebars between words are sometimes not recognized in computer scripts, resulting in errors and potential data loss. Therefore, every filename and directory name in the dataset is described with underscores.

In the newly created DANS_Data directory, two files and one directory are generated at the end of the procedure of making the dataset archive ready. The new directory addresses specific project related information, that starts as broad as possible and gets more specific with each layer of directories. For example, the first directory layer is named after the project site, followed by the layer of directories for ceramic types. This directory layer of ceramic types is followed up by a directory layer of 3D models which is subdivided in the different file sizes for each object. The structure of the directory structure of DANS_Data is displayed in figure 5.2.

Having many directory layers helps users of datasets to read the data in organized and easily understandable portions. Datasets often comprise a lot of data and it can be difficult to find the object or information you are looking for without an organized file structure. However, having many directories can accidentally lead to moving data into wrong directories. Therefore, data sorting checkups should be performed before depositing a dataset in a digital archive.

The file and directory naming are advisable to follow a structure that is coherent across all and within all directories. It should be clearly evident what a directory contains, based on the name. Therefore, each directory should start with the naming of the aspect type, followed by the actual aspect. For example, 'Ceramics-BRB' or 'Site-JA'. Abbreviations can be used, because in the one of the first layers of directories, the project codebook is located that explains, among other things, each abbreviation. The file list is also located in the same layer as the codebook because it should be easily accessible as well and help users of the dataset as a guide through the dataset. There are limitations that cannot be used for filenames, such as punctuation marks. If filenames contain any punctuation marks, the files will generate naming errors and file errors during the ingest procedure in the infrastructure of the digital archive, resulting in data loss.

file names of the exported 3D data are structured according to the file structure, followed by the type of format used. The dataset of this thesis has the file naming of figure 5.3. An advantage of this file naming technique is that 3D files are easily retraceable and findable in the file structure. Another advantage is the file format in which the model is exported is still visible when the filetype is (temporarily) changed. Filetypes are often changed to be viewed in a text editor such as Notepad++ to examine the internal binary or ASCII structure a 3D model. Having the 3D file format at the end is a reminder of what the original file format is.

JA_3Dmodel_BRB_2_temp_100000f_BLEND.blend
JA_3Dmodel_BRB_2_temp_100000f_DAE.dae
JA_3Dmodel_BRB_2_temp_100000f_FBX.fbx
JA_3Dmodel_BRB_2_temp_100000f_OBJ.mtl
JA_3Dmodel_BRB_2_temp_100000f_OBJ.obj
JA_3Dmodel_BRB_2_temp_100000f_PLY.ply
JA_3Dmodel_BRB_2_temp_100000f_X3D.x3d
JA_3Dmodel_BRB_2_temp_JPG.jpg

Figure 5.2: Filenames of the directory of JA_3Dmodel_BRB_2_temp_100000f. The filenames indicate that this folder contains files that 3D models of the Jebel Aruda excavation, from the pottery type BRB with the name '2 temp'. The file is decimated, based on the number of f (faces).

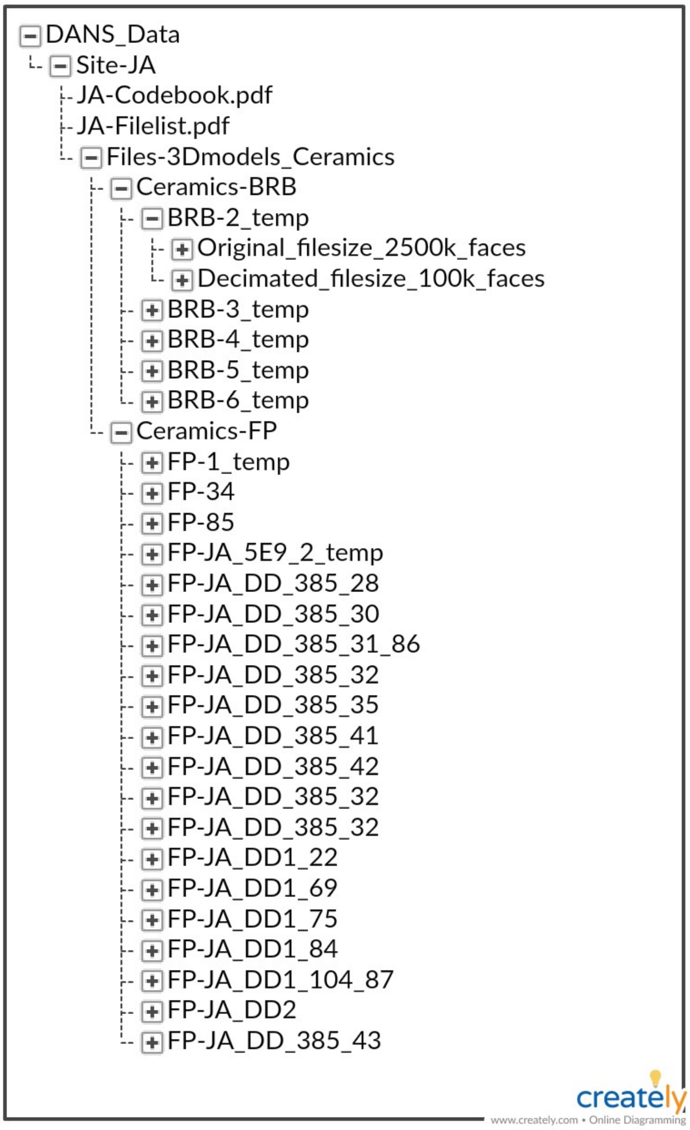


FIGURE 5.3: FILE STRUCTURE OF THE DATASET. MADE USING CREATELY (HTTPS://WWW.CREATELY.COM).

5.3.3 Import and export of recommended 3D file formats

Next, the 3D file formats of the 3D models are addressed. For the implementation, the directory of flowerpot 'JA_DD_385_28' is used. Only one specific object is chosen, because the procedure for each 3D model is practically the same. The only differences are the distinct name of each object in each directory, and the fact that a few models have different original sizes (2,000,000 faces or 1,000,000 faces instead of the standard 2,500,000 faces). The import and export step is arguably the most important part of the workflow because this step has the most influence on the longevity of the 3D models. The procedure of part three is visualized in the workflow of figure 5.4.

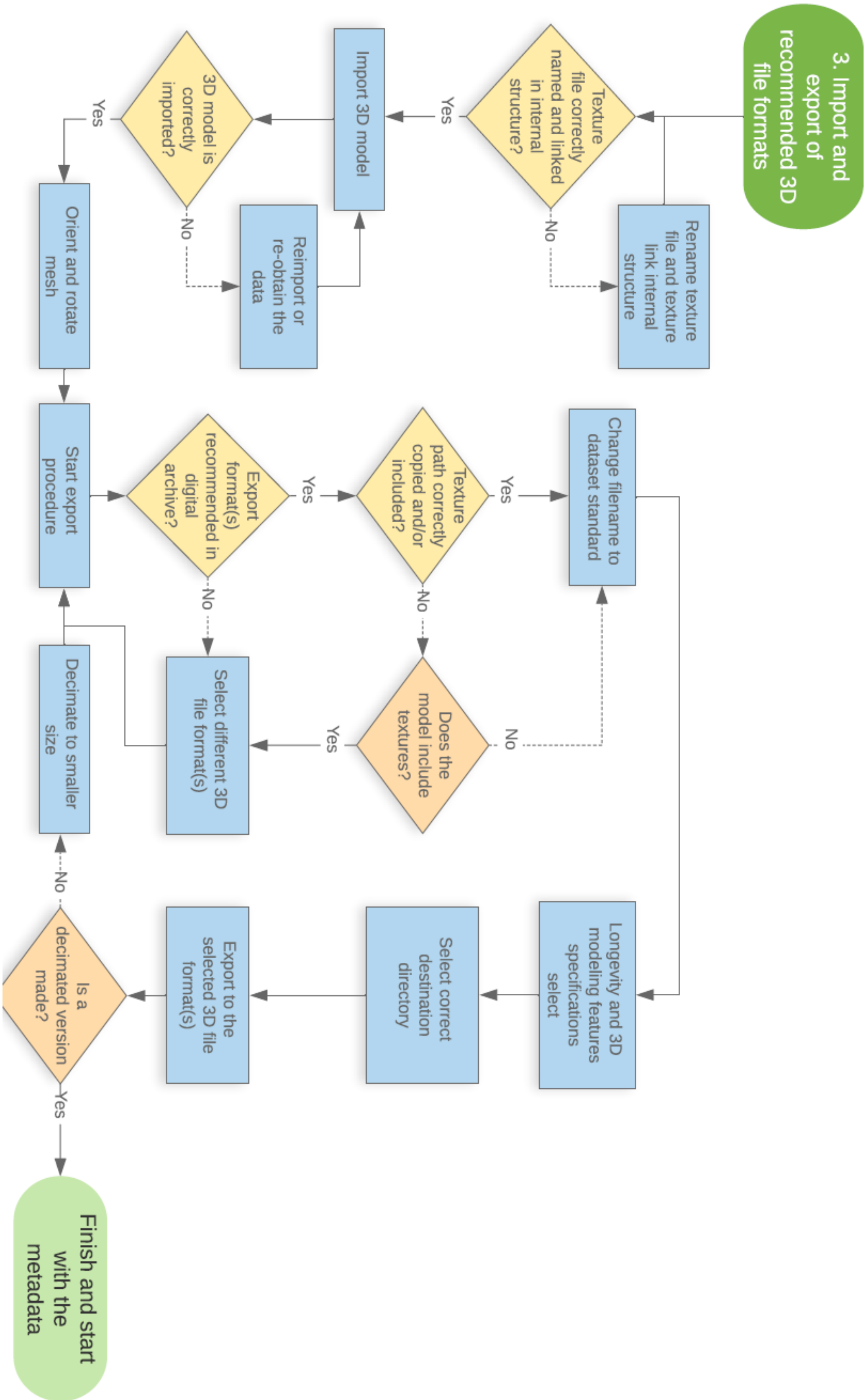


FIGURE 5.4: PART THREE: SPECIFIC IMPORT AND EXPORT WORKFLOW OF RECOMMENDED 3D FILE FORMATS (MADE BY THE AUTHOR).

First, the structure of the original file of 2,500,000 faces is inspected, to check if the texture is applied correctly. This action has to be performed before importing or exporting, because after importing small errors now can cause large mistakes in the outcome. The main reason for inspecting the internal structure of the file is to conform the texture filename to the described file name structure of part two. The name of textures cannot be changed during or after import, thus it should be changed before import.

The JPG file name is changed to *"JA_3Dmodel_FP_JA_DD_385_28_JPG.jpg"*. The MTL file is opened, as is visible in figure 5.5. The last line displays *map_Kd* followed by the original name of the JPG file. The original JPG file name must be changed to describe the newly generated name for the JPG file. This name change guarantees the correct appliance of the texture on the 3D model and the correct file naming of the texture file.

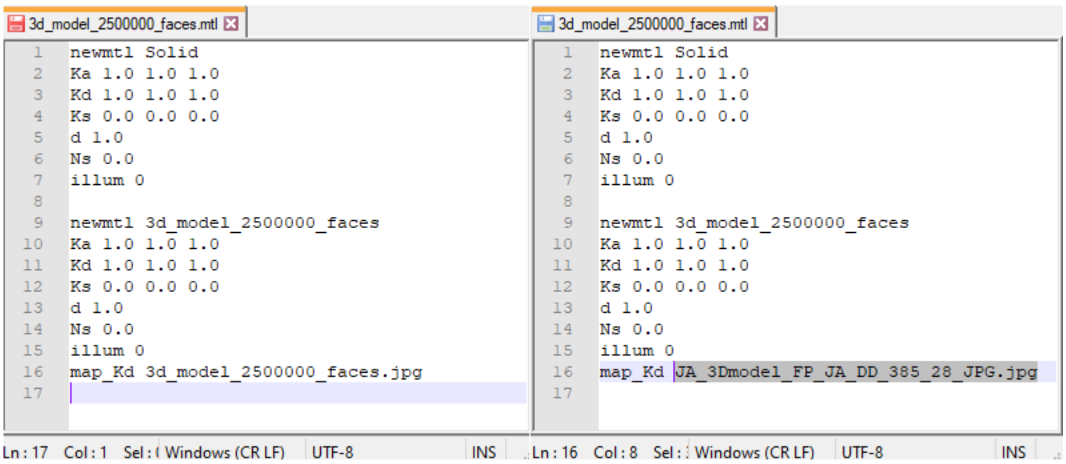


FIGURE 5. 4: LEFT: THE ORIGINAL STRUCTURE OF THE MTL FILE OF JA_DD_385_28 IN NOTEPAD++. RIGHT: THE HIGHLIGHTED AND CHANGED NAME OF THE TEXTURE FILE IN LINE 16.

After the inspection of the internal structure of the MTL file is complete, the import procedure of the 3D model is started.

Blender is opened with a clean 3D viewport. The 3D object is imported using its option in the import screen. The specifications of the imported file, in this case an OBJ file, are displayed in figure 5.6. The specifications ensure, using *image search*, that the correct texture is applied, even if it is accidentally located in a subdirectory. The 'Keep Vert order' ensures that the internal structure of the file remains the same, as well as maintaining unused vertices. Maintaining an identical internal structure is a major target of data import and export, as it will result in the least data deviation.

The clamp size is only required when multiple 3D objects are imported. 3D Files can vary a lot in size and by using the clamp size, all the objects are imported to a fixed size.

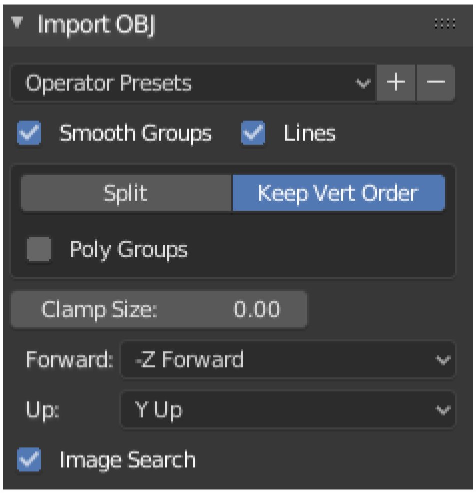


FIGURE 5.5: THE IMPORT SPECIFICATION SCREEN OF OBJ FILES IN BLENDER 2.8.

After import, the 3D object should be moved and rotated to represent the model from its best angle. This action is performed to let users view the object with ease and with the right orientation. Which is especially useful for future users who are less familiar with 3D objects and 3D applications or users unable to locate and orient the 3D object themselves.

3D models generated in other 3D applications often have vertex coordinates that do not always overlap with (local) vertex coordinates of Blender. The 'Forward' and 'Up' axis in figure 5.6 changes the orientation of the object, but it can also happen that the origin of an object is not at the center of the geometry. What this means is that the location that Blender thinks is the center of the object is not actually the center of the object. This can easily be changed by selecting: Set origin -> Geometry to Origin. The object is now at the center of the 3D viewport and can, if it is necessary, be rotated to stand upright. During this procedure, it is also important is to inspect if the texture is applied correctly during the import procedure. This part is the last component before the objects are exported to extend longevity and if the texture of a 3D model is not present during that phase, the model will not be exported correctly. The final orientation, rotation and texture should look like the bottom figure of figure 5.7.

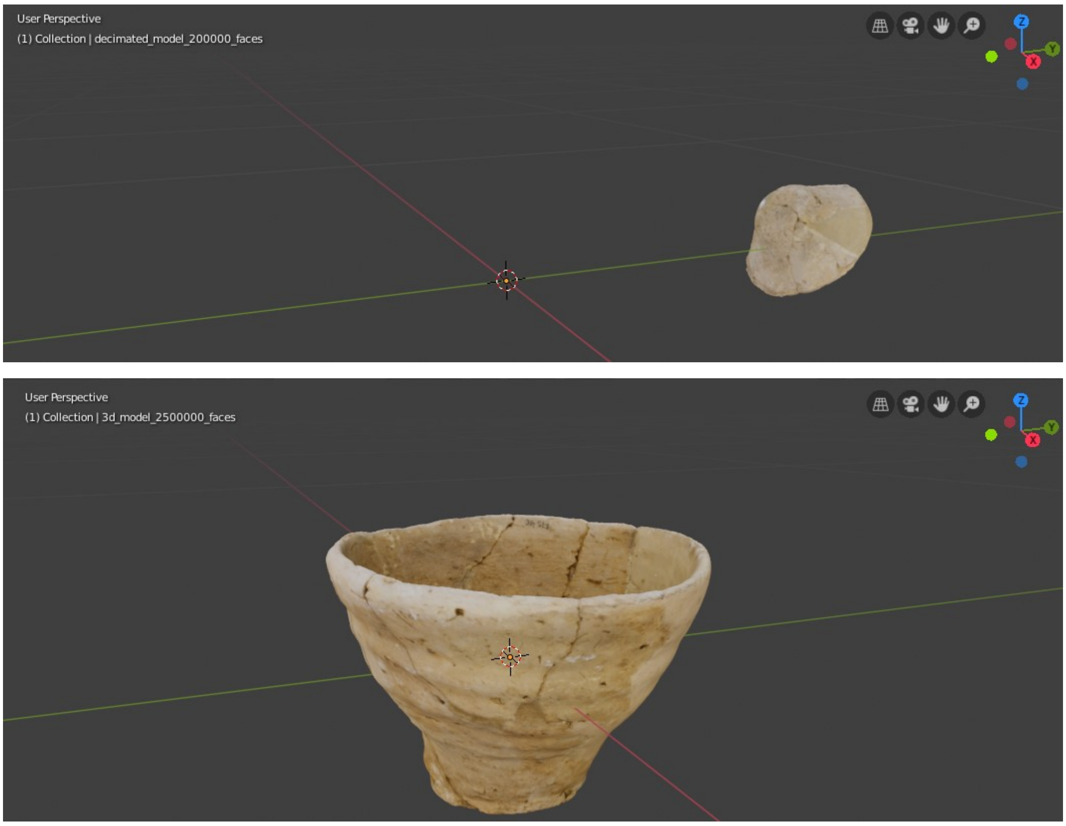


FIGURE 5.6: THE TOP FIGURE DISPLAYS A POSSIBLE LOCATION AND ORIENTATION OF A 3D MODEL IN BLENDER 2.80. THE BOTTOM FIGURE DISPLAYS THE PREFERRED, CENTERED AND REPRESENTABLE ORIENTATION AND ROTATION OF THE 3D MODEL IN BLENDER 2.80.

The 3D model is now representable for future users and is ready to be exported. The file is exported to one or multiple preservation friendly formats. In this case the six 3D file formats of:

- BLEND
- DAE
- FBX
- PLY
- OBJ
- X3D

Essential elements for the export procedure of each 3D format is shortly discussed. After the export procedure, the decimation process is addressed. Decimation is useful for many purposes described in chapter 4.1, such as previewing and comparing 3D models.

For each of these 3D file formats it is essential to export the format with the least alterations and in its most uncomplicated form. A couple of these formats can store the image texture within the file format. This is preferred over having a file format and an additional texture file. Incorporating the texture allows ease of use for beginning 3D model users and avoids accidental loss of data during transfer.

File formats should also not be compressed or stored in other shapes than its original geometry type (which are described in chapter 4.3), compression and different geometries alter the internal structure excessively and lead to potential data loss during export.

The BLEND format requires that all external data is packed in the BLEND file, which is an additional feature enabled using 'File->External Data->Automatically Pack Into .blend'. whilst saving the 3D model in the BLEND FORMAT, the filename should correspond to the structure given in part two and should look similar to: *JA_3Dmodel_FP_JA_DD_385_28_2500000f_BLEND.blend*. The export specifications of BLEND are simple, but that is because of its unique storing mechanic, described in chapter 4.4.6. The 'Remap Relative' function can be selected if the BLEND file will be stored in a different directory than the original (OBJ) file, see figure 5.8.

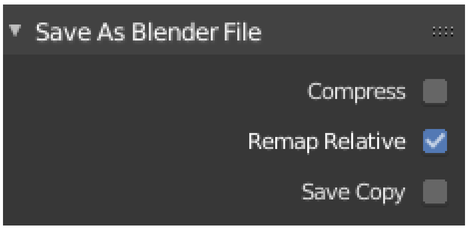


FIGURE 5.7: THE SAVE/EXPORT SPECIFICATIONS SCREEN OF BLEND FILES IN BLENDER 2.8

The DAE format is to be exported in a similar way as the BLEND file. However, DAE in Blender has different storage specifications. The DAE format is used in multiple 3D applications compared to the BLEND format, because DAE is an exchangeable file format. The export procedure must generate the most value for the exchange attribute of the format.
The object is exported using Collada (Default) (.dae) with the filename *JA_3Dmodel_FP_JA_DD_385_28_2500000f_DAE.dae*. The specifications of the export of DAE files are displayed in figure 5.9. The specifications in Blender provide the option the 'copy' the texture file with the export file. 'Copy' ensures that the DAE file and its required texture file are located near each other and guarantees the relative reference to the texture file. Relative references change when the files are located to another directory. The DAE file will refer to the newly generated image texture file and will keep doing that even when both files are transferred or moved. The opposite of relative referencing is absolute referencing, which means that the DAE file would still attempt to use the original image texture file. Absolute referencing is not preferable when files are transferred, because the reference of file path within the DAE file is not correct anymore.

The 'Triangulate' option under the geometry tab protects the triangulation of data and maintains the approximate geometry structure of the file.

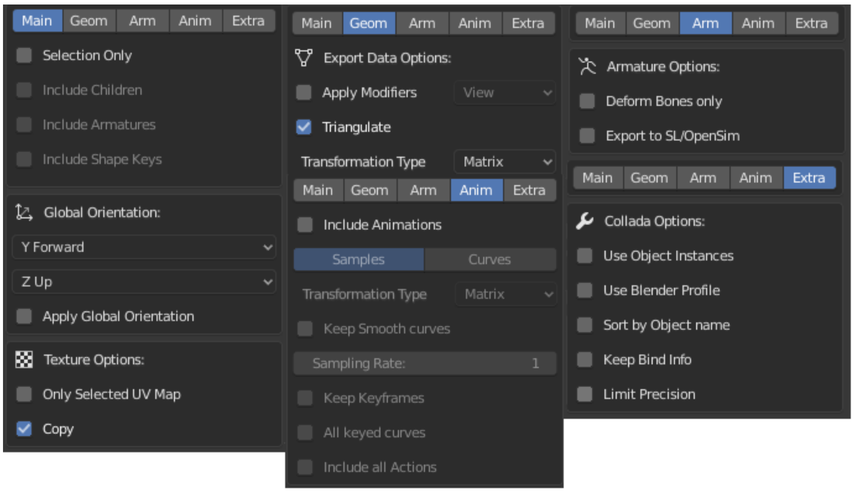


FIGURE 5.8: THE EXPORT SPECIFICATIONS SCREEN OF DAE FILES IN BLENDER 2.8.

The FBX file is exported in FBX (.fbx) with the filename *JA_3Dmodel_FP_JA_DD_385_28_2500000f_FBX.fbx*. FBX files can store the texture file within the file format, as described in chapter 4.4.5. This action is applicable in the specifications in the 'Main' tab. The red encircled symbol in figure 5.10 displays the option to embed the texture in the FBX file. The path mode must be 'Copy', or this function does not work. The other specifications for FBX files are unaltered.

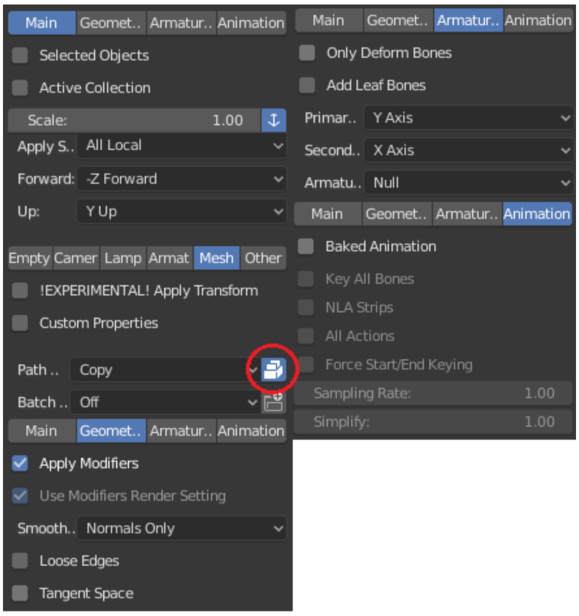


FIGURE 5.9: THE EXPORT SPECIFICATIONS SCREEN OF FBX FILES IN BLENDER 2.8. THE RED ENCIRCLED OPTION ENSURES THAT THE TEXTURE IS EMBEDDED WITHIN THE FBX FILE.

The export of the PLY format, Stanford (.ply), has the filename *JA_3Dmodel_FP_JA_DD_385_28_2500000f_PLY.ply*. PLY files have limited export specifications. Only the 'UVs' function can be useful if the model utilizes UV coordinates, addressed in chapter 4.3.2. The specifications of PLY files in Blender are visible in figure 5.11.

The export of OBJ files is via Wavefront (.obj) and the filename of the OBJ format of this file is *JA_3Dmodel_FP_JA_DD_385_28_2500000f_OBJ.obj*. The specifications of the OBJ export procedure have to be altered to assure that: the faces are stored as triangulate faces, the edges and UVs are included, the path mode to the texture file is copied and lastly, the materials file (MTL file) is generated. The use of the MTL file is an essential addition to every OBJ file with texture, as addressed in chapter 4.4.1. The specification screen of the OBJ export is visible in figure 5.12.

The file export in X3D extensible 3D (.x3d) has the filename *JA_3Dmodel_FP_JA_DD_385_28_2500000f_X3D.x3d*. The specifications of X3D formats must include triangulation for the approximate geometry. The specifications of X3D export are visible in figure 5.13.

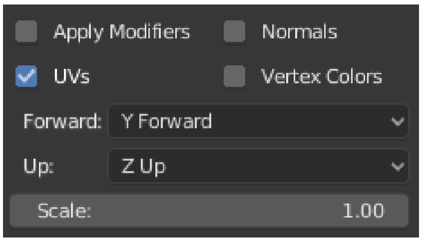


FIGURE 5.10: THE EXPORT SPECIFICATIONS SCREEN OF PLY FILES IN BLENDER 2.8.

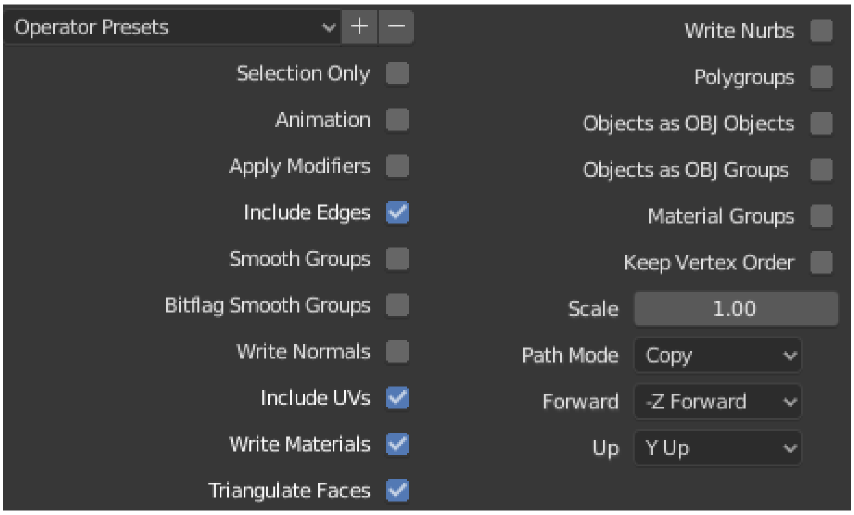


FIGURE 5.11: OBJ EXPORT SPECIFICATIONS IN BLENDER 2.8.

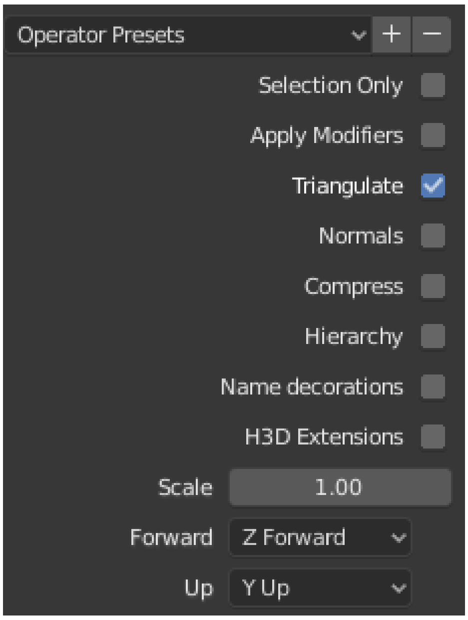


FIGURE 5.12: THE EXPORT SPECIFICATIONS SCREEN OF X3D FILES IN BLENDER 2.8.

After finishing all exports in original sizes, the decimation of files to 100,000 faces is started. Decimated files require less storage space and can be easily rotated and manipulated in 3D software. Decimated files should be stored as separate files in the *Decimated_filesize_100k_faces* directory but are for many archaeological purposes not enough to be stored as the only file representing a 3D object. It enhances the performance, accessibility and adaptability of models and are therefore useful additions to store. However, decimated files are far less accurate than the original models.

Decimation is performed by reducing the faces with minimal shape changes. This action is performed after exporting all the files in the original faces number. Once decimated, the file cannot be un-decimated. In Blender, the decimation modifier under properties is selected, whilst ensuring that the triangulation faces remain. The ratio is reduced to maintain only 100,000 (100K) faces for each 3D model. When the decimation process is completed, the export procedure can start again in a new directory with filenames including *100000f* instead of *2500000f*. The generated FP directories with the decimated files are displayed in figure 5.14.

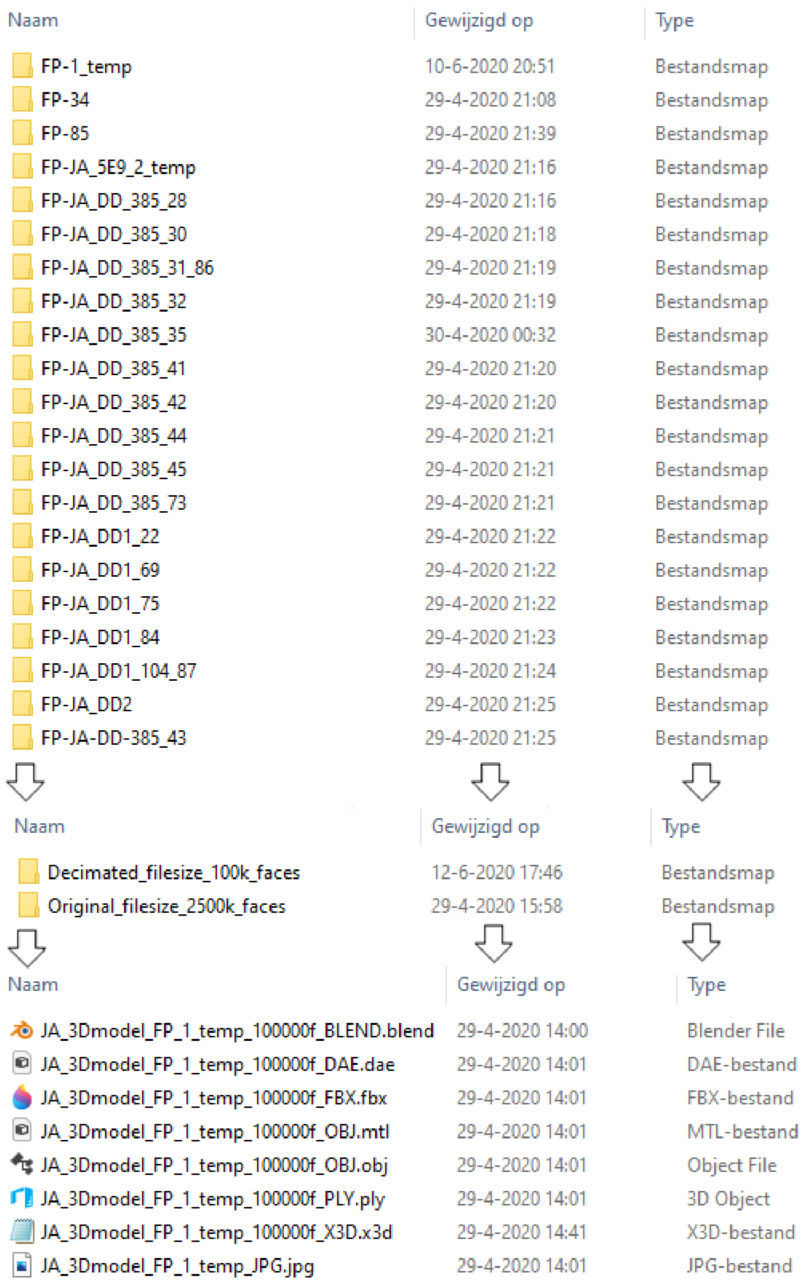


FIGURE 5.13: THE THREE LAYERS OF DIRECTORIES OF THE MODELS.

In summary, this part of the workflow explains how 3D models are exported in multiple formats. Each generated file has specifications for each format that are best suited for object-based archaeological 3D models and for extended longevity. The files are stored with their original number of faces and with a decimated number of faces. All the adjustments to the 3D models are now performed. Thus, part four describing metadata, the last process to finalize the dataset, can be initiated.

5.3.4 Codebook

The first part of the documentation of the dataset is the codebook, which is explained in chapter 3.2.3 and is the start of the fourth part of the workflow. Codebooks are created in datasets to describe used abbreviations, the background of the original data and the generation of 3D data. Although a codebook for each separate file is possible, only one codebook is used for this dataset. Each 3D file uses the same abbreviations and is generated under the same circumstances. It is important for the readability of the dataset that this, as well as the other metadata, are stored in longevity-friendly formats, such as PDF/A, TXT or CSV. The complete codebook of this dataset named *JA_Codebook.pdf* is visible in appendix F.

The codebook for the dataset with 3D models starts with a short instruction of which 3D files require which additional files to be opened properly. This is useful for OBJ and DAE formats that require additional files to view the texture. The short instruction is followed with a list of abbreviations and first addresses abbreviations on the project level, followed by the 3D object level. 3D files or any part of this dataset, do not require much text or abbreviations, thus the abbreviation list is short and concise.

After the abbreviations, specific project metadata is described. This described information partially overlaps with information depicted in the project-level metadata for EASY. The codebook can therefore be referred to when generating the project-level metadata. The question can be raised why it is necessary to store the information twice. The same information needs to be displayed, because a dataset can be downloaded and passed to different people using other sharing methods. In that case, the dataset is stripped from its connection to DANS and EASY, but still needs a source to the original data location. That is where the codebook is used for. A similar guiding mechanism is used in articles through the use DOIs for referencing to the original publisher, which is obtainable during the EASY metadata procedure that is described in chapter 5.3.6. A couple of elements of the codebook are derived from the CARARE metadata schema.

Project metadata consists of general questions that inform the reader:

- Who is involved in the making this dataset?
- When are the project dates?
- Where is the data generated?
- How can the data be accessed and under which license?
- Is the data linked to other data(sets)?

After the project metadata, the more content specific metadata is addressed that is, in this case, specified to archaeological objects information. If artefacts vary to much and cannot be described in these few questions, multiple codebooks are required, with each codebook based on each artefact or artefact type. However, because all artefacts of the case study can be categorized to two Uruk pottery types, it will suffice.

Typical object related questions about the archaeological object information are:

- What is the classification of the material?
- What are the physical characteristics of this specific material in this project?
- What are the number of artefacts?
- To which period can the artefacts be dated to?

The third part of this codebook described the technical aspects and addresses the description of the generation and alteration processes of the 3D data, also known as the paradata. This part also addresses who(m) generated the data and what storage possibilities have been created for it. In chapter 2.2.2 an accurate display of the technical part of the codebook is depicted in table 2.1. The complete overview of the codebook is given in appendix F.

5.3.5 File level metadata

The last file to be added to the dataset and the second metadata part is the metadata file list. Generating file lists can take a lot of time, but is facilitated by file list applications such as Dirlister V2, figure 5.15. File level metadata is explained in chapter 3.2.2 and are generally created for longevity reasons. The file list is also used as a check to control if there are punctuation marks within filenames, described in the file naming procedure of chapter 5.3.3.

Dirlister partially automates the generation of file lists. However, there is still some manual labor to perform after the automation. The DANS_Data directory is selected for the automation process and every file that is located within this directory will be formed into a list. That includes the codebook as well. As mentioned in chapter 3.2.2, not all columns are necessary in EASY, thus only the essential columns ‘name’, ‘File size’ and ‘Created/modified dates’ are selected.

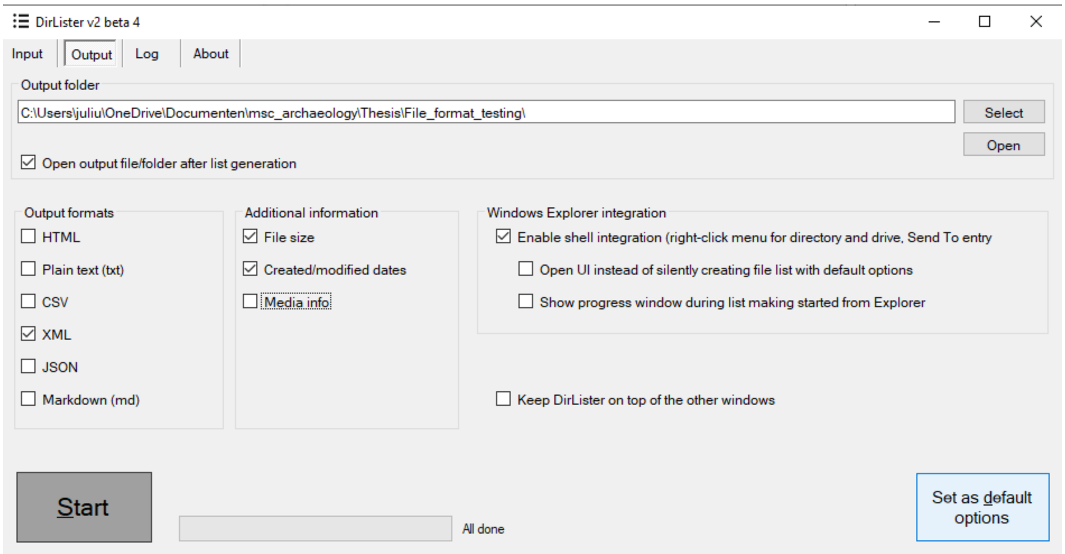
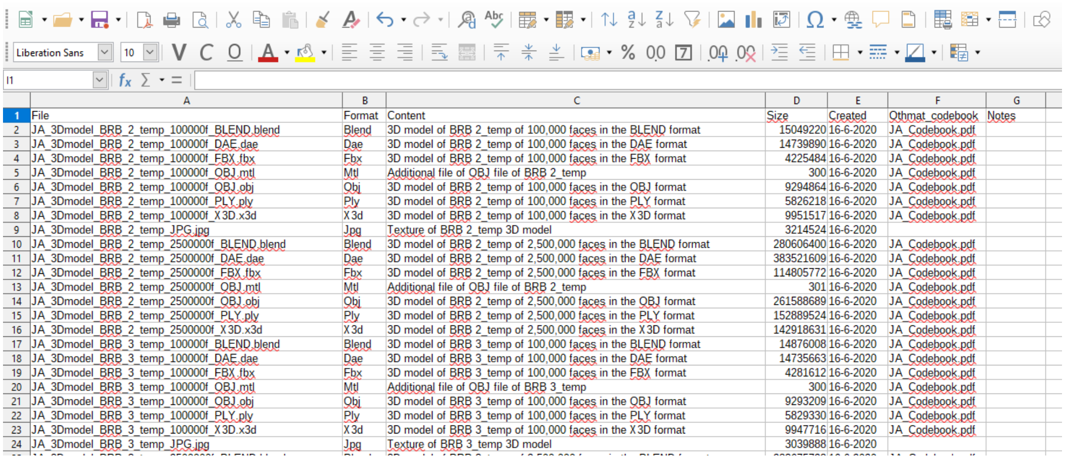


FIGURE 5.14: THE INTERFACE OF DIRLISTER V2, THIS FOSS IS USED TO GENERATE THE METADATA FILE LIST.

The file is saved with the name *JA-Metadata.csv* but is immediately imported in a spreadsheet application, for example, LibreOffice Calc, to add additional information. Four extra columns are added: 'Format', 'Content', 'Othmat_Codebook' and 'Notes'. The format is derived from the filenames of the 3D models and in the content column each file gets a short description of what each object is. All files that include references and abbreviations are referred to the codebook in the Othmat_Codebook column. The codebook is described in the previous subchapter.

Although not obligated by DANS, another column 'notes' was added. The notes column is very useful for adding additional information in the form of text if future users of the dataset need to know if data is incorrectly generated or adjusted. The resulted metadata list is visible in appendix G and a screenshot of the file list is visible in figure 5.16.



	A	B	C	D	E	F	G
1	File	Format	Content	Size	Created	Othmat_codebook	Notes
2	JA_3Dmodel_BRB_2_temp_100000f_BLEND.blend	Blend	3D model of BRB 2_temp of 100,000 faces in the BLEND format	15049220	16-6-2020	JA_Codebook.pdf	
3	JA_3Dmodel_BRB_2_temp_100000f_DAE.dae	Dae	3D model of BRB 2_temp of 100,000 faces in the DAE format	14739890	16-6-2020	JA_Codebook.pdf	
4	JA_3Dmodel_BRB_2_temp_100000f_FBX.fbx	Fbx	3D model of BRB 2_temp of 100,000 faces in the FBX format	4225484	16-6-2020	JA_Codebook.pdf	
5	JA_3Dmodel_BRB_2_temp_100000f_OBJ.mtl	Mtl	Additional file of OBJ file of BRB 2_temp	300	16-6-2020	JA_Codebook.pdf	
6	JA_3Dmodel_BRB_2_temp_100000f_OBJ.obj	Obj	3D model of BRB 2_temp of 100,000 faces in the OBJ format	9294864	16-6-2020	JA_Codebook.pdf	
7	JA_3Dmodel_BRB_2_temp_100000f_PLY.ply	Ply	3D model of BRB 2_temp of 100,000 faces in the PLY format	5826218	16-6-2020	JA_Codebook.pdf	
8	JA_3Dmodel_BRB_2_temp_100000f_X3D.x3d	X3d	3D model of BRB 2_temp of 100,000 faces in the X3D format	9951517	16-6-2020	JA_Codebook.pdf	
9	JA_3Dmodel_BRB_2_temp_JPG.jpg	Jpg	Texture of BRB 2_temp 3D model	3214524	16-6-2020		
10	JA_3Dmodel_BRB_2_temp_2500000f_BLEND.blend	Blend	3D model of BRB 2_temp of 2,500,000 faces in the BLEND format	280606400	16-6-2020	JA_Codebook.pdf	
11	JA_3Dmodel_BRB_2_temp_2500000f_DAE.dae	Dae	3D model of BRB 2_temp of 2,500,000 faces in the DAE format	383521609	16-6-2020	JA_Codebook.pdf	
12	JA_3Dmodel_BRB_2_temp_2500000f_FBX.fbx	Fbx	3D model of BRB 2_temp of 2,500,000 faces in the FBX format	114805772	16-6-2020	JA_Codebook.pdf	
13	JA_3Dmodel_BRB_2_temp_2500000f_OBJ.mtl	Mtl	Additional file of OBJ file of BRB 2_temp	301	16-6-2020	JA_Codebook.pdf	
14	JA_3Dmodel_BRB_2_temp_2500000f_OBJ.obj	Obj	3D model of BRB 2_temp of 2,500,000 faces in the OBJ format	261588689	16-6-2020	JA_Codebook.pdf	
15	JA_3Dmodel_BRB_2_temp_2500000f_PLY.ply	Ply	3D model of BRB 2_temp of 2,500,000 faces in the PLY format	152889524	16-6-2020	JA_Codebook.pdf	
16	JA_3Dmodel_BRB_2_temp_2500000f_X3D.x3d	X3d	3D model of BRB 2_temp of 2,500,000 faces in the X3D format	142918631	16-6-2020	JA_Codebook.pdf	
17	JA_3Dmodel_BRB_3_temp_100000f_BLEND.blend	Blend	3D model of BRB 3_temp of 100,000 faces in the BLEND format	14876008	16-6-2020	JA_Codebook.pdf	
18	JA_3Dmodel_BRB_3_temp_100000f_DAE.dae	Dae	3D model of BRB 3_temp of 100,000 faces in the DAE format	14735663	16-6-2020	JA_Codebook.pdf	
19	JA_3Dmodel_BRB_3_temp_100000f_FBX.fbx	Fbx	3D model of BRB 3_temp of 100,000 faces in the FBX format	4281612	16-6-2020	JA_Codebook.pdf	
20	JA_3Dmodel_BRB_3_temp_100000f_OBJ.mtl	Mtl	Additional file of OBJ file of BRB 3_temp	300	16-6-2020	JA_Codebook.pdf	
21	JA_3Dmodel_BRB_3_temp_100000f_OBJ.obj	Obj	3D model of BRB 3_temp of 100,000 faces in the OBJ format	9293209	16-6-2020	JA_Codebook.pdf	
22	JA_3Dmodel_BRB_3_temp_100000f_PLY.ply	Ply	3D model of BRB 3_temp of 100,000 faces in the PLY format	5829330	16-6-2020	JA_Codebook.pdf	
23	JA_3Dmodel_BRB_3_temp_100000f_X3D.x3d	X3d	3D model of BRB 3_temp of 100,000 faces in the X3D format	9947716	16-6-2020	JA_Codebook.pdf	
24	JA_3Dmodel_BRB_3_temp_JPG.jpg	Jpg	Texture of BRB 3_temp 3D model	3039888	16-6-2020		

FIGURE 5.15: SCREENSHOT OF THE FILE LIST IN LIBREOFFICE CALC.

5.3.6 EASY Metadata

The last metadata part for any dataset is the generation of metadata on the project level for EASY. EASY metadata is partially already described in the codebook of chapter 5.3.4; thus, much information is already obtainable. The generation of EASY metadata is the last action to perform, but is not performed within the dataset, but at the website of www.easy.dans.knaw.nl. This information can be stored in a separate file but must be filled in during the deposit procedure in EASY. This metadata is used by digital archive and is also visible for search engines using the OAI-PMH, descrbed in chapter 3.3.1. Therefore, the information should be as extensive and accurate as possible to make the dataset more findable and accessible.

In chapter 3.2.1 is explained that project level metadata in EASY is based on the DC metadata schema and that six elements of this scheme are obligatory in EASY: Title, Creator, Description, Date (created), Rights and Audience. Some of these obligatory fields do go more in-depth. For example, the creator element requires the addition of organization of the creator. Licensing rights involve open access or restricted access, as well as the addition of one of the creative common licenses (chapter 3.2.1) for open access, visible in figure 5.17. The information bars during the deposition procedure are visible in figure 5.18.

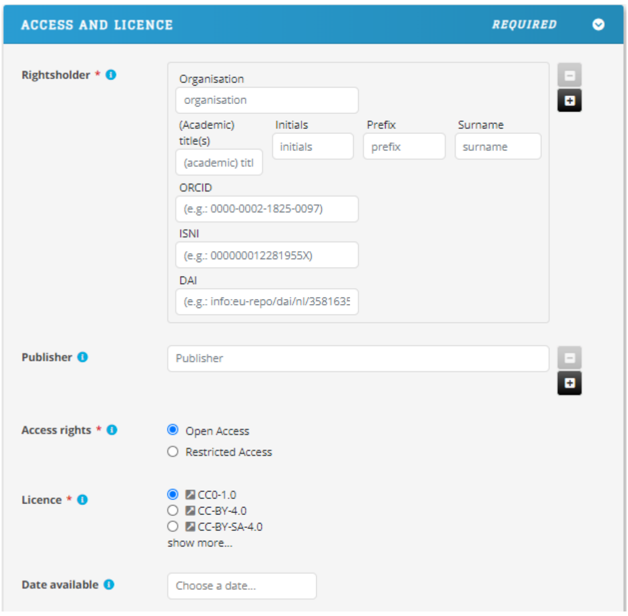


FIGURE 5.17: THE ACCESS AND LICENSE BAR IN THE DEPOSITION PROCEDURE IN EASY. BESIDES DESCRIBING THE OBLIGATORY RIGHTSHOLDER, ACCESS RIGHTS AND LICENSE, THE DATE AVAILABLE ENABLES THE OPTION TO EMBARGO THE DATASET.

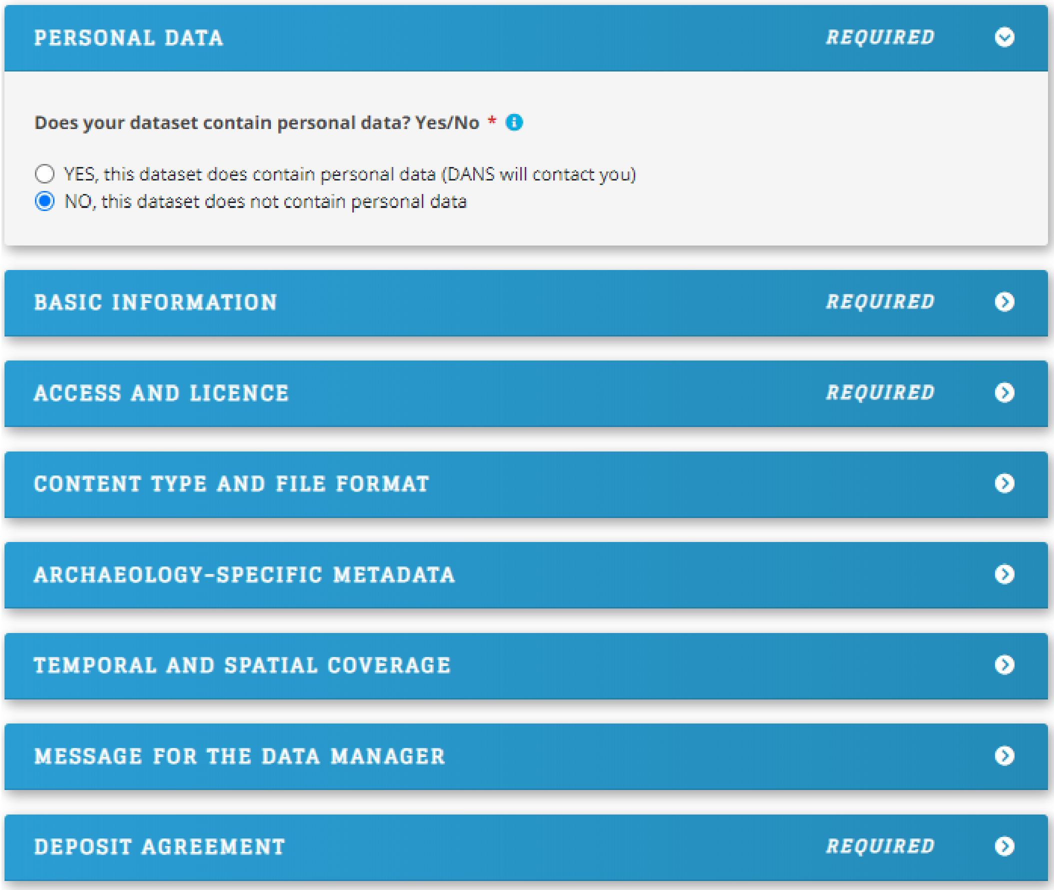


FIGURE 5.18: THE INFORMATION BARS DURING THE DEPOSITION OF PROJECT LEVEL METADATA IN EASY. THE BARS THAT ARE NECESSARY TO FILL IN ARE APPOINTED WITH A 'REQUIRED' AT THE END OF THE BAR.

Project level metadata includes a few other elements that are required to fill in during the deposition procedure in EASY: a question about the inclusion or exclusion of personal data and a deposit agreement with DANS. These actions protect DANS and ensures that the researcher/depositor knows the consequences of depositing a dataset in EASY. However, the question is already asked during part one, the ethical considerations. Thus, the dataset has already addressed the situation of personal data. It also functions as a reminder for each depositor to acknowledge and check the deposited content for mistakes. EASY also provides the depositor with a DOI in the basic information tab, which can be copied to the codebook.

There are a few additional elements that, although not obligatory, are very useful for a dataset. Some of them relate to archaeology specific metadata, such as the subject and object or temporal period. But the most useful option is the 'date available' that offers the option of an embargo on the dataset. If a publication within the dataset still needs to be published, the dataset can get an embargo imposed to delay the accessibility of the dataset. The maximum embargo period is two years. When these elements are fulfilled, the data is ready to be ingested in EASY.

To summarize, this chapter introduced the workflow for generating an object-based 3D dataset. But before the workflow was addressed, the software and hardware specifications to produce the workflow were displayed. The workflow consists of four parts: Ethical considerations, file structuring and file naming, 3D import and export, and finally metadata. the 3D models used are according to DANS recommended standards and are BLEND, DAE, FBX, OBJ, PLY, and X3D. The metadata exists of three smaller components: The codebook, the file list and the EASY metadata.

Now the workflow and the implementation of the workflow on the case study are addressed, it is time to discuss the results of the implementation. The results focus on the impact of choices made during the generation of the 3D dataset. With a focus on accuracy of the generated 3D models and a focus on adherence to the FAIR principles. This focus is necessary because the target of the dataset is to accurately generate a dataset that is suitable for long-term archiving.

Chapter 6: 3D Archiving results

In this chapter, the efficiency and time management of the dataset, the quality of each 3D file format and the adherence of the generated dataset to the FAIR principles (explained in chapter 3.3) are discussed. The efficiency of the generation of both 3D models and the 3D dataset, influences the fit for purpose of object-based archaeological 3D models. The quality of each 3D file format discusses both the implementation of 3D features and the accuracy of the files. The adherence of the dataset to the FAIR principles will revisit each principle by assessing if the dataset has resulted in becoming FAIR.

6.1 Dataset results

In the workflow, a total of six different 3D formats were used to export each 3D model. These six formats each have their own internal structure, adherence to specific 3D modeling features, and requirement of specific additional files to be opened properly as described in chapter 4.3 and 4.4. The workflow enables the possibility to reflect on the different storage sizes of each file and time consumption of generating each 3D format in Blender. However, before addressing each 3D file format in specific, the dataset in its entirety is discussed.

The complete import and export procedure for each 3D model separately required approximately 15 to 20 minutes. This includes the texture name editing in the original OBJ file, the orientation, and the decimation. The file naming and correct use of specifications were the most error prone procedures of the workflow.

The 30 exported models indicate in a time investment of around 10 hours if every performed step would go as exactly as planned. However, the error prone parts of the workflow, especially the texture name adjustment in the MTL file, was forgotten often. Resulting in an unnecessary and time-consuming repetition of the procedure. In reality, an estimation of 12 hours is more accurate. Most of the hours spent was processing time of the exporting of the large 3D models. The time the export of each format required depended on the format and varied much. The export times of eight 3D models have been clocked and the time spent on the export procedure of each format is visible in table 6.1.

	ORIGINAL NUMBER OF FACES	BLEND	DAE	FBX	PLY	OBJ	X3D
9_TEMP	2.5 million	5	20	43	71	118	40
34	2.5 million	5	18	42	71	108	38
85	2.5 million	4	16	41	70	109	69
JA_DD_385_28	2.5 million	5	19	42	71	105	61
JA_DD_385_43	2.5 million	5	18	40	73	100	38
JA_DD_385_44	2 million	4	14	34	58	86	28
JA_DD1_22	1 million	3	7	17	32	45	20
JA_DD1_75	1 million	3	7	16	29	44	16

TABLE 6.1: TIME MEASUREMENTS OF THE EXPORT OF EACH FORMAT IN SECONDS. THE NUMBER OF FACES OF EACH 3D MODEL IS DEPICTED IN THE FIRST COLUMN.

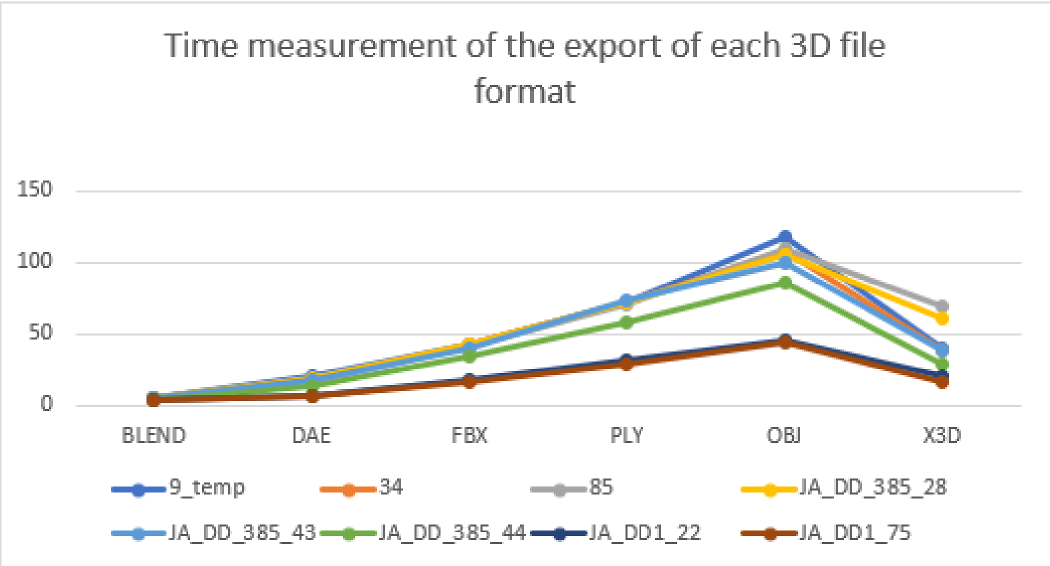


FIGURE 6.1: GRAPH OF TABLE 6.1. DEMONSTRATING THE CONSISTENCY OF THE EXPORT PROCEDURE IN SECONDS.

The last three files in table 6.1, show smaller number of faces than the other five. These smaller face numbers were already smaller when the data was presented to me, as described in chapter 2.2.2. The twelve files that have a different original file size than 2,500,000 faces are displayed in table 6.2. The less than 2,500,000 faces directories demonstrate in figure 6.1 that time for export is longer for files with more faces. Figure 6.1 also shows that for the standard 2,500,000 faces files, the export time varies little and is relatively consistent for the BLEND, DAE, FBX and PLY files. The OBJ and X3D formats vary much. The BLEND, DAE, FBX and PLY formats can be considered more stable formats if time is limited.

For each format, figure 6.1 shows that the OBJ format is by far the slowest to export, followed by PLY and X3D. X3D varies a lot, but also has files that export as fast as FBX files. The BLEND and DAE formats are the fastest to export. The time consumption of each export is presented, but how does it compare to the file sizes of each format?

The complete dataset of the thirty 3D models has a total size of 36.4 GB. The average directory with the original sized 3D models (2,500,000 faces) is about 1.25 GB and deviates for most of the files a maximum of 21 MB, as is visible in figure 6.2 and 6.3. The directories of table 6.2 are not considered in these figures, because they do not have the original 2,500,000 faces and would alter the figures substantially. It becomes evident that the directory sizes are not fully consistent, although they vary relatively little. The reason cannot be linked to the number of faces, because they are all the same for each directory. Therefore, the deviations must be caused by the file formats within the directories.

DIRECTORY NAME	ORIGINAL FILE SIZE	DIRECTORY SIZE
BRB-5_TEMP	2476K faces	1211 MB
BRB-4_TEMP	2468K faces	1204 MB
BRB-8_TEMP	2465K faces	1202 MB
BRB-6_TEMP	2462K faces	1204 MB
BRB-7_TEMP	2447K faces	1192 MB
FP-JA_DD2	2400K faces	1174 MB
FP-JA_DD_385_31_86	2100K faces	1098 MB
FP-JA_DD1_104_87	2020K faces	1011 MB
FP-JA_DD_385_44	2000K faces	1011 MB
FP-JA_DD1_22	1000K faces	508 MB
FP-JA_DD1_75	1000K faces	506 MB
FP-JA_DD_385_45	1000K faces	507 MB

TABLE 6.2: DIRECTORIES WITH ORIGINAL FILE SIZES THAT DEVIATE FROM THE STANDARD 2,500,000 FACES.

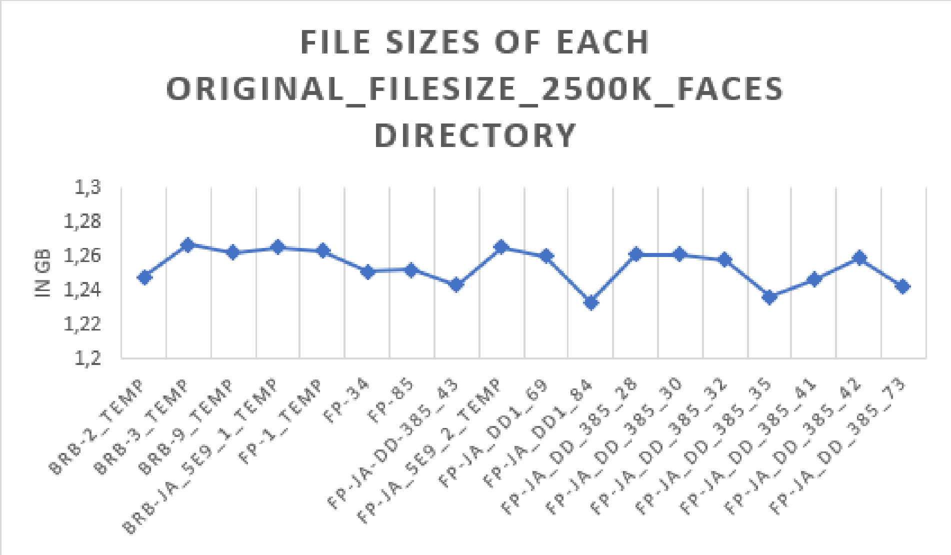


FIGURE 6.2: GRAPH OF FILE SIZES OF EACH ORIGINAL_FILESIZE_2500K_FACES DIRECTORY.

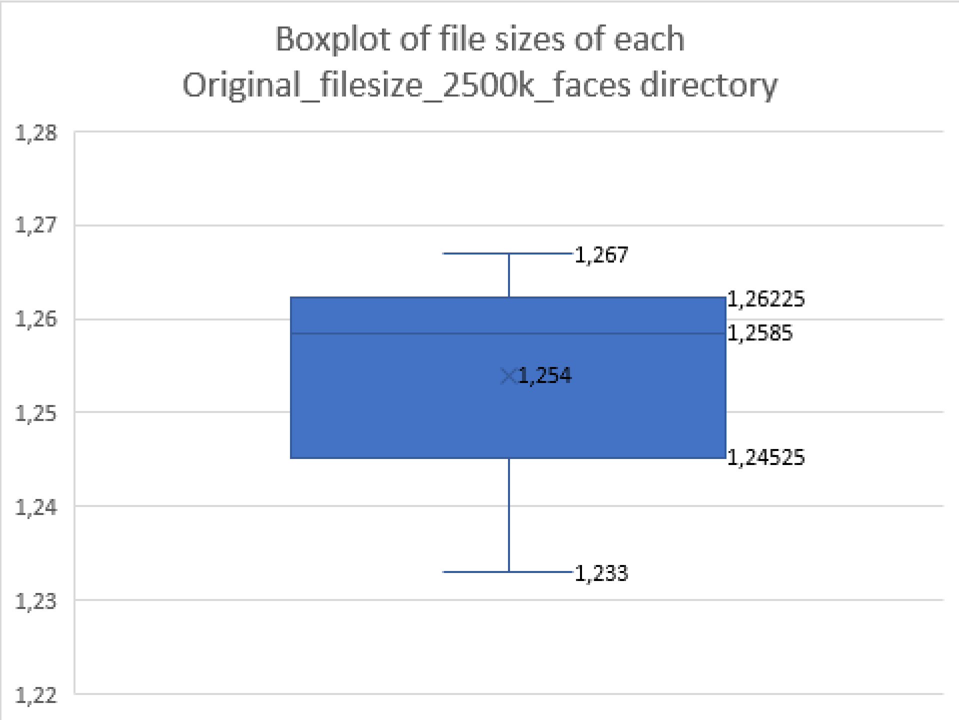


FIGURE 6.3: BOXPLOT OF THE FILE SIZES OF EACH ORIGINAL_FILESIZE_2500K_FACES DIRECTORY.

The average file sizes of each file formats for each 3D model in the original file size and the decimated file size are displayed in figure 6.4 and 6.5. Appendix H presents the varying file sizes of each 3D model for each 3D file format. Figure 6.6 displays an overview of the varying sizes of the original sized 3D models for each 3D file format.

The figures and appendix indicate that each format requires a different amount of storage space. The average DAE file requires more than 350 MB for each 3D model of pottery, while the FBX format only requires about 110 MB per file. The six formats can be divided into two groups: the large files of DAE, BLEND and OBJ, and the small files of PLY, X3D and FBX. When comparing the export storage results between the original file size and the decimated file size, the six formats indicate relative similar storage patterns for most formats, except for BLEND and X3D, as is visible in figure 6.7. These two formats require more storage for the decimated file in relation to the original file size.

The deviation within file formats is large for the DAE, but especially for the FBX format. The maximum deviation of the DAE format is more than 6 MB, while the maximum deviation of the FBX format is more than 15 MB. The BLEND format differentiates for only four files that deviate about 2.5 MB but stays extremely consistent for the other twelve files. The other three formats, OBJ, PLY and X3D show very little deviations and only deviate less than 1 MB. The OBJ, PLY and X3D files display a similar pattern of deviation between the files. The similarities stay relative consistent for all the 18 files, meaning that the deviations cannot effectively be linked to the export procedure of the files. It rather indicates that the deviations can be linked to the different sizes of the files than to the 3D file formats. The other three formats, DAE, BLEND and FBX do not display a similar pattern.

In short, the file sizes of the original sized files and the decimated files variate quite much between each format, but can be divided in a large storage space requirement group (DAE, BLEND, OBJ) and a small storage space requirement group (PLY, X3D, FBX). However, both FBX and DAE display inconsistent file sizes between the files, as is visible in figure 6.7.

When the export time is linked to the file sizes, figure 6.8, the DAE and BLEND formats show that they are able to store a remarkable amount of data in a short period of time. The other four formats show a more persistent storage space to time requirement ratio. The short time requirement of the BLEND format is probably caused by the way Blender stores the internal structure of the proprietary format, described in chapter 4.4.6. The DAE ratio cannot be explained by the internal structure of the file format but might be explained by the storage of the features of each format. These features will be addressed in the next part of the chapter.

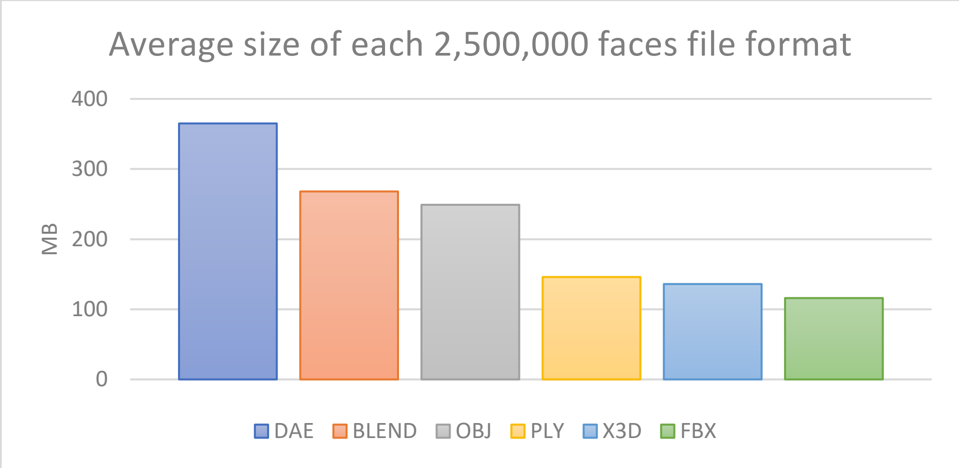


FIGURE 6.4: AVERAGE SIZE OF EACH 2,500,000 FACES FILE FORMAT.

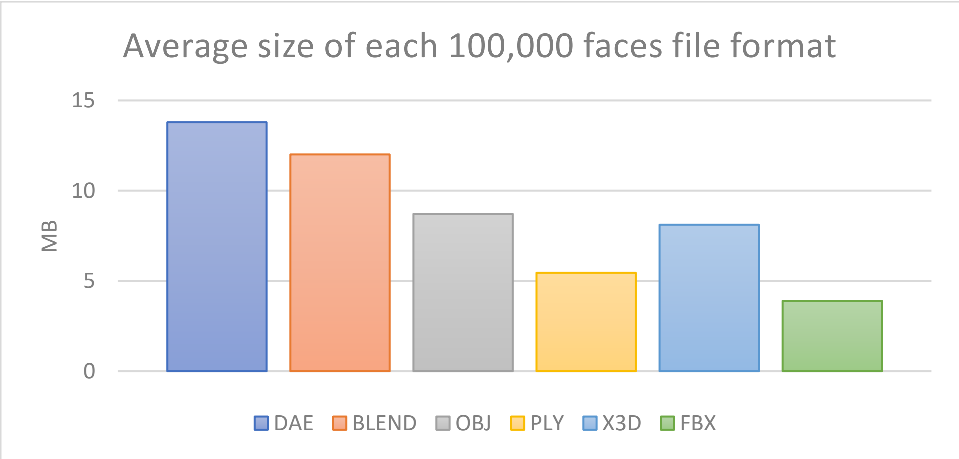


FIGURE 6.5: AVERAGE SIZE OF EACH 100,000 FACES FILE FORMAT.

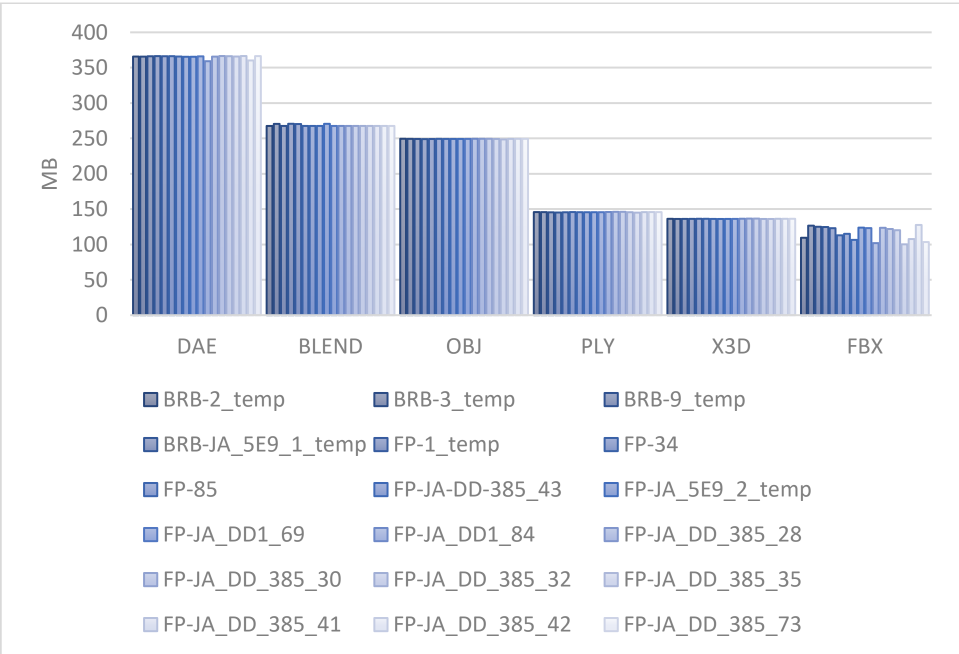


FIGURE 6.6: OVERVIEW OF EACH ORIGINAL SIZED 3D MODEL IN EACH 3D FILE FORMAT.

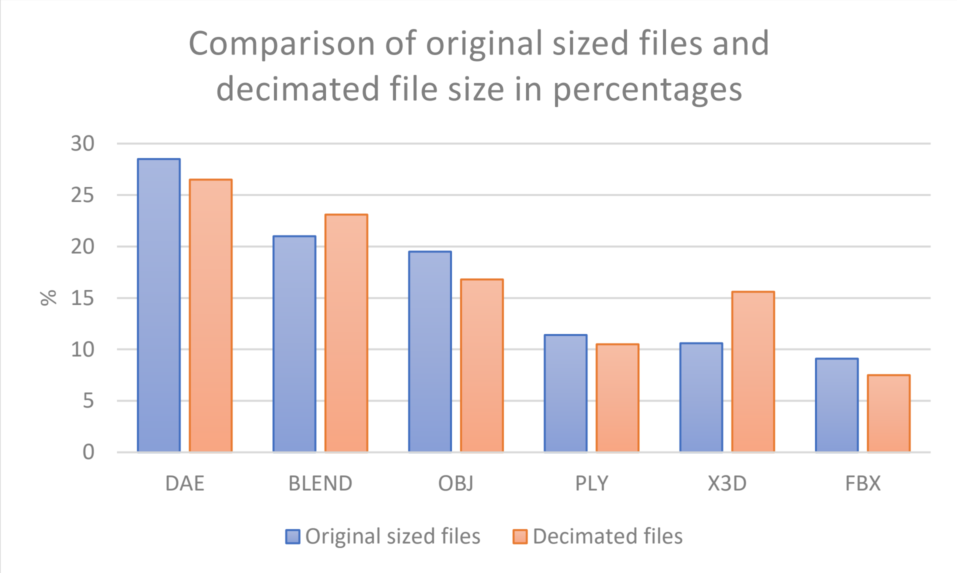


FIGURE 6.7: COMPARISON OF ORIGINAL SIZED FILES AND DECIMATED FILE SIZE IN PERCENTAGES.

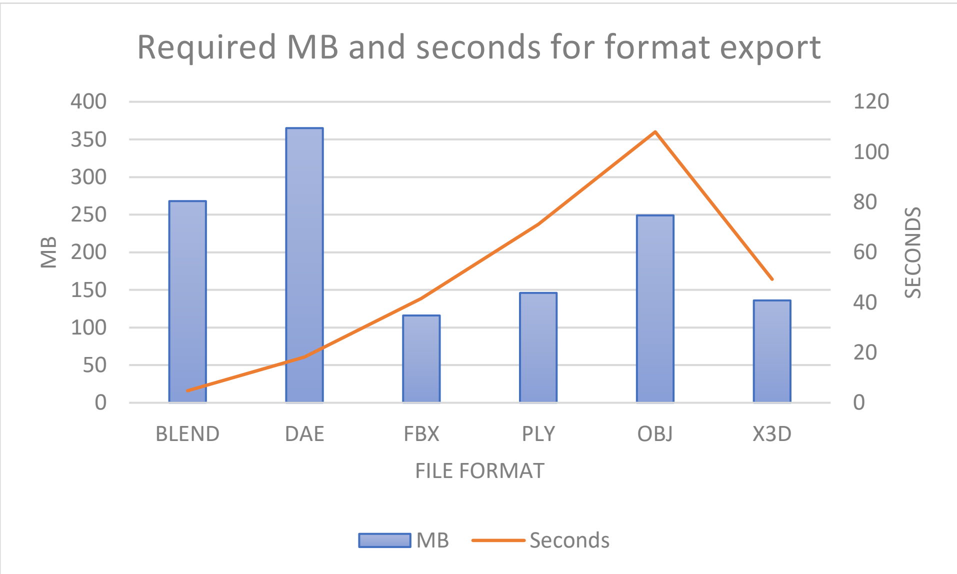


FIGURE 6.8: REQUIRED MB AND SECONDS FOR EACH FORMAT EXPORT.

6.2 Individual format structure and analysis

The next step addresses each format individually and assesses the quality of the 3D modeling features for each 3D file format, described in chapter 4.3. The quality consists of the implementation of 3D features and the accuracy of files. The implementation of 3D features and size measurements are performed on the files in FP-JA_5E9_2_temp and are displayed in table 6.3.

The results of table 6.3 are obtained in Blender 2.80 by importing each file individually and counting the vertices and faces. The files of the BLEND, DAE, FBX and OBJ format indicate a similar number of vertices to the original unaltered OBJ file. The PLY and X3D have slightly different vertices than the original OBJ file, but a similar number of vertices to each other. The number of faces of the exported formats and the original OBJ are analogous. The decimated files are, apart from the BLEND file, identical as well.

The visual aspects in Blender as well as the internal structures of each format of the files of FP-JA_5E9_2_TEMP influence other quality aspects. Appendix I contains figures of the visual analysis of each 3D file format of the meshes and, if included in the file, the textures. The internal structures displayed in Notepad++ are also included.

	VERTICES	FACES
ORIGINAL FILE	1,250,014	2,500,000
BLEND	1,250,014	2,500,000
BLEND DEC.	50,014	100,000
DAE	1,250,014	2,500,00
DAE DEC	50,014	99,994
FBX	1,250,014	2,500,00
FBX DEC	50,014	99,994
PLY	1,257,957	2,500,000
PLY DEC	51,666	99,994
OBJ	1,250,014	2,500,000
OBJ DEC	50,014	99,994
X3D	1,257,957	2,500,000
X3D DEC	51,666	99,994

TABLE 2.3: VERTICES AND FACES COUNT OF THE FILE FORMATS OF FP-JA_5E9_2_TEMP. BOTH THE ORIGINAL FILE SIZE OF 2,500,000 FACES AND THE DECIMATED FILE SIZE OF 100,000 FACES ARE INCLUDED.

For the import of the BLEND file, only the packaged BLEND file is necessary to open. The texture is applied successfully around the mesh. The resulted mesh and the applied texture of the original files can accurately portray details such as notches, dents, inclusions, material damage and other pottery traits, as is visible in Appendix I. The decimated file does show the pottery traits as well, but blurrier. The internal structure of the BLEND file demonstrates the proprietary structure and uses a combination of Blender related terminology and random symbols. However, the bootstrap (chapter 4.4.2) *BLENDER* of the file is visible at the beginning of the internal structure.

The DAE import requires the additional JPG file for the textures. The import in Blender has been successful and the texture has been applied correctly. The DAE format of the original sized file indicates similar accuracy of pottery traits as the BLEND format, but the decimated file is just as blurred. Analyzing the internal structure of the DAE file indicates the human readable structure in XML.

The FBX import do not require any additional files for import. The import in Blender has been successful and the texture has been applied correctly. The FBX format of the 2,500,000 faces file and the decimated 100,000 faces file indicate similar accuracy of pottery traits as the above-mentioned formats. Analyzing the internal structure of FBX indicates the unreadable proprietary structure. However, it starts with the human readable bootstrap *Kaydara FBX Binary*.

The PLY import in Blender is successful, but no textures are applied on the mesh. PLY import only requires the PLY format file. The mesh accurately displays the pottery traits for the original file size. The decimated file size is not very accurate and shows visible triangular faces. However, most of the pottery traits are somewhat visible. Analyzing the internal structure of the PLY format indicates the human readable structure, the *PLY* bootstrap, and the PDI (described in chapter 3.1.2) in the header.

The OBJ import in Blender requires two additional files to open properly, the JPG image and the MTL metadata file. The import has been successful, and the texture has been applied correctly. The pottery traits are evidently visible in the mesh of the original file size, but the decimated file is very vague and blurry. The internal structures of the OBJ and MTL format are both human readable and start with the PDI in the header.

The X3D import in Blender was successful. However, no textures have been applied, because Blender 2.80 does not apply textures on X3D formats. The pottery traits are visible in the mesh of the original file size, but do not indicate the traits as detailed as the other formats. The decimated file is extremely grainy, and faces are becoming distinctly visible. Analyzing the internal structure indicates the XML human readable structure.

From the analysis of the faces and measurements of accuracy, the visual outcomes of the decimated files are blurrier and less capable to deal with physical pottery traits. The original file sized formats display the pottery traits much more accurately and all seem as useful options for pottery analysis. However, for more visually pleasing results, the BLEND, DAE, FBX and OBJ file formats suit better, as they can display the textures by themselves as well. Next, are additional accuracy measurements between the original and decimated formats of the OBJ, PLY and FBX formats. Additional accuracy measurements have been performed on the OBJ, PLY and FBX formats in CloudCompare V2.11 and will be further explained in the next part of this chapter. The BLEND, DAE and X3D formats were not importable in CloudCompare, thus these additional accuracy measurements were not assessed for these formats.

6.3 Geometrical comparison between original and decimated exported formats

Visual measurements are already discussed in the previous part of this chapter, but for OBJ, PLY and FBX additional highly accurate measurement are possible. These accurate measurements are possible because these three formats are importable in CloudCompare V2.11. CloudCompare a FOSS that can compare point clouds and meshes and is used to indicate distance comparisons between the original model and the export generated models.

The OBJ, PLY and FBX file meshes of FP-1_temp and BRB-5_temp are compared to the original point cloud of the OBJ model in CloudCompare V2.11. This comparison indicates alterations that occurred the export and decimation procedure in Blender. The comparison is based on the evaluation of relative geometry quality using CloudCompare by Khalil and Grussenmeyer (2019).

The measurements are very accurate, because the alignment between objects is calculated to a precision of 1.0E-5 cm for 100% of the object surface. Once the objects are aligned, the variations between the original object and the exported object can be compared by measuring the cloud-to-mesh (C2M) distance between the vertices of the objects, as depicted in figure 6.9.

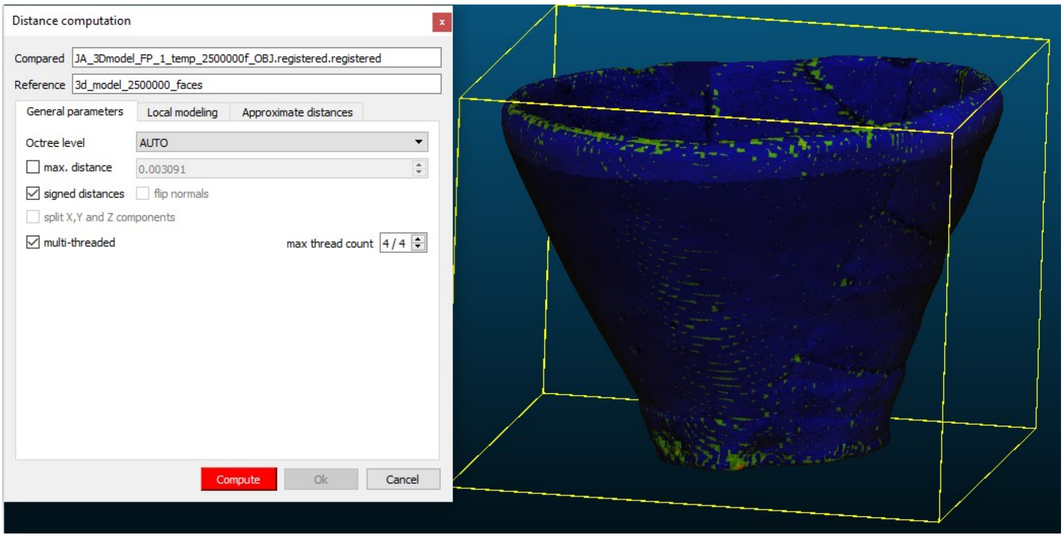


FIGURE 6.9: THE CLOUD-TO-MESH PROCEDURE IN CLOUDCOMPARE V2.11, WITH ON THE RIGHT THE TWO OVERLAPPING MODELS IN GREEN AND BLUE.

The Gaussian (normal) distribution of six comparisons to the original file of BRB-5_temp, are visible in figure 6.10. These distributions indicate that the 2,500,000 faces FBX file has the most similarities to the original file, based on its high central peak and thin distribution. However, the decimated FBX indicates the least similarities. The outcomes of the OBJ and PLY format are very comparable in pattern, mean and standard deviation. The distributions indicate that all three files have altered data. For the original sized files FBX Is the least altered and for the decimated files PLY is narrowly the least altered.

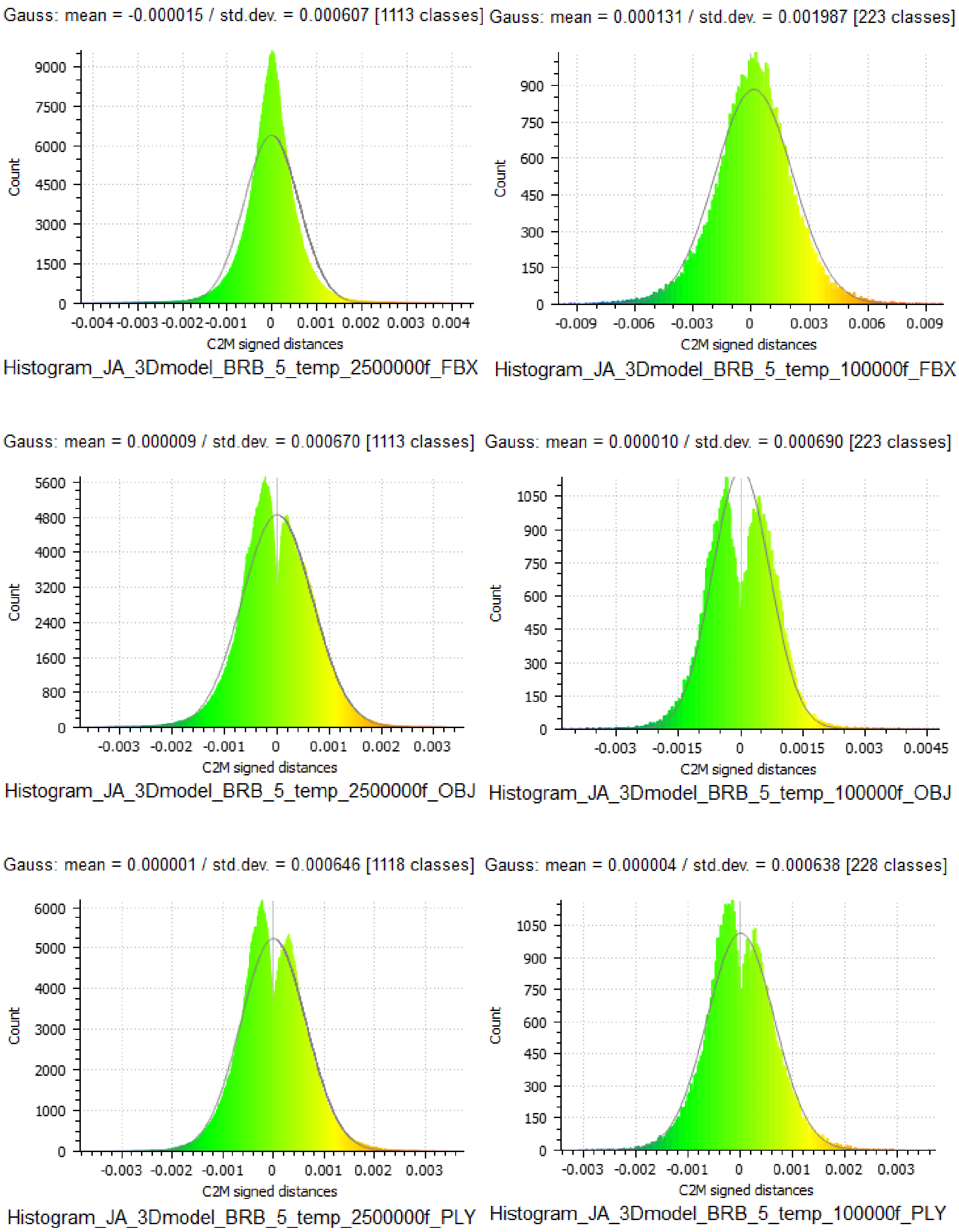


FIGURE 6.10: GAUSSIAN DISTRIBUTIONS OF THE SIMILARITY COMPARISON BETWEEN THE ORIGINAL BRB-5_TEMP AND THE ORIGINAL FILE SIZE EXPORT FILES ON THE LEFT. THE RIGHT CONTAINS COMPARISONS BETWEEN THE ORIGINAL AND THE DECIMATED FILES. THE SELECTED FORMAT ARE FBX, OBJ AND PLY AND THE DISTANCES ARE IN METERS.

The next part presents figure 6.11 to 6.16, which demonstrate deviating points in blue and red in the 3D models. The first three models are formats of BRB-5_temp, while the last three models are decimated formats of FP-1_temp. The distances between the objects are displayed in a blue-green-yellow-red color scale. Gray was used to indicate deviations larger than the set parameters and indicates extreme deviations. All the files were set to have similar distance parameters. The calculations are in meters.

The figures illustrate that most of the deviations were in the notches along the rim and at the base of the model. It also becomes evident that the original file sized models do not deviate very much. However, deviations are much more present in the decimated files, which is probably caused by the decimation procedure in Blender.

What can be concluded is that all files show variations when imported and exported. In the normal distributions, the large FBX format file varies the least from the original. However, the differences are not very large. All in all, the formats confine themselves to relatively small deviations. The deviations that are visible, are mainly located in the notches along the rim and at the base of the model.

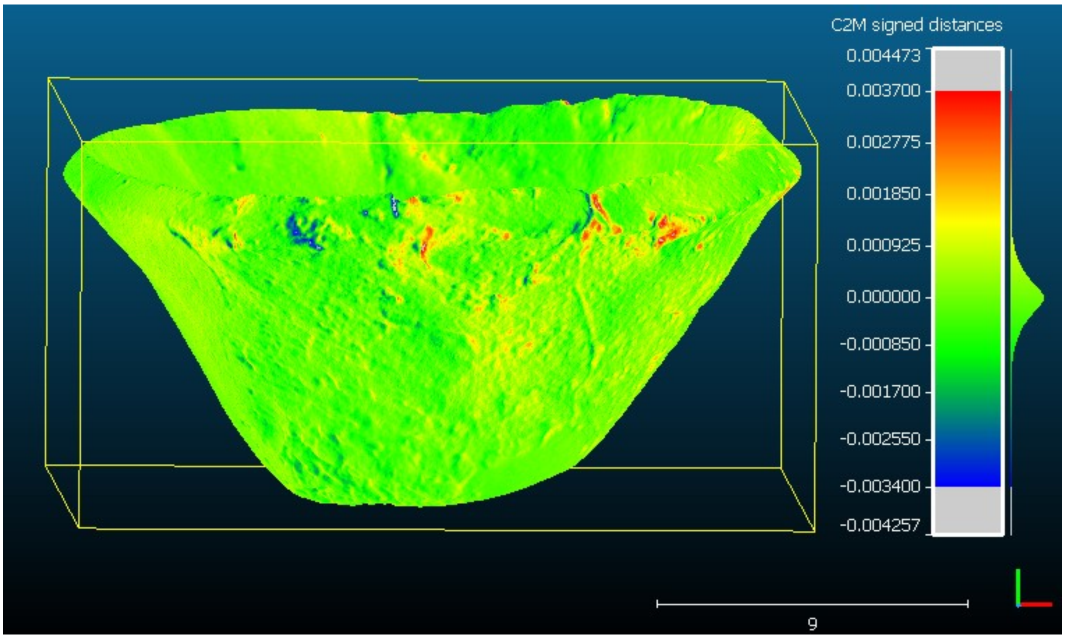


FIGURE 6.11: JA_3DMODEL_BRB_5_TEMP_2500000F_FBX COMPARED TO THE ORIGINAL OBJ FILE IN CLOUDCOMPARE.

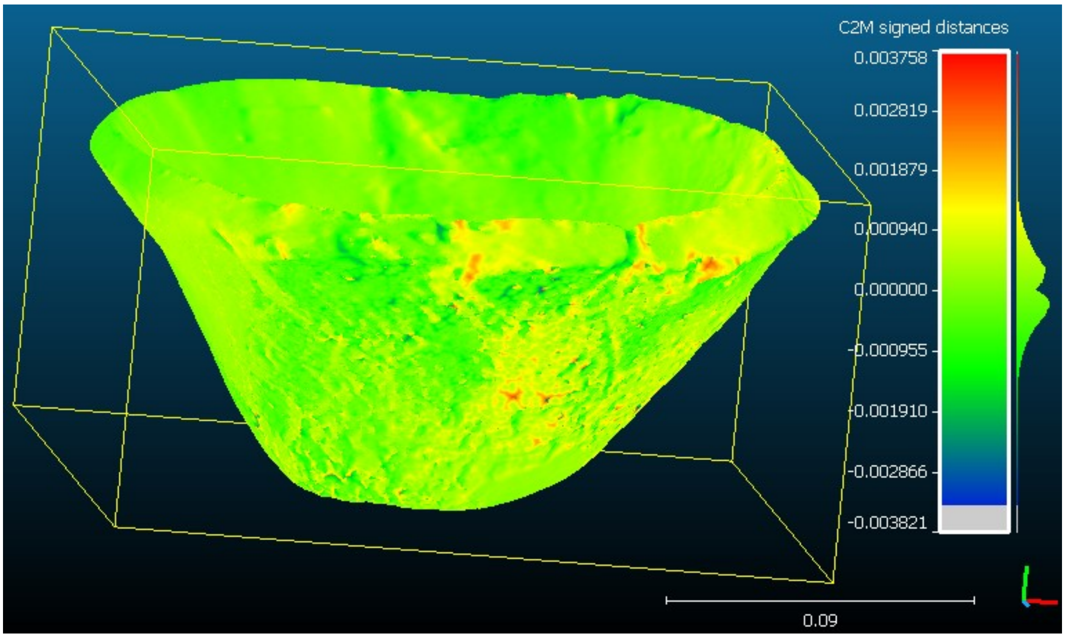


FIGURE 6.12: JA_3DMODEL_BRB_5_TEMP_2500000F_PLY COMPARED TO THE ORIGINAL OBJ FILE IN CLOUDCOMPARE.

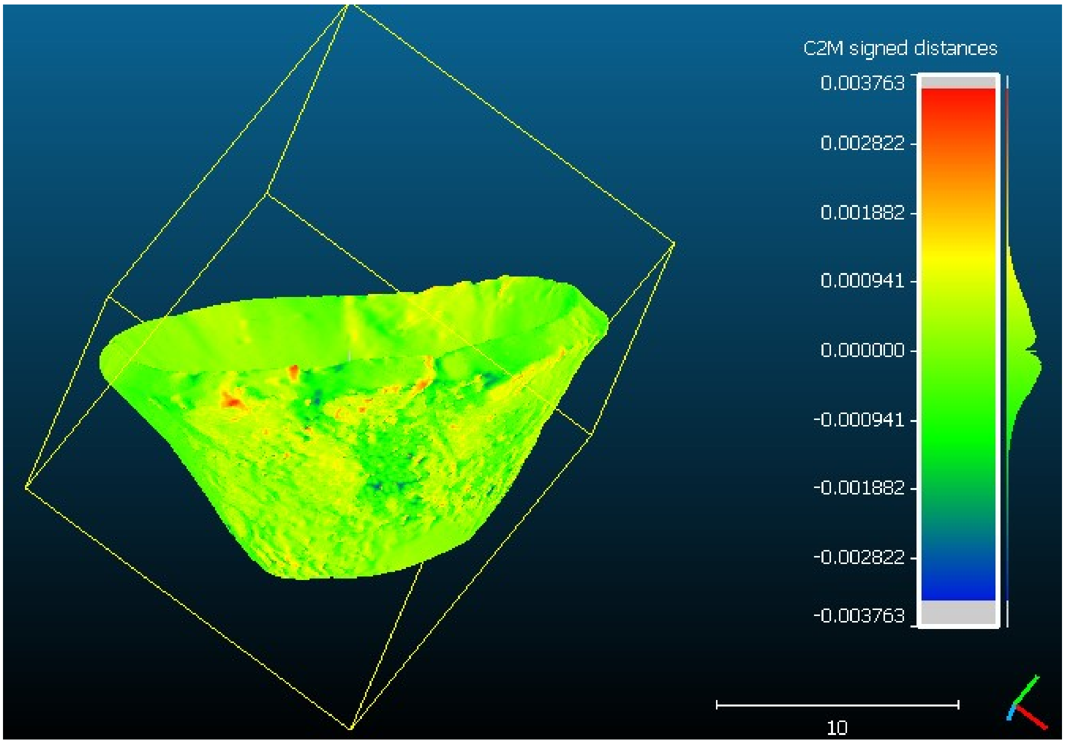


FIGURE 6.13: JA_3DMODEL_BRB_5_TEMP_2500000F_OBJ COMPARED TO THE ORIGINAL OBJ FILE IN CLOUDCOMPARE.

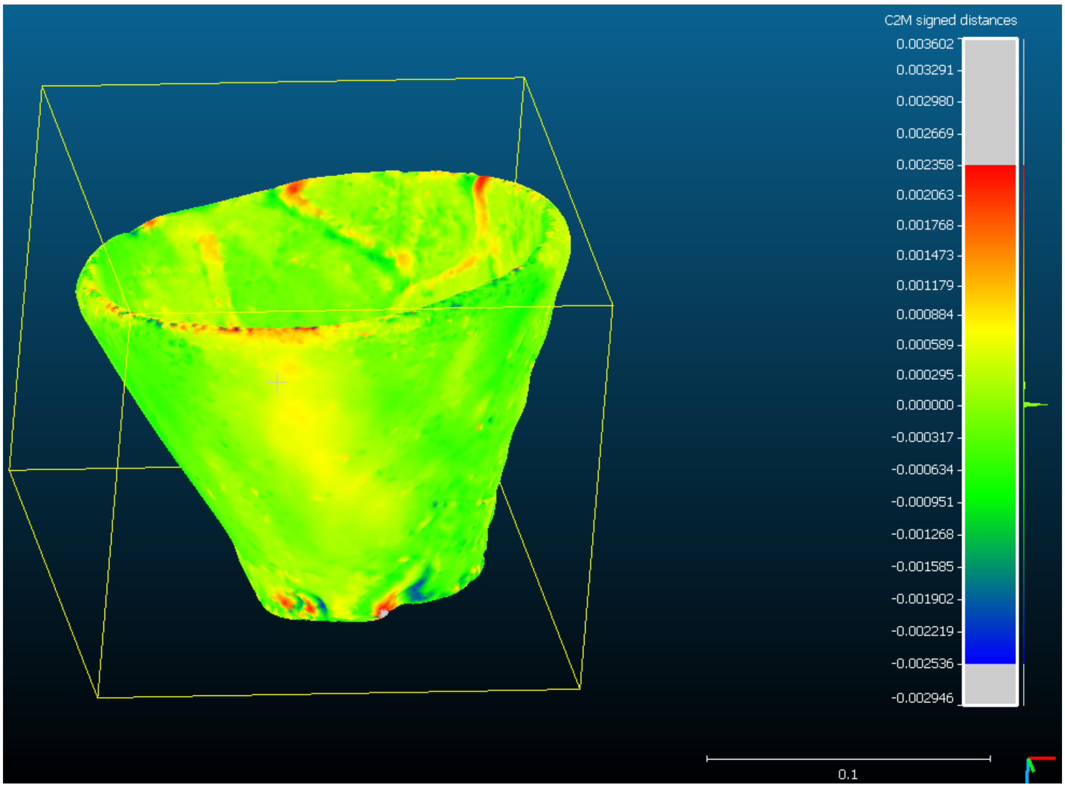


Figure 6.14: JA_3Dmodel_FP_1_temp_100000f_OBJ compared to the original OBJ file in CloudCompare.

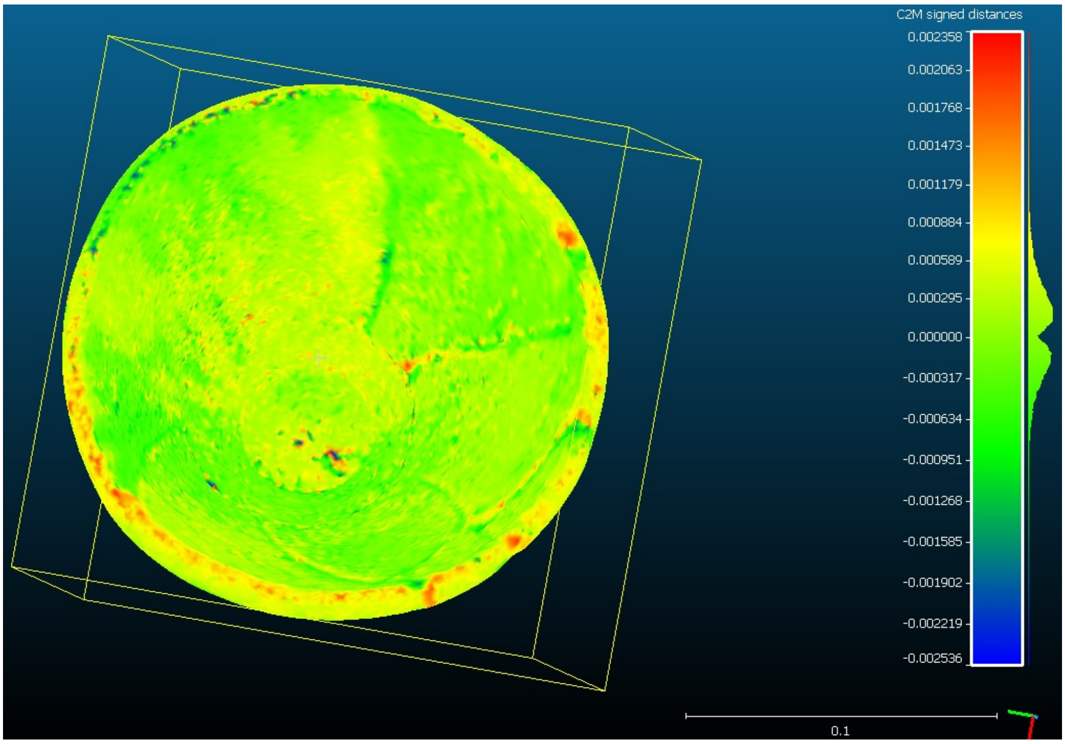


Figure 6.15: JA_3Dmodel_FP_1_temp_100000f_FBX compared to the original OBJ file in CloudCompare.

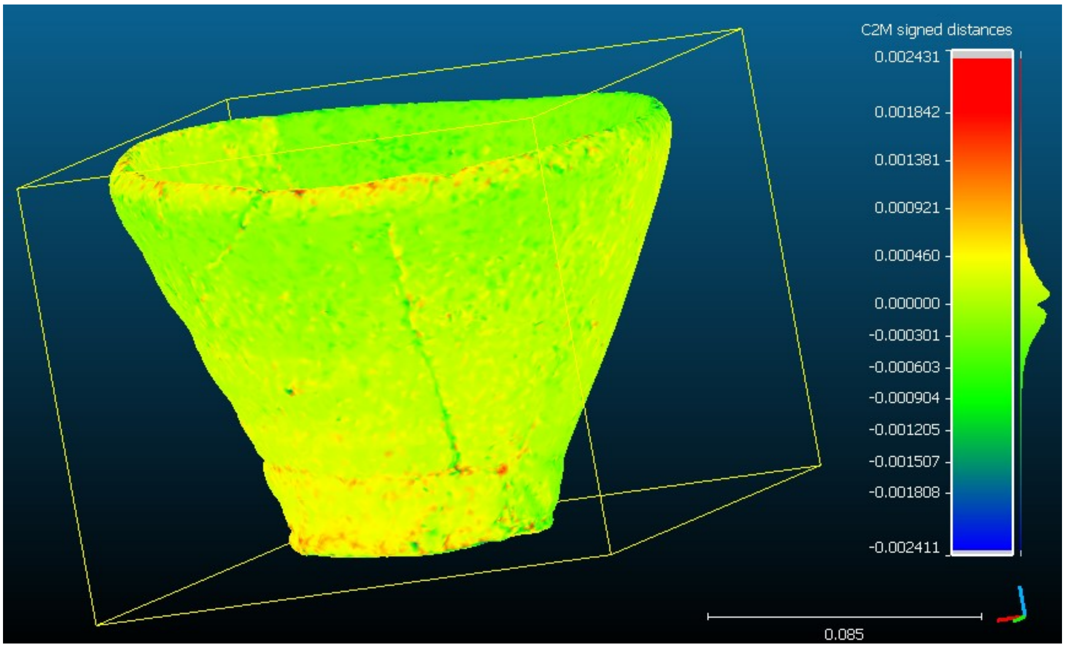


FIGURE 6.16: JA_3DMODEL_FP_1_TEMP_100000F_PLY COMPARED TO THE ORIGINAL OBJ FILE IN CLOUDCOMPARE.

6.4 3D dataset adherence to the FAIR principles

In chapter 3.3 the FAIR principles were discussed. Their target is to bring clarity around the goals and urgencies of good data management and stewardship (Wilkinson *et al.* 2016, 1). Although the principles are perceived as guidelines, they are intended to make the data more Findable, Accessible, Interoperable and Reusable. The results of the generated 3D dataset and EASY are matched to the in chapter 3.3 described parts of the FAIR principles, which are also visible in figure 6.17. EASY has been evaluated on its performance of FAIR principles in the past, but this analysis has been mainly for datasets in general and not for 3D datasets in specific. The matching procedure indicates whether the dataset is long-term sustainable and if the generated data and digital archive provides enough longevity. Table 6.4 presents the results of the matching procedure.

F3, F4 and A1.1 are not influenced by adding or adjusting information within the 3D dataset. The findability and accessibility are mainly targeted to metadata, while the reusability and interoperability is focused on both the data and metadata.

In general, the generated 3D dataset adheres to all the FAIR principles to a certain extend. The 3D data is compliant to the FAIR principles because the data follows the recommendations of EASY. The 3D data is generated to retain as much of its original value and with the combination of the 3D file formats, the interoperability is very good. The only thing that are not fully addressed in the 3D data are scenery or other geometrical features, as described in R1. These additional information sources are not generated during the acquisition method but could provide a better representation of the data. However, the 3D dataset is constructed to retain most of its original value and to only add essential information. Therefore, the option is chosen to not add the scenery information, because it influences the outcome of the storage procedure.

The metadata is according to I2, I3 and R1 not rich enough in metadata elements. Although the documentation is provided in three different metadata levels, the metadata does not fully include all the elements of any of the mentioned metadata schemas. However, only the essential elements of metadata schemas in EASY are addressed.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

FIGURE 6.17: THE FAIR GUIDING PRINCIPLES (WILKINSON *ET AL.* 2016, 4).

Findable	
F1	The dataset is provided a DOI by DANS that is included in both the EASY metadata and the JA_Codebook.pdf. EASY also incorporates the URN:NBN identifier. The DOI is a globally unique and persistent identifier.
F2	The described metadata and paradata of the dataset address the project level and the object level and display information to understand how the data is generated and which problems occurred during the generation. The codebook also facilitates the use of 3D models, with the short introduction at the start, explaining which files are required for opening each format. All metadata files are stored in both human- and computer-readable languages.

F3	F3 is not a 3D dataset specific point. However, as mentioned in F1, the JA_codebook within the dataset includes the DANS generated DOI persistent identifier.
F4	F4 is not a 3D dataset specific point. F4 is dependent on the metadata harvest of EASY, which allows harvesting of metadata records through the OAI-PMH (Tsoupra *et al.* 2018, 7). Only the EASY metadata is harvested, not the dataset itself.
Accessible	
A1.1	A1.1 is not a 3D dataset specific point. However, the metadata in EASY is retrievable through open, free, and universal protocols (Tsoupra *et al.* 2018, 8).
A1.2	The licensing in both the codebook and in the EASY metadata in combination with the ethical considerations generates dstinct and concise indications of authorization. The authorization is dependent on the contributors, authors, as well as the presence of personal and legal information within the dataset.
A2	A2 does not suffice to the data within the dataset, but rather to the EASY metadata and the open privacy policy of EASY. Which will remain accessible if the data is no longer available. The file list and codebook metadata and paradata are contained within the dataset and will therefore be no longer accessible if the data is not available.
Interoperable	
I1	The combination of the six recommended file formats of EASY provide both computer- and human-readable formats of the 3D data that are useable in many 3D modeling applications. However individually, the formats of BLEND and FBX are not human readable and only computer readable in specific 3D modeling applications. The metadata of all levels is stored in XML, CSV, or TXT formats, which are human and computer readable formats as well.
I2	The described metadata levels mainly incorporate the required elements of the DC metadata schema (described in chapter 3.2.1). The codebook also incorporates certain elements of the CARARE metadata schema and the codebook itself is an indicator of vocabulary and jargon of the dataset. Both DC and CARARE metadata are integrated in the policy of EASY.
I3	Although not obligatory, the codebook and EASY metadata are both able to indicate references to related data and datasets. It is one of the elements of the DC metadata schema. However, the referred data is not really addressed in the EASY metadata, because it is not an essential element.
Reusable	
R1	This dataset is intended to be generated from a practical point of view and maintains a short and concise list of metadata elements. However, the 3D data is generated to maintain as much of its original value. The variety of the 3D file formats provide sufficient

	information for many 3D features and data elements and can therefore be considered as richly addressed. The lack of different geometrical features (NURBS and CSG) and without lighting or scenery can be considered a limiting factor on reusability. However, these additional features can also be perceived as preventing the essential goal of the dataset, which is to display object-based archaeological 3D models. Only certain metadata elements of DC or CARARE schema are fully addressed in the metadata. However, the licenses and the provenance are provided.

TABLE 6.4: EVALUATION OF THE 3D DATASET TO THE FAIR PRINCIPLES.

In summary, the result chapter gave an in-depth perspective of the generation of the 3D dataset, with a focus on time management, content quality, consistency of 3D file formats and Blender, and lastly, adherence to the FAIR principles.

Of the six exported formats, the results indicate that the BLEND and DAE formats are the quickest to export and the file sizes of the original sized files and the decimated files variate quite much between each format. The formats can be divided in a large storage space requirement group (DAE, BLEND, OBJ) and a small storage space requirement group (PLY, X3D, FBX).

For more visually pleasing results, the BLEND, DAE, FBX and OBJ file formats suit better, as they are able to display the textures as well BLEND and FBX are even better, because they have the texture embedded within the format file.

The three formats of OBJ, PLY and FBX were applicable in CloudCompare and were compared to their original OBJ file for two pots. It became evident that all the three file format have a little bit of data loss/deviation, but indicate relatively similar patterns for deviations.
The adherence of the 3D dataset to the FAIR principles indicate that the 3D data mostly complies to the principles, but that the metadata of the dataset could be more extensive and cover full metadata schemes.

Chapter 7: Discussion

This chapter discusses the interpretations and implications of the presented results in the previous chapter. The first part focusses on how the dataset and its 3D components fit to the purposes of 3D models in archaeology and which format(s) suits the longevity aspects best. The discussion on the purpose establishes the user and digital archive aspects of this thesis. Secondly, the digital archive of DANS is discussed for its usefulness of preserving 3D data, the usefulness of EASY establishes the digital archive perspective. The third part discusses the implications of the resulted adherence of the dataset to the FAIR principles. Lastly, the tools are reviewed on their use, followed by an assessment of the (lack of) benefits of FOSS.

7.1 3D Data aspects and fit for purpose

In the workflow, the 3D data has been exported to six different formats: DAE, FBX, OBJ, PLY and X3D. In the results chapter, each format is addressed regarding its internal structure, output, adherence to 3D modeling features and time and storage requirements. The results imply that all the tested 3D file formats can store object-based archaeological 3D data. Each format seems to have its own perks and drawbacks that suit specific purposes and uses within 3D applications. For archaeological purposes, 3D models would have the most value if they are stored in two sizes as described in chapter 4.1. One 3D model size that is used for research that is focused on realism and requires high-resolution, photorealism and interactivity and where storage space is not the biggest issue (Remondino and El Hakim 2006, 274). And another 3D model size that is used for smooth interaction, fast loading, general ease in accessibility and functions appropriately in 3D assemblages. The first one obviously requires large file sizes and the second one prefers smaller files. For most of the museum purposes, the smaller files are more useful as they do not require much storage space and are therefore more affordable to maintain and preserve. Regarding 3D modeling features, most of the original values should be retained. Meaning that object-based 3D models in archaeology are dependent on approximate geometries and heavily depend on textures. The colors of the models should be preserved as well, incase uncertainty must be visualized. Scenery or animation elements can be added but are not necessary. Mainly because generated 3D models during digital data acquisition techniques are originally not stored with any of these two 3D modeling features. The comparison in CloudCompare of only three of the six formats generate a difficult assessment of accuracy between the formats. BLEND, DAE and X3D did not get an in-depth accuracy testing, thus accuracy is not as delineated and evaluated as the other three formats. The impossibility of importing BLEND, DAE and X3D indicate an inferior interoperability for necessary 3D modeling software.

The open and human readable 3D file formats provide the most preservation potential, although as it currently seems, most or all of the tested 3D file formats will still be used in the next 10 to 20 years. Unless a revolution within the 3D modeling community appears that completely reinvents the structure and use of 3D models, all selected 3D file formats will still be useful. The selected 3D file formats are all at least 15 years old and are to a certain extend still popular today. It seems that all 3D file formats might still be around a while. And all formats together will provide the best longevity for the content of the 3D dataset. However, some of the formats are distinctly better equipped for the purposes of object-based archaeological 3D models than others.

Based on the results of the previous chapter and purposes presented in chapter 4.1, the FBX format suits the most requirements. The accuracy comparison with the original 3D file in CloudCompare indicated that the original FBX file was the most accurate between the three formats. The FBX format also has low storage requirements and does not require much storage time, which are advantageous for the interactivity and visualization aspects of 3D models for museums. The possibility of storing the texture file within the format has huge advantages for beginner 3D modeling users and for transferring purposes, because no additional files are required that can get lost or corrupted. Potentially, all 3D formats generated in the future might automatically incorporate the image texture in the file format, just as the FBX and BLEND files currently do. 3D file formats such as DAE and OBJ require multiple files to function properly and do not adhere to the one-file-fits-all format concept. If the DAE or OBJ format are used in the future, users might think it is corrupt because the texture is not applied, only because they are unknown to required additional files.

Other advantages of the FBX format are the interoperability and the bootstrap within the internal structure that ensure correct usage in 3D modeling software. However, FBX has three disadvantages. The format is proprietary, is the least consistent in storing data and handles the decimation procedure not very well. The decimation process is less of an issue, because the storage requirements of the file are so small already, even without decimation. Though, when these three disadvantages combined are considered, preserving only FBX files of 3D data seem unwise. Therefore, generating OBJ formats of 3D files as well, will counter the disadvantages of FBX. OBJ files do require more storage space, require the longest time to export and cannot store CSG, scenery or animation features of files. The advantages of OBJ files are the formats ability is able to store data very consistently, its ability to link the texture to the 3D model, incorporation of a human readable structure, and the ability to keep the accuracy of the original file relatively well. Although there is much overlap between the DAE format and the OBJ format, the OBJ format gets preferences based on the lower storage requirements and its interoperability within 3D modeling software.

The combination of the FBX and OBJ files have the disadvantage that the files take up more storage space. However, the storage space the two formats combined require is almost exactly as much as the DAE format on its own requires.

The other formats each have some value for the fit for purpose, depending on the situation. The DAE format is like the OBJ format and the main advantages are its relative interoperability and the accuracy the 3D file format displays. The disadvantage of using DAE is the size of the files and its impossibility to be opened in CloudCompare.

The PLY format would be the most useful for object-based archaeological 3D models if 3D objects have no textures and only depends on vertex colors. All current 3D data acquisition techniques used in archaeology use textures to accurately display the objects surface. Textures can display much more detail than vertex colors. PLY has the advantages of needing very little storage space, portraying the original file accurately and consisting of a human readable file structure.

X3D has the same texture problem as PLY and is similar as well in export time and storage requirement. However, X3D indicates a less accurate display of the geometry for visual purposes. The X3D format is mainly designed to support virtual environments, which in this case are nonexistent. Of all the formats, X3D is the least accurate and therefore the least capable to store object-based archaeological 3D models.

The BLEND format has the disadvantage of being a proprietary format and containing a human unreadable structure. The BLEND format also requires much storage space and is not interoperable, unlike the other proprietary 3D file format FBX. An advantage of the BLEND format is its ability to store data quickly and inclusion of textures within BLEND files. The bootstrap within the internal structure can ensure correct usage in Blender. The severe limitation of the BLEND format only being usable in Blender is the main reason why the format is not very useful for preservation. The format is not applicable in other 3D software, meaning that it is impossible to display the format on tablets or in a virtual environment.

The advisable file structure of a 3D dataset for object-based archaeological 3D models should, according to the 3D formats that fit the archaeological purpose best, exists of the two 3D file formats of FBX and OBJ, instead of six formats. The MTL file and a JPG texture file for the OBJ file have to be included as well. While using all six different formats in the dataset will certainly ensure very good longevity, the inconvenience and effort to generate four additional formats outweighs the probable extended longevity.

7.2 Digital archive considerations and durability

DANS has considerable requests before a dataset can be deposited in EASY. Fortunately, most of these requests have a low minimum requirement and can be performed relatively quick when the correct tools are used. Although the framework for deposition of 3D data in EASY is indicated within the guidelines of DANS, the experience DANS has with storing 3D models could be better (V. Gillissen, personal communication). Gillissen showcases alternatives for 3D models such as videos as indicated in the dataset '*Animated, interactive 3D visualization of a corn mill, after a description by Ramelli'* (https://easy.dans.knaw.nl). Videos can highlight specific parts of the 3D model but eliminate the ability to freely scale and move the model and remove the possibility to accurately measure distances for archaeological research. Both depicted as essential purposes of 3D models. Transferring 3D models to video also means deviating and altering the original model, creating the potential data loss if the video does not display all the needed elements of the 3D model.

An alternative to storing 3D data in digital archives is storing data in one of the increasing numbers of digital 3D libraries. *TurboSquid* (www.Turbosquid.com) and *SketchFab* (www.sketchfab.com) are substantial digital libraries for specifically 3D modeling, but many smaller and specialized alternatives are available too, such as the journals *Internet Archaeology (*https://intarch.ac.uk/) and *Studies in Digital Heritage* (https://scholarworks.iu.edu/). These alternatives allow embedded 3D models in publications. And the focus of the digital 3D libraries is to turn 3D into a mainstream media format or to save time on modeling. Embedding 3D models in publications requires the U3D or PRC file format. U3D and PRC are currently the only 3D file formats that are embeddable in PDF and only U3D is recommended in EASY. However, U3D is yet to be implemented 3D file format in Blender or in other FOSS software. The U3D is a format that is likely to decline over time and lacks support of essential elements (https://www.loc.gov). Even though its use might be declining, it would still be interesting to match the quality aspects and usefulness of the U3D format to the other six recommended formats in EASY.

The purpose of a digital archive is to preserve research data for future reuse. Until now, all digital 3D libraries are not institutionally supported and have a commercial viewpoint. Resulting in the facilitation of uploading by supporting a wide variety of file formats and focusing on obtaining a wide audience and discoverability by adding tags and categories. 3D digital archive alternatives lack structured metadata and quality harvesting options, because the only available fields of data entry in digital 3D libraries are the description, the model information description, and the optional README file within the download itself. Commercial alternatives are not able to guarantee long-term maintenance or archiving of models. EASY is not a digital archive that is specifically focused on 3D models and does not support interactivity and direct display of 3D models. Many digital 3D libraries do support these interactivity elements by using a previewer. The lack of a previewer is a big negative aspect for preserving 3D models in EASY.

Complex actions are required for depositing in EASY and each action for data deposition is highly dataset specific. However, the guidelines for depositing created by DANS contribute enormously to get through this complex process. The huge number of open access datasets already within EASY also help immensely for quality measuring. The guidelines of EASY help researchers to contribute to help future users of the dataset to easily find what they want through the extensive metadata and provide support for ethical considerations by offering a processing agreement with DANS. It is difficult for digital archives to acquire extensive datasets of researchers if the researchers themselves are required to do the difficult and boring additional work. That is the reason why a wide variety in extensiveness of datasets exists.

For preserving 3D models in general, EASY regulations seems slightly outdated. DANS does not seem to accurately distinguish the different aspects of 3D modeling features in their guidelines of preferred formats and does not consider increasingly popular formats such as STL or C4D. STL is hugely popular in the 3D printing community where it is advocated as a defacto standard within 3D printing. STL contains a data structure that is archive suitable because of the choice to export STL files in a neutral ASCII format or in a binary format. C4D is a proprietary format of the 3D modeling application Cinema4D, that has got much traction recently within the 3D modeling community. The lack of these formats suggests that the 3D formats options within DANS are established at least some years ago and are not updated or maintained frequently as they do not adhere to contemporary usage. They are not considered in this thesis, because the lack of recommendation in EASY indicates that the versions will not be maintained as sufficiently as the recommended 3D file formats. A quality and usefulness comparison between these formats and the recommended file formats would be an interesting endeavor as another research topic. Although DANS implies it regularly updates the recommended file format lists for its infrastructure and preservation strategy, the lack of C4D and STL as recommended 3D file formats indicate that the obsolescence-prevention is not very active.

For object-based archaeological 3D models specifically, the options DANS offers are sufficient. Within 3D modeling, object-based archaeological models are relatively uncomplicated for storing, compared to scenery 3D models with ultra precision and animations. The 3D modeling features that this thesis include for object-based 3D models, such as triangulation, color or texture incorporation and high-quality details created by the acquisition techniques, are represented within the offered formats by DANS. The biggest drawback of using EASY is that 3D models must be downloaded from EASY to visualize or manipulate them. However, with the existence of decimated models, the metadata and the coherent and structured filenames, the choice of downloading a specific 3D model is sufficiently easier and more accurate.

7.3 FAIR-ness of the 3D dataset in EASY

In the results, the 3D dataset was evaluated based on the FAIR principles. The 3D data in the dataset did comply to the FAIR principles, but a dichotomy appeared during the analysis of the metadata adherence to the FAIR principles. The considerations from the FAIR principles expect rich metadata and extensive use of vocabularies (metadata schemes). By approaching the generation of the 3D dataset from a practical perspective, only the essential metadata elements were addressed. The essential metadata elements covered parts of the DC and CARARE metadata schemas but not in its entirety. However, what does incomplete adherence imply for 3D datasets?

First, the 3D dataset complies with the 3D data and metadata requests that EASY has. EASY provides a few data management requirements and even with only these obligatory elements, DANS acknowledges that the data is suitable for preserving. EASY sends the generated EASY metadata to be harvested by other digital archives using the OAI-PMH, ensuring that the findability and accessibility is good.

Secondly, the terminology of rich metadata within the FAIR principles is vague. How rich is rich enough? There are many different metadata schemes currently available, with many of them addressing broad or highly specific information sources. Metadata and paradata should enhance the data, but if all the different metadata schematics are used and all the information is provided, at some point the usefulness backfires. Too much information about a dataset can negatively surprise and overwhelm future users and therefore lose its core and essential element, providing valuable and appropriate information that improves the use of the data. The essential elements given in this 3D dataset, provide enough information to enhance the data to not overwhelm the reader. They also comply to the request of the digital archive.

In short, the parts of the DC and CARARE schemas that the metadata and paradata of this 3D dataset include, are general enough to provide adequate information about the dataset. The documentation of the data is implemented using three levels of metadata that are generous and adequate for other people to understand. With the current given information, other and future users are able to replicate the data(set). The codebook helps future users to find the documentation on the generation of the model, incorporates structured information on how to use the 3D models and provides general information and a persistent identifier to link the data to the digital archive. The file list and codebook function as a map and legend that guide future users through the structure of the dataset, whilst providing content information for each file. The EASY metadata shows the right content information that gets the attention of people that use EASY or other digital archives. Together, the provided levels of metadata do not overwhelm the reader and in the very limited case that the provided information is not enough, associated human- and information-resources are specified within the EASY metadata and codebook and can guide future users further.

7.4 Tools and FOSS review

Throughout the thesis, five applications were used and all of them were FOSS. Four of them:

Blender, Notepad++, Dirlister and LibreOffice Calc were used during the creation of the 3D dataset, while CloudCompare was only used during the 3D model quality analysis. Blender and Notepad++ were, besides during the workflow, also used during the 3D model quality assessment in the results, but their appearance in the results was relatively minor.

The next part of this chapter first assesses the confusing perspective of using FOSS for the workflow. This is followed by an evaluation of the purpose and overall usefulness of each applied FOSS separately. Lastly, a statement about operating and using FOSS in general to generate an archive ready 3D dataset is formulated.

The perspective of using FOSS for this workflow and topic can be perceived as confusing, as the original models were generated in the proprietary application Agisoft Metashape.

The first arguments for using FOSS is that this workflow is for a wider audience than only for users who generate photogrammetric models using Metashape. There is a steadily increasing number of alternatives of FOSS applications that generate 3D models through photogrammetry and laser scanning, such as COLMAP and Regard3D. However, these applications do not always have the variety of 3D formats available that digital archives expect of them. These applications also do not provide extensive export specifications that Blender can provide.

Secondly, the target of the workflow is to demonstrate the procedure of making data archive-ready, which indicates that 3D data acquisition software in itself cannot provide enough modeling to generate an archive ready 3D dataset, as became evident during the four steps of the workflow. Thirdly, Archaeology increasingly uses FOSS for the concept of Open Archaeology. As mentioned in the introduction and chapter 4, affordability is a primary motivation for choosing applications and archaeologists are not known for their abundance in capital. FOSS will undoubtedly be used increasingly within archaeology, thus critically assessing FOSS ensures that advantages and disadvantages for specific purposes are clarified. Resulting in feedback and improvements over the long run.

Blender v2.80 offered an abundance of possibilities for configurations for many of the 3D file formats DANS prefers. 3D modeling software are the only approach to accurately alter 3D data and is therefore considered indispensable during the workflow. Of the 3D modeling software currently available, Blender is one of the most used 3D modeling applications and shows no indications that the application will be abandoned anytime soon. Blender can be used to model, sculpt, UV edit, texture paint, alter shading, animation, UV composite, script in Python interactive console 3.7.0 and render 3D models (http://Blender.org). However, only the modeling aspect as well as its capabilities of import and export options were extensively used. For those specific functions, Blender is a valuable addition for 3D modeling after the generation of 3D models. The 3D format import and export options can include and exclude highly specific 3D modeling features and storing concepts, which diminishes the required storage space significantly. However, some minor issues occurred during the use of Blender for the workflow and analysis. One major issue is the lack of tools for accurately comparing different 3D objects. The X3D export in Blender also incorrectly exported the texture, resulting in X3D files without texture.

Of all FOSS aspects, the cost-free aspect of Blender was its most valuable asset. The community of Blender that is derived from the transparency of the application, is used in numerous occasions as well. The abundance of YouTube videos and forums offered many answers to Blender related questions that are not always findable within proprietary software.

Notepad++ fulfilled the requests asked of the FOSS appropriately. Notepad++ is very similar to Microsoft Notepad but is even more structured and user friendly than its MS doppelganger. Some of the additional features the application has, quickly come forward during usage. The very extensive options within the options panel showcase accurate possibilities of selecting and editing texts within lines. The numbering on the left alleviates and structures the data and is very helpful in quickly assessing in which row information is described, as visible in figure 7.1. Blue balls next to the rows also mitigate the ease of use. Lastly, the automatic saving option indicates another convenient and structural deviation from MS Notepad. Although most of the utilities of this application were not utilized in its full potential, it still offered abundant information options for its task within the workflow.

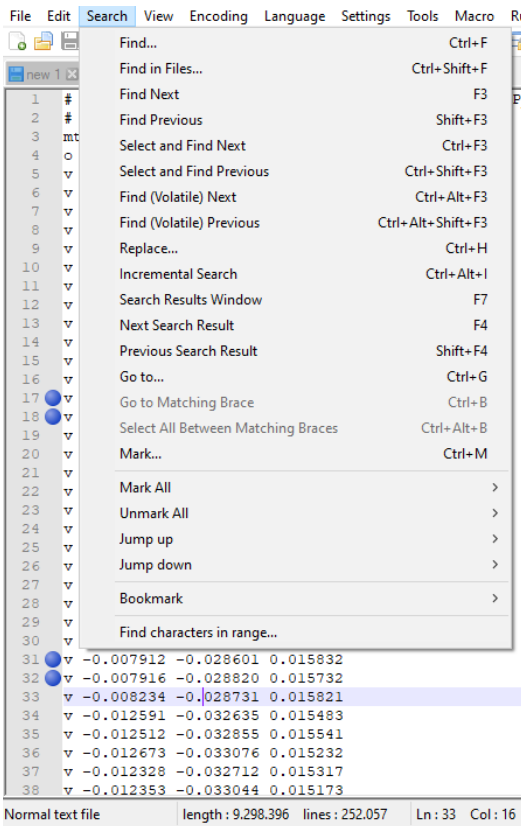


FIGURE 7.1: ALL THE OPTIONS WITHIN THE SEARCH PANEL IN NOTEPAD++ (SCREENSHOT). ALSO VISIBLE ARE THE LISTING OF THE ROW NUMBERS AND THE SELECTION OF ROWS THROUGH BLUE BALLS.

Dirlister V2 beta 4 proved somewhat useful for generating the metadata file list, although its capabilities in offering the correct data columns were very limited. The reason for the limited column options is the recent first version of the applications in 2019. Since then, four new beta versions have been published and steadily increase the column options and export outputs. Unfortunately, the application does not completely adhere to archaeological or 3D modeling specific descriptions. While Dirlister offers the option to include *Mediatype* as a column, it can only include general mediatypes, such as images, audio, or video. In short, Dirlister is currently a less qualified application than contemporary proprietary freeware or paid software but can potentially offer great value in the future if the community remains supportive of the developments within the application.

Another tool, Calc of the LibreOffice V6.4.4.2 (64 bit), was required to finish the last conditions of the metadata file list. LibreOffice Calc performed appropriately and is a useful FOSS alternative to Microsoft Excel. The changes made of the file list during the workflow did not require much of the capabilities of the application, hereby indicating that many functionalities of LibreOffice Calc have not been fully tested or evaluated as a FOSS alternative to Microsoft Excel. The only few options that were required of Calc were to import the produced Dirlister CSV file and to change the column names and information. At least for those functions, LibreOffice Calc is a very adequate tool.

CloudCompare v2.11 beta, licensed under the GNU GPL, was used for accurately comparing and analyzing the 3D data deviations between formats. Although not incorporated in the workflow, CloudCompare was very useful for generating highly accurate comparisons between the original and exported formats of the archaeological 3D data. The results indicate that CloudCompare has much more accurate measuring techniques to assess the loss of data between 3D models than Blender. The visual analysis is also better and can be substantiated by the detectible distance in Blender between the viewport and the objects. This detectible distance in Blender prohibits the user to analyze object-based models up close, as displayed in figure 7.2. CloudCompare and other FOSS such as MeshLab, are therefore better for detailed visualizations of small dents or cracks, that have archaeological value. CloudCompare was useful to obtain highly accurate comparisons but proved sometimes troublesome during the alignment of 3D objects. CloudCompare can be used, just as Blender, as an additional tool for archaeological 3D modeling that attends the quality measurements between point-clouds and meshes of object-based 3D models.

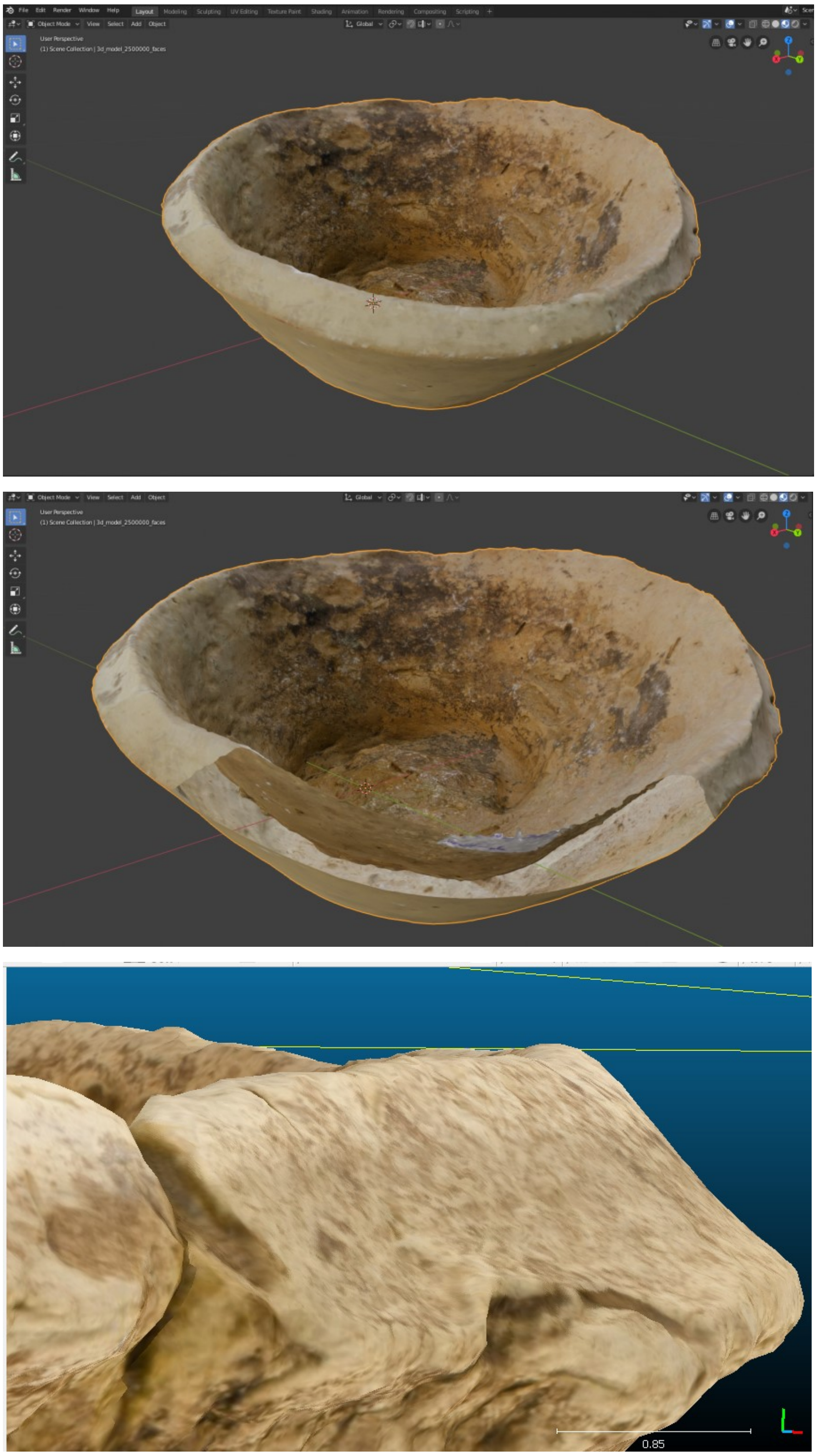


FIGURE 7.2: ZOOMING IN FROM A TO B RESULTS IN THE REMOVAL OF A CHUNK OF THE NEAREST RIM. THIS MEANS THAT THE TOP FIGURE DISPLAYS THE CLOSEST VISUALIZING DISTANCE THAT IS POSSIBLE IN BLENDER 2.8. THE BOTTOM FIGURE DISPLAYS THE ZOOM OPTION IN CLOUDCOMPARE, WHICH IS ABLE TO ZOOM IN MUCH CLOSER.

In general, FOSS applications provide ample support to generate an archive ready 3D dataset. The biggest advantage of using FOSS are the favorable licensing and transparency compared to paid and proprietary software (Vella 2018, 84). The free licensing and openness of the FOSS proved especially beneficial for the time spend on getting acquainted with 3D modeling. The transparency contributed immensely to the active user community for Blender and CloudCompare and provided useful recommendations during the handling of the applications. For subscription-based or fee-based applications, it can become aggravating when constant updates are required or application layouts change. And would have impacted the results of the thesis if Blender was a proprietary application. Blender updated its version during the making of this thesis, but still offered the, by now, older version of 2.80. Employing this older version ensured consistency for the outcome of the 3D import and export procedure.

Chapter 8: Conclusion

The main target of this thesis was to discuss and provide a constructive and clear procedure of what archaeological researchers should do with object-based archaeological 3D models before depositing them in digital archives. The procedure had to deal with considerations of the requirements of digital archives, the requirements of the users of the data, the legal and institutional requirements and had to fit in what is currently technically possible in archaeology.

To suffice to all these requirements, three additional issues were addressed as well. First, multiple 3D file formats were analyzed for their ability to adhere to the purpose of generating 3D models in archaeology. Secondly, the 3D dataset that was generated during the thesis was evaluated to the FAIR principles, which strive to bring clarity around the goals and urgencies of good data management and stewardship. And lastly, the usefulness of applying certain tools for preparing a 3D dataset was assessed, with a focus on the benefits of employing FOSS tools. The digital archive that was used was EASY, a certified digital archiving system in the Netherlands that assumes responsibility for long-term preservation of research data and accessibility of digital objects.

Based on the assessment of the different requirements, it can be concluded that a 3D dataset has to adhere to four steps to store the essential requirements for digitally preserving a dataset of object-based archaeological 3D models for the long-term in digital archives.

The first step addressed ethical considerations and fulfilled the requirements of the legal and institutional perspectives. These perspectives do not influence 3D data much, except if personal data is included within the (3D) dataset. Personal data must always be anonymized and if that is not possible, a processing agreement with DANS has to be conducted. In general, the legal and institutional perspectives had a lot of influence on the outcome of the thesis but have very few requirements that need to be conformed to.

The second step involved structuring the directories of the dataset and (re)naming the files to ensure human and computer readability. This step was considered essential for both the (future) user experience when exploring the dataset and the digital archive perspective. The directory structure must be ordered coherently across all and within all directories of a 3D dataset. The structure guides future users through the dataset. Filenames need be altered to comply to the directory structure and adherence to the ingest requests of the digital archive.

The third step involved the import and export of the 3D file format to 3D file formats that were fit for purpose and were simultaneously preservation friendly formats. Six 3D file formats that were recommended in EASY and implementable in Blender were used for analyses on time management, storage space, quality and consistency. The analyses demonstrated that, for the purposes of 3D models, archaeological researchers should consider using FBX format. However, for preserving 3D data, the combination of FBX and OBJ formats is more appropriate. The FBX format has the best overall attributes but has certain digital preservation limitations that can circumvented by also storing 3D files in the OBJ format. The formats together do require more storage space, which is the only drawback of combining the two formats. Other 3D file formats, such as PLY, BLEND, DAE and X3D are very useful in specific situations as well, but do not achieve the overall level of usefulness and versatility the combination of FBX and OBJ formats provide. Regardless of the optimal 3D file formats, exchange between different formats should be avoided as much as possible. Import and export of 3D models will always cause data loss and decrease the accuracy of 3D models.

The fourth and last step of the workflow involved three levels of metadata: the codebook, the file list and digital archive specific metadata. The three levels of metadata combined provide generous and adequate information for other researchers to understand and replicate the 3D data(set). The metadata and paradata does not overwhelm the reader and by using these three levels, the dataset adheres to the metadata standards of EASY as well.

The biggest obstacle that appeared during the assessment of the 3D dataset to the FAIR principles was the occurrence of the dichotomy between rich metadata and practical implementations. The described 'essential' elements of the 3D dataset did not fully cover the interoperability and reusability principles, because of the incomplete use of standard metadata schemas and vague terminology of rich metadata. What can be concluded from the assessment is that although metadata schemas were not fully adopted, the metadata in the dataset still incorporated elements that provide sufficient project information. And in the extreme situation that more information of the dataset is required, the metadata also provides relations to affiliated human and computer resources. It can be concluded that the complete 3D dataset mostly adheres to the FAIR principles and the elements which are not fully addressed do not inherently alter the outcome of the interoperability or reusability. The FAIR principles are guidelines that are deliberately broad and slightly vague. EASY ensures good findability and accessibility and provides enough information for 3D data to be interoperable and reusable, even if not all the elements of the principles are fully adhered to.

The four utilized applications for the workflow were all, to a certain extend, required to establish every step of the workflow. A 3D modeling software was indispensable for import and exporting 3D file formats and file list generators alleviate much of the effort and time consumption of the procedure. However, it would have been even more useful if these applications were slightly more extensive in the capabilities of altering texture file names and offering additional information columns of the file list. To accomplish these additional little although essential tasks, two additional applications were required. Thus, concluding that currently, four applications are essentially required to correctly execute the workflow.

It can be concluded that the licensing of FOSS for this thesis proved advantageous for the results and outcome of this thesis. The transparency of the FOSS tools is not addressed much. However, the financially free aspects of FOSS were highly beneficial for the affordability, which is a main limitation of museums and students. Blender, CloudCompare, Notepad++ and LibreOffice Calc proved to be useful 3D FOSS for the generation of preservation functionable 3D file formats and for accurately comparing 3D meshes. The only less useful FOSS was Dirlister. The application cannot yet be recommended for the purpose, but partially because of the application is still in its relative infancy. I expect that in the future this application will be more useful for generating metadata file lists. This thesis has demonstrated that, because of the utility of FOSS, it is possible to generate a preservation-ready object-based archaeological 3D dataset from 3D data, without funding.

While currently more and more alternatives for online storage of 3D models are possible, the thesis established that depositing in the digital archive EASY offers the best longevity for 3D data. EASY does not allow 3D models to be previewed, which is a useful function that digital libraries and online 3D journals have. However, these libraries and journals do not institutionally maintain all the 3D file formats and documentation in the way that EASY does and EASY does not have a commercial viewpoint. Longevity is a concept that is difficult to assess because the long-term purpose and use of 3D files are still somewhat uncertain. Therefore, the human and computer readability should be facilitated as much as possible.

The potential for future prospects of the research is abundant. While only object-based archaeological 3D models are attended, a comparison with local- and regional-based 3D models influences the choice of 3D file format considerably. Using archaeological 3D data that utilizes NURBS or CSG is another path of study that can be followed in the future. Although NURBS and CSG are not yet implemented in a similar scale as triangulation, I expect archaeologists to use these geometries in a larger scale in the relative near future. The limitations are mainly funding and 3D data acquisition. Once an affordable NURBS acquisition technique will be developed, I anticipate archaeologists to "hop on" quickly and find ways to extract even more accurate and additional information from 3D models. Another research potential is to select other 3D file formats that DANS does not recommend. These other 3D file formats can be assessed in 3D modeling features and internal structures, which can alter the dataset structure and the data outcome. There is a good possibility other 3D file formats will soon be added to the current recommended list of DANS.

This thesis has demonstrated a four-step workflow that provides for all the requirements of the different stakeholders. The workflow ensures that the essential elements of a 3D dataset are preserved, while only applying FOSS. This presented workflow indicates that it is currently possible to generate an object-based archaeological 3D dataset using FOSS. However, the process of the workflow can definitely be improved by developing a one-size-fits-all 3D file format for data preservation and by adapting 3D FOSS to be more archaeologically accurate. Dealing with these issues will establish the full potential of 3D modeling in future archaeology.

Abstract

The use of 3D models has steadily increased within archaeology, leading to the adoption of many 3D digital data acquisition techniques and 3D quality assessments. Yet the problem remains of how 3D models and 3D file formats can be opened 10 or 20+ years after they have been generated. 3D Data preservation for the long-term is a predicament that current archaeological digital archives are dealing with. Correct preservation has to consider the purpose of current 3D models and what 3D models can offer for future research aspects. Legal, institutional and technical aspects influence preservation as well and combined with the digital archive and user perspectives, form the stakeholders of 3D data preservation. The used 3D data of this thesis are photogrammetric-generated 3D models of Uruk pottery of Jebel Aruda. The focus of this thesis is to demonstrate how existing object-based archaeological 3D data should be converted and presented within a dataset to consider the requirements of all the stakeholders.

This thesis presents a workflow for generating a dataset of object-based archaeological 3D models for EASY. EASY is a certified digital archiving system in the Netherlands that assumes responsibility for long-term preservation of research data and accessibility of digital objects. The workflow only addresses stakeholder requirements that are essential for digitally preserving 3D data for the long-term. This workflow is assessed on its adherence to the FAIR principles, on the fit for purpose of 3D file formats for object-based archaeological 3D models and lastly, the required tools and applicability of Free and Open-Source Software (FOSS).
The workflow consists of four steps with the first step addressing ethical considerations and the second step directory structuring and file naming. The third step involves importing the original 3D file in Blender and exporting the file to six preservation recommended 3D file formats. The last step of the workflow is to generate three levels of documentation using metadata and paradata.

Three evident results emerged from this thesis. The research indicates that of the six recommended 3D file formats, the combination of the FBX and OBJ file formats provide the most value for preserving object-based archaeological 3D models. The research also suggests that the generated 3D dataset not completely adheres to the FAIR principles, but that elements which are not fully addressed do not inherently alter the outcome of the interoperability or reusability. Lastly, the thesis has shown that it is possible to generate a preservation-ready 3D dataset using only FOSS. Although archaeological and preservation specifications can be researched further, object-based archaeological 3D datasets can be generated in a cost-free and transparent production.

Samenvatting (Dutch)

3D modellen worden steeds vaker gebruikt in archeologie, wat ervoor gezorgd heeft dat veel digitale 3D acquisitie technieken en kwaliteitsmetingen door middel van 3D modellen mogelijk zijn. Helaas bestaat het probleem dat huidige 3D modellen en 3D bestandsformaten niet altijd meer te openen of bruikbaar zijn over 10 of 20+ jaar. 3D data preserveren voor de lange termijn is één van de ingewikkelde situaties waarmee huidige archeologische digitale archieven mee zitten te kampen. Correcte preservatie behoort rekening te houden met het doel van huidige 3D modellen en welke waarden huidige 3D gegevens in de toekomst kunnen hebben. Legale, institutionele en technische aspecten beïnvloeden het preserveerproces ook en gezamenlijk met de perspectieven en waarden van het digital archief en de gebruiker, vormen ze de belanghebbende actoren van 3D gegevens preservering. De gebruikte 3D gegevens in deze scriptie zijn gegeneerde 3D modellen door middel van fotogrammetrie, van Uruk potten uit Jebel Aruda. De focus van de scriptie is om te demonstreren hoe huidige object-gebaseerde archeologische 3D gegevens geconverteerd en gepresenteerd kunnen worden in een dataset om aan de voorwaarden van alle belanghebbenden te voldoen.

De scriptie presenteert een workflow voor het genereren van een dataset van object-gebaseerde archeologische 3D modellen in EASY. EASY is een gecertificeerd digitaal archiveringssysteem dat de verantwoordelijkheid draagt voor het op lange termijn behouden van onderzoeksgegevens en toegang tot digitale objecten. De workflow adresseert enkel de vereisten van de belanghebbenden die als essentieel beschouwd kunnen worden voor het digitaal behouden van 3D gegevens voor de lange termijn. De workflow is daarnaast beoordeeld op de aanhankelijkheid tot de FAIR principes, op het passen bij het doel van het genereren van 3D modellen en als laatste, de benodigde applicaties en toepasbaarheid van gratis en opensource software (FOSS).

De workflow bestaat uit vier stappen, waarbij de eerste stap ethische overwegingen adresseert en de tweede stap bestandsnamen en de directorystructuur. De derde stap bestaat uit het importeren van het originele 3D bestands in Blender en het exporteren naar zes preservatie aanbevolen 3D bestandsformaten. De laatste stap van de workflow bedraagt het genereren van drie niveaus van documentatie door middel van metadata en paradata.

Drie duidelijke onderzoeksresultaten kwamen naar voren tijdens de scriptie. Het onderzoek impliceert dat van de zes aanbevolen 3D bestandsformaten, de combinatie van de FBX- en OBJ-bestandsformaten de meeste waarde bieden voor het preserveren van object-gebaseerde archeologische 3D modellen. Ook suggereert het onderzoek dat de gegenereerde 3D dataset niet volledig voldoet aan FAIR principes, maar dat de ontbrekende elementen het resultaat van de interoperabiliteit en herbruikbaarheid niet inherent aanpassen. Als laatste laat de scriptie zien dat het mogelijk is om preservatie-gerede 3D datasets te generen enkel door middel van FOSS. Alhoewel archeologische en preservatie specificaties nog verder onderzocht moeten worden,

kunnen object-gebaseerde archeologische 3D datasets kosteloos en transparant genereerd worden.

Internet Sources

https://www.3dnatives.com/en/laser-3d-scanner-vs-structured-light-3d-scanner-080820194/, accessed on 1 May 2020.

http://3doc.i3dconverter.com/formats.html, accessed on 8 June 2020.

https://3dscanexpert.com/agisoft-photoscan-photogrammetry-3d-scanning-review/, accessed on 28 May 2020.

https://www.blender.org/about/, accessed on 20 April 2020.

https://www.blend4web.com/en/community/article/350/, accessed on 20 April 2020.

https://www.coretrustseal.org/why-certification/certified-repositories/, accessed on 1 April 2020.

https://creately.com/, accessed on 10 June 2020.

https://www.dans.knaw.nl, accessed on 1 April 2020.

https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing, accessed on 1 April 2020.

https://dans.knaw.nl/nl/actueel/nieuws/dans-maakt-open-access-mogelijk, accessed on 13 June 2020.

https://dans.knaw.nl/nl/over/organisatie-beleid/beleid-en-strategie/preservatie-plan-data-archiving-and-networked-services-dans, accessed on 2 March 2020.

https://docs.blender.org/manual/en/latest/render/shader_nodes/vector/bump.html, accessed on 1 May 2020.

https://www.easy.dans.knaw.nl, accessed on 10 April 2020.

https://github.com/notepad-plus-plus/notepad-plus-plus, accessed on 20 June 2020.

https://github.com/SanderSade/DirLister, accessed on 13 June 2020.

www.go-fair.org, accessed on 1 May 2020.

https://guides.archaeologydataservice.ac.uk/g2gp/3d_2-2, accessed on 7 January 2020.

https://guides.library.ucsc.edu/c.php?g=618773&p=4306386#:~:text=The%20Dublin%20Core%20Metadata%20Element,of%20physical%20and%20digital%20resources., accessed on 27 June 2020.

https://intarch.ac.uk/, accessed on 22 June 2020.

https://www.loc.gov/preservation/digital/formats/fdd/fdd000507.shtml, accessed on 17 February 2020.

https://www.loc.gov/preservation/digital/formats/fdd/fdd000491.shtml, accessed on 22 June 2020.

https://m.all3dp.com/3d-file-format-3d-files-3d-printer-3d-cad-vrml-stl-obj/, accessed on 8 January 2020.

https://news.artnet.com/exhibitions/microsoft-ar-vr-art-museum-1737159, accessed on 20 June 2020.

https://www.organisatiegids.universiteitleiden.nl/binaries/content/assets/geesteswetenschappen/commissies/flowchart-checklist-ethics-committee.pdf, accessed on 27 June 2020.

https://www.organisatiegids.universiteitleiden.nl/en/faculties-and-institutes/archaeology/committees/ethics-committee?_ga=2.225792026.1252079990.1582021957-1451730350.1497179607, accessed on 1 May 2020.

https://www.pngitem.com/middle/ibbTTwi_picture-vertex-edge-and-face-hd-png-download/, accessed on 21 June 2020.

https://pro.carare.eu/doku.php?id=support:metadata-schema, accessed on 20 April 2020.

https://www.reddit.com/r/blender/, accessed on 18 March 2020.

https://www.rmo.nl/collectie/collectiezoeker/collectiestuk/?object=228841, accessed on 30 June 2020.

https://www.rmo.nl/collectie/collectiezoeker/collectiestuk/?object=28205, accessed on 30 June 2020.

https://scholarworks.iu.edu/journals/index.php/sdh/index, accessed on 22 June 2020.

https://www.sttmedia.com/filelistcreator, accessed on 13 June 2020.

https://www.youtube.com/watch?v=_-WohCkKVF0 visited on 7 April 2020.

https://github.com/notepad-plus-plus/notepad-plus-plus, accessed on 20 June 2020.

Bibliography

Aboali, M., N.A. Manap, A.M. Darsono and Z.M. Yusof, 2017. Review on three-dimensional (3-D) acquisition and range imaging techniques. *International Journal of Applied Engineering Research* 12 (10), 2409-2421.

Akkermans, P.M.M.G., and G.M. Schwartz, 2003. *The archaeology of Syria: from complex hunter-gatherers to early urban societies (c. 16,000-300 BC)*. Cambridge: Cambridge University Press.

Algaze, G., 2008. *Ancient Mesopotamia at the dawn of civilization: the evolution of an urban landscape*. Chicago: University of Chicago Press.

Bakker, J.A., J. Kruk, A.E. Lanting and S. Milisauskas, 1999. The earliest evidence of wheeled vehicles in Europe and the Near East. *Antiquity* 73(282), 778-790. https://doi.org/10.1017/S0003598X00065522

Bell, T. and S. Zhang, 2015. Multiwavelength depth encoding method for 3D range geometry compression. *Applied Optics* 54 (36), 10684–10691. https://doi.org/10.1364/AO.54.010684

Bennett, M.J., 2015. Evaluating the Creation and Preservation Challenges of Photogrammetry-based 3D Models. *Archiving Conference* 1, 78-82. https://opencommons.uconn.edu/libr_pubs/52

Berchum M. van and M. GROOTVELD, 2016. Het beheren van onderzoeksdata, in A.S.M. De Jong, M. Koren, M. de Niet and E. Sieverts (eds), *Handboek Informatiewetenschap*, Alphen aan de Rijn: Vakmedianet (77), 1-28.

Betts, M.W., H. Maschner, C. Schou, R. Schlader, J. Holmes, N. Clement and M. Smuin, 2011. Virtual zooarchaeology: building a web-based reference collection of northern vertebrates for archaeofaunal research and education. *Journal of Archaeological Science*, 38 (4), 755-762. https://doi.org/10.1016/j.jas.2010.06.021

Boardman, C., and P. Bryan, 2018. *3D Laser Scanning for Heritage: Advice and Guidance on the Use of Laser Scanning in Archaeology and Architecture*. Swindon. Historic England.

Boeckhout, M., G.A. Zielhuis and A.L. Bredenoord, 2018. The FAIR guiding principles for data stewardship: fair enough?. *European Journal of Human Genetics* 26, 931–936. https://doi.org/10.1038/s41431-018-0160-0

Boehler, W. and A. Marbs, 2002. 3D Scanning Instruments, in W. Boehler (ed), *Proceedings of the CIP AWG 6 International Workshop on Scanning For Cultural Heritage Recording: September, 1-2, 2002*. Thessaloniki: Corfu, 9-12.

Brandsen, A., K. Lambers, S. Verberne and M. Wansleeben, 2019. User Requirement Solicitation for an Information Retrieval System Applied to Dutch Grey Literature in the Archaeology Domain. *Journal of Computer Applications*, 2(10), 21-30. https://doi.org/10.5334/jcaa.33

Buccellati, G., 1990. Salt at the dawn of history: The case of the beveled-rim bowls, in P. Matthiae, M. Van Loon and H. Weiss (eds), *Resurrecting the Past. A Joint Tribute to Adnan Bounni*, Istanbul: Nederlands Historisch-Archeologisch Instituut Te Istanbul, 17-40. https://doi.org/10.6084/m9.figshare.8282495

Buchele, S.F. and R.H. Crawford, 2003. Three-dimensional halfspace constructive solid geometry tree construction from implicit boundary representations, in SM '03: *Proceedings of the eighth ACM symposium on Solid modeling and applications*. New York: Association for Computing Machinery, 135-144. https://doi.org/10.1145/781606.781629

CCSDS Secretariat, 2012. *Reference Model for an Open Archival Information System (OAIS)*. Wasshington DC: Space Operations Mission Directorate.

Chadha, R.K., 2009. Digital preservation: strategies for Indian libraries, in P. Chand and A.K. Rai (eds), *7th international Convention on Automation of Libraries in Education and Research, February 25-27, 2009*. Pondicherry: Pondicherry University, 1-13.

Chandler, J.H., S. Buckley, M.B. Carpenter and C.M. Keane, 2016. Structure from motion (SFM) photogrammetry vs terrestrial laser scanning. *Geoscience handbook*. Alexandria: American Geosciences Institute, 1-4.

Corns, A., A. Deevy, G. Devlin, L. Kennedy and R. Shaw, 2015. 3D-ICONS: Digitizing Cultural Heritage Structures. *New Review of Information Networking: Shaping our Legacy: Preserving the Social and Cultural Record*, 20(1-2), 59–65. https://doi.org/10.1080/13614576.2015.1115232

Crawford, H., 1991. *Sumer and the Sumerians*. Cambridge: Cambridge University Press.

Crawford, H., 2015. *Ur : the city of the moon god*. London: Bloomsbury Academic.

DANS, 2011. *Archeologische metadata. Uitleg van het invullen van de metagegevens van archeologische datasets.* Handleiding DANS, Den Haag.

Decker, S. and J. Ford, 2017. Management of 3D Image Data in D. Errickson and T. Thompson (eds), *Human remains: Another dimension*. Cambridge: Academic Press, 175-184.

Dell'Unto, N., G. Landeschi, J. Apel and G. Poggi, 2017. 4D recording at the trowel's edge: Using three-dimensional simulation platforms to support field interpretation. *Journal of Archaeological Science: Reports* 12, 632–645. https://doi.org/10.1016/j.jasrep.2017.03.011

Diara, F., 2019. Cultural Heritage digital data: future and ethics, in H. Kremers (ed), *Digital Cultural Heritage*. Cham: Springer, 303-314. https://doi.org/10.1007/978-3-030-15200-0_20

Driel, G. van, 1977. De Uruk-Nederzetting op de Jebel Aruda. Een voorlopig bericht (stand eind 1976), *Phoenix* 23, 42-64.

Driel, G. van, 2002. Jebel Aruda: variations on a Late Uruk domestic theme, in J.N. Postgate (ed), *Artefacts of Complexity: Tracking the Uruk in the Near East*. Warminster: Aris and Phillips Ltd., 191-206.

Du, T., J.P. Inala, Y. Pu, A. Spielberg, A. Schulz, D. Rus, A. Solar-Lezama and W. Matusik, 2018. InverseCSG: Automatic conversion of 3D models to CSG trees. *ACM Transactions on Graphics (TOG)*, 37(6), 1-16. https://doi.org/10.1145/3272127.3275006

Ducke, B., 2012. Natives of a connected world: free and open source software in archaeology. *World Archaeology*, 44(4), 571-579. http://doi.org/10.1080/00438243.2012.743259

El Idrissi, B., 2019. Long-Term Digital Preservation: A Preliminary Study on Software and Format Obsolescence, in K. El Maghraoui and D. Chiadmi (eds), *Proceedings of the ArabWIC 6th Annual International Conference Research Track*. New York: Association for Computing Machinery, 1-6. https://doi.org/10.1145/3333165.3333178

Errickson, D., 2017. Shedding light on skeletal remains: The use of structured light scanning for 3D archiving, in D. Errickson and T. Thompson (eds), *Human remains: Another dimension*. Cambridge: Academic Press, 93-101.

Fielden, K., 1981. A Late Uruk Pottery Group from Tell Brak, 1978. *Iraq* 43 (2), 157-166. https://doi.org/10.2307/4200143

Fink, J. K., 2017. *Chemicals and Methods for Conservation and Restoration: Paintings, Textiles, Fossils, Wood, Stones, Metals, and Glass*. Hoboken: John Wiley & Sons.

Geng, J., 2011. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2), 128–160. https://doi.org/10.1364/AOP.3.000128

Goulder, J., 2010. Administrators' bread: an experiment-based re-assessment of the functional and cultural role of the Uruk bevel-rim bowl. *Antiquity* 84, 351-362. https://doi.org/10.1017/S0003598X0006662X

Guery, J., M. Hess and A. Mathys, 2017. Photogrammetry, in A. Bentkowska-Kafel and L. MacDonald (eds), *Digital Techniques for Documenting and Preserving Cultural Heritage*. York: Arc Humanities Press, 229-235.

Grimaud, V. and S. Cassen, 2019. Implementing a protocol for employing three-dimensional representations in archaeology (PETRA) for the documentation of neolithic funeral architecture in Western France. *Digital Applications in Archaeology and Cultural Heritage* 13, 1-18. https://doi.org/10.1016/j.daach.2019.e00096

Grussenmeyer, P., T. Landes, M. Doneus and J. Lermat, 2018. Basics of range-based modelling techniques in Cultural Heritage, in R. Eppich (ed), *3D Recording, Documentation and Management of Cultural Heritage*. Dunbeath: Whittles Publishing, 305-368.

Hamilakis Y. and P. Duke, 2007. *Archaeology and Capitalism: From Ethics to Politics, Taylor and Francis*. Walnut Creek: Routledge.

Hess, M., and A. Baik, 2017. 3D laser scanning, in A. Bentkowska-Kafel and L. MacDonald (eds), *Digital Techniques for Documenting and Preserving Cultural Heritage.* York: Arc Humanities Press, 199-206.

Horik, R. van, 2005. *Permanent pixels: Building blocks for the longevity of digital surrogates of historical photographs*. Den Haag: DANS. http://doi.org/10.26530/OAPEN_353253

Ikeuchi, K., and Y. Sato, 2001. Modeling from Reality, in D.C. Young (ed), *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. Quebec: Springer Science & Business Media, 117-124. https://doi.org/10.1109/IM.2001.924417

Jecić, S., and N. Drvar, 2003. The assessment of structured light and laser scanning methods in 3D shape measurements, in F. Matjiček (ed), *4th International Congress of Croatian Society of Mechanics, September 18-20, 2003*. Zagreb: University of Zagreb, 99-100.

Kansa, E. and S.W. Kansa, 2013. We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication. *Journal of Eastern Mediterranean Archaeology and Heritage Studies*. 1(1), 88-97. http://dx.doi.org/10.1353/ema.2013.0007

Kansa, S. W., L. Atici, E.C. Kansa and R.H. Meadow, 2020. Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data. *Advances in Archaeological Practice* 8(1), 40-52. https://doi.org/10.1017/aap.2019.36

Khalil, O. and P. Grussenmeyer, 2019. 2D & 3D Reconstruction Workflows From Archive Images, Case Study Of Damaged Monuments In Bosra Al-Sham City (Syria). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2-W15 (2), 55-62. https://doi.org/10.5194/isprs-archives-XLII-2-W15-55-2019

Khosrowshahi, F., 2017. Building Information Modelling (BIM) a Paradigm Shift in Construction, in M. Dastbaz, C. Gorse and A. Moncaster (eds), *Building Information Modelling, Building Performance, Design and Smart Construction*. Cham: Springer, 47-64.

Kuroczyński, P., O. Hauck and D. Dworak, 2016. 3D models on triple paths-new pathways for documenting and visualizing virtual reconstructions, in S. Münster, M. Pfarr-Harfst, P. Kuroczyński and M. Ioannides (eds), *3D Research Challenges in Cultural Heritage II*. Cham: Springer, 149-172.

Kuzminsky, S. and M. Gardiner, 2012. Three-dimensional laser scanning: potential uses for museum conservation and scientific research. *Journal of Archaeological Science*, 39 (8), 2744-2751. https://doi.org/10.1016/j.jas.2012.04.020

Lambers, K. and F. Remondino, 2008. Optical 3D measurement techniques in archaeology: recent developments and applications, in A. Posluschny, K. Lambers and I. Herzog (eds), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), April 2–6, 2007.* Bonn: Habelt Verlag, 27-35.

Lawrence, D., 2012. *Early Urbanism in the Northern Fertile Crescent: A Comparison of Regional Settlement Trajectories and Millennial Landscape Change*. Durham: Durham University.

Lee, K.H., O. Slattery, R. Lu, X. Tang and V. McCrary, 2002. The state of the art and practice in digital preservation. *Journal of research of the National institute of standards and technology*, 107 (1), 93-106. http://doi.org/10.6028/jres.107.010

Lee, C. A., 2010. Open archival information system (OAIS) reference model, in M.J. Bates and M.N. Maack (eds), *Encyclopedia of Library and Information Sciences 3rd Edition*. Boca Raton: CRC Press, 4020-4030. http://doi.org/10.1081/E-ELIS3-120044377

Luhmann, T., S. Robson, S. Kyle and J. Boehm, 2013. *Close-range photogrammetry and 3D imaging*. Berlin: Walter de Gruyter (De Gruyter Textbook 2nd Edition). https://doi.org/10.1515/9783110302783

M'Darhri, A., V. Baillet, B. Bourineau, A. Calantropio, G. Carpentiero, M. Chayani, L. de Luca, I. Dudek, B. Dutailly H. Gautier, E. Grilli, V. Grimaud, C. Hoffmann, A. Joffres, N. Jonçic, M. Jordan, J. Kimball, A. Manuel, P. Mcinerney, I. Muñoz Pandiella, A. Neroulidis, E. Nocerino, A. Pamart, C. Papadopoulos, M. Potenziani, E. Saubestre, R. Scopigno, D. Seillier, S. Tournon-Valiente, M. Trognitz, J. Vallet and C. Zuanni, 2019. *Share-Publish-Store-Preserve. Methodologies, Tools and Challenges for 3D Use in Social Sciences and Humanities*. Marseille: Parthenos.

Max, N.L. and B.G. Becker, 1994. Bump shading for volume textures. *IEEE Computer Graphics and Applications* 14 (4), 18–20. https://doi.org/10.1109/38.291525

McHenry, K. and P. Bajcsy, 2008. An overview of 3d data content, file formats and viewers. *National Center for Supercomputing Applications* 1205, 1-22.

McPherron, S., T. Gernat, and J. Hublin, 2009. Structured light scanning for high-resolution documentation of in situ archaeological finds. *Journal of Archaeological Science* 36 (2009), 19-24. https://doi.org/10.1016/j.jas.2008.06.028.

Michelson, A. and J. Rothenberg, 1992. Scholarly communication and information technology: exploring the impact of changes in the research process on archives. *The American Archivist* 55 (2), 236-315. https://doi.org/10.17723/aarc.55.2.52274215u65j75pg

Mons, B., C. Neylon, J. Velterop, M. Dumontier, L.O.B. da Silva Santos, and M.D. Wilkinson, 2017. Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services and Use* 37 (1), 49-56. http://doi.org/10.3233/ISU-170824

Münster, S., P. Kuroczyński, M. Pfarr-Harfst, M. Grellert and D. Lengyel, 2015. Research Challenges for a Computer-Based Interpretative 3D Reconstruction of Cultural Heritage - A German Community's View, in Y. Yen, K. Weng, and H. Cheng (eds), *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 25th International CIPA Symposium 2015*. Taipei: Taiwan, 207-213. https://doi.org/10.5194/isprsannals-II-5-W3-207-2015

Niven, K. and J. Richards, 2017. The storage and long-term preservation of 3D data, in D. Errickson and T. Thompson (eds), *Human remains: Another dimension*. Cambridge: Academic Press, 175-184.

Niven, L., T. Steele, H. Finke, T. Gernat and J.J. Hublin, 2009. Virtual skeletons: using a structured light scanner to create a 3D faunal comparative collection. *Journal of Archaeological Science*, *36* (9), 2018-2023. https://doi.org/10.1016/j.jas.2009.05.021.

Oates, J., 1985. Tell Brak: Uruk Pottery from the 1984 Season. *Iraq* 47, 175-186. http://doi.org/10.2307/4200238

Olsen, B., R. Placchetti, J. Quartermaine and A. Killebrew, 2013. The Tell Akko Total Archaeology Project (Akko, Israel): Assessing the suitability of multi-scale 3D field recording in archaeology. *Journal of Field Archaeology* 38 (3), 244-262. http://doi.org/10.1179/0093469013Z.00000000056

Pearson D. and C. Webb, 2008. Defining File Format Obsolescence: A Risky Journey. *International Journal of Digital Curation* 3 (1), 89-106. https://doi.org/10.2218/ijdc.v3i1.44

Pfarr-Harfst, M., 2016. Typical workflows, documentation approaches and principles of 3D digital reconstruction of cultural heritage, in S. Münster, M. Pfarr-Harfst, P. Kuroczyński and M. Ioannides (eds), *3D Research Challenges in Cultural Heritage II*. Cham: Springer, 32-46.

Pletinckx, D., 2012. How to make sustainable visualizations of the past: an EPOCH common infrastructure tool for interpretation management, in A. Bentkowska-Kafel and H. Denard (eds), *Paradata and transparency in virtual heritage*. London: Routledge, 223-232.

Rothman, M. S., 2002. Tepe Gawra: chronology and socio-economic change in the foothills of Northern Iraq, in J.N. Postgate (ed), *Artefacts of Complexity: Tracking the Uruk in the Near East*. Warminster: Aris and Phillips Ltd., 49-78.

Pejšová, P, and M. Vaska, 2014. *Free Licenses & Creative Commons: A Powerful Tool for Open Access Publishing in Grey Literature 16*, 1-16. https://doi.org/10.17026/dans-zss-naep

Potts, D., 2009. Bevel-rim bowls and bakeries: evidence and explanations from Iran and the Indo-Iranian borderlands. *Journal of Cuneiform Studies* 61 (1), 1-23. https://doi.org/10.1086/JCS25608631

Poux, F., R. Neuville, L. Van Wersch, G.A. Nys and R. Billen, 2017. 3D Point Clouds in Archaeology: Advances in Acquisition, Processing and Knowledge Integration Applied to Quasi-Planar Objects. *Geosciences* 7 (96), 1-32. https://doi.org/10.3390/geosciences7040096

Rabin, S. (Ed), 2009. *Introduction to game development*. Scarborough: Nelson Education.

Remondino, F. and S. El-Hakim, 2006. Image-based 3D modelling: a review. *The photogrammetric record* 21 (115), 269-291. https://doi.org/10.1111/j.1477-9730.2006.00383.x

Reu, de J., G. Plets, G. Verhoeven, P. De Smedt, M. Bats, B. Cherretté, W. De Maeyer, J. Deconynck, D. Herremans, P. Laloo, M. Van Meirvenne and W. De Clercq, 2013. Towards a three-dimensional cost-effective registration of the archaeological heritage. *Journal of Archaeological Science* 40 (2), 1108-1121. https://doi.org/10.1016/j.jas.2012.08.040

Revello Lami, M., L, Opgenhaffen, and I. Kisjes, 2016. Pottery goes digital: 3D laser scanning technology and the study of archaeological ceramics, in S. Campana, R. Scopigno, G. Carpentiero and M. Cirillo (eds), *Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology.* Oxford: Archaeopress, 421-430.

Rieke-Zapp, D. and S. Royo, 2017. Structured light 3D scanning, in A. Bentkowska-Kafel and L. MacDonald (eds), *Digital Techniques for Documenting and Preserving Cultural Heritage*. York: Arc Humanities Press, 245-249.

Robson, S., S. MacDonald, G. Were and M. Hess, 2012. 3D recording and museums, in C. Warwick, M.M. Terras, J. Nyhan (eds), *Digital humanities in practice*. London: Facet Publishing, 91-116.

Rothenberg, J., 1998. Ensuring the longevity of digital information. *International Journal of Legal Information* 26 (1-3), 1-22. https://doi.org/10.1017/S0731126500000469

Rothenberg, J., 1999. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington D.C.: Council on Library and Information Resources.

Rothman, M. S. and B. Peasnall, 2002. *Tepe Gawra: the evolution of a small, prehistoric center in northern Iraq*. Pennsylvania: University of Pennsylvania (Museum of Archaeology).

Sanjurjo, J. and J.L. Fenollós, 2012. Restudying the Beveled Rim Bowls: new preliminary data from two Uruk sites in the Syrian Middle Euphrates, in F. Caramelo, M.L. Paiva Monte and D. Paiva (eds), *Res Antiquitatis: Journal of Ancient History* 3, Braga: Estudio de Artes Graficas, 263-277.

Shapiro, V. and D.L. Vossler, 1991. Construction and optimization of CSG representations. *Computer-Aided Design* 23 (11), 4-20. https://doi.org/10.1016/0010-4485(91)90077-A

Shimray, S.R. and C.K. Ramaiah, 2018. Digital Preservation Strategies: An Overview, in S. Rajeshwari, E. Soundararajan, V. Sundararajan, V. Rajendran (eds), *11th National Conference on*

Recent Advances in Information Technology (READIT) August 8-9 2018, Kalpakam: Tamilnadu, 46-54.

Tjalsma, H., 2006. *Archiveren van Digitaal Academisch Erfgoed, een verslag als voorbeeld*. Den Haag: DANS.

Tsirliganis, N., G. Pavlidis, A. Koutsoudis, D. Papadopoulou, A. Tsompanopoulos, K. Stavroglou, Z. Loukou, and C. Chamzas, 2002. Archiving 3D cultural objects with surface point-wise database information, in B. Werner (ed), *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*. Piscataway: The Institute of Electrical and Electronics Engineers, 766-769. https://doi.org/10.1109/TDPVT.2002.1024032

Tsoupra E., P. Doorn and E. Fankhauser, 2018. *Evaluation of DANS EASY Repository Based on the FAIR Principles*. Den Haag: DANS. http://doi.org/10.17026/dans-2au-j2z6

Turk, G., 1994. *The PLY Polygon File Format*. Stanford: The Leland Stanford Junior University.

Vatanen, I., 2003. Deconstructing the (Re)Constructed: Issues in Conceptualising the Annotation of Archaeological Virtual Realities, in M. Doerr and S. Apostolis (eds), *Proceedings of the conference on computer applications and quantitive methods in archaeology, Crete, 30th April, 2002,* Athene: Hellenic Ministry of culture, 69-74.

Vella, E., 2018. *Foss Forward: Using Open Data and Free and Open Source Software to Document the Terraces in the Lower Engadine, Switzerland*. Leiden: Universiteit Leiden.

Waugh, A., R. Wilkinson, B. Hills and J. Dell'Oro, 2000. Preserving digital information forever. In R. Furuta, K.M. Anderson, D.L. Hicks and P. Nürnberg (eds), *DL '00 Proceedings of the fifth ACM conference on Digital libraries*, New York: Association for Computing Machinery, 175-184. https://doi.org/10.1145/336597.336659

Wilkinson, M., M. Dumontier, I. Aalsberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmuns, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. gray, P. Growth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swrtz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Nature* 3, 1-9. https://doi.org/10.1038/sdata.2016.18

Wu, Q., K. Xu and J. Wang, 2018. Constructing 3D CSG Models from 3D Raw Point Clouds. *Computer Graphics Forum* 37 (5), 221–232. https://doi.org/10.1111/cgf.13504

List of figures, tables, appendices:

Figures

Tables

Appendices