ABSTRACT

The dataset consists of 3D scans of one cuneiform tablet from Haft Tappeh Iran and one cuneiform tablet of the Hilprecht Collection as well as 3D annotations on these 3D meshes, including metadata. The 3D annotations were created with the annotation software Annotorious on 2D renderings and reprojected to the original 3D model. The respective 2D renderings and annotations in 2D are also part of this data publication. The annotations might be used in machine learning tasks for character recognition, linguistic studies, or visualization in Assyriology. We publish these data in different formats and give guidance on how to use them in different usage scenarios and with several software applications. The data serve as the basis for a detailed description, reasoning, and elaboration of a recommendation for the state-of-the-art handling of 3D data in cuneiform research. The data is stored as an archive on Zenodo and may serve as an example for replication by similar 3D scanning.
(1) OVERVIEW

CONTEXT
The cuneiform script has been the dominating script in the Ancient Near East. Its use has been attested from approximately the 31st century BC to the second century AD primarily in the area of ancient Mesopotamia and has been primarily written on clay.

The dataset contains two different representations of cuneiform clay tablets from two different periods of time with different archaeological backgrounds.

The cuneiform tablet HT 07-31-95 has been excavated in Haft Tappeh, Iran. Haft Tappeh, the ancient city of Kabnak is located about 15km southeast of the ancient city of Susa and is considered a historic center between Mesopotamia and the Iranian plateau. The object is part of a collection of 800 clay tablets consisting mainly of administrative texts written in the Akkadian language and labeled objects dated to the Middle Assyrian period. The contents of these texts can be described as lists of materials, lexical lists and scholarly texts. The collection has been found in the so-called “Terrace Complex I and II” complex excavated by Ezatollah Negahban between 1965 and 1978. The excavation is documented in [1] and has been continued by Mofidi-Nasrabadi since 2005 [2]. About 500 cuneiform tablets out of this latest excavation have been 3D scanned and are published in 2022 in the research project “The Digital Edition of texts from Haft Tappeh”.

The second cuneiform tablet HS 1174 is part of the Hilprecht Collection of Babylonian Antiquities [3]. This collection includes a variety of cuneiform and further objects gathered from expeditions of Hermann Volrath Hilprecht with additions from further contributors. The archaeological provenience and context of this cuneiform tablet is not known, but we know of its administrative content. It is written in the Sumerian language and has been created during the Ur III period. The Hilprecht Collection has been 3D-scanned two times, the last time in the year 2009. In 2019, all scanned cuneiform tablets of the Hilprecht Collection have been oriented, filtered, texturized and 3D renderings have been generated in the HeiCuBeDa dataset.

Hence, our data consists of two typical administrative cuneiform text tablets from different areas in time which can serve as a showcase of how 3D annotations might be realised on a typical cuneiform tablet 3D scan.

Spatial coverage
The dataset contains 3D scans of cuneiform tablets from an excavation site in Haft Tappeh (Iran) and of the Hilprecht Collection, which were scanned at the following places, respectively:

Frau Professor Hilprecht Collection of Babylonian Antiquities
Friedrich-Schiller Universität Jena

Fürstengraben 6, 07743 Jena, Thuringia, Germany
Northern Boundary: 50.9300953
Southern Boundary: 50.9302309
Eastern Boundary: 11.5893116
Western Boundary: 11.589562

Haft Tappeh and Choghazanbil Museum
Susa-Shushtar Road, Haft Tappeh, Khuzestan Province, Iran
Northern Boundary: 32.0800798
Southern Boundary: 32.0804943
Eastern Boundary: 48.3296285
Western Boundary: 48.3301068

Temporal coverage
The concerned dataset reflects cuneiform tablets from the Middle Assyrian time period 1400-1000BC and the Ur III period 2112BC-2004BC, as defined in the Middle Chronology, in the case of the cuneiform tablet from the Hilprecht Collection.

(2) METHODS

The dataset adheres to a catalogue of criteria for the publication of 3D models in Assyriology which we elaborate on in the following description of the creation steps.

[4] provides a comprehensive overview on the state of the art in Visual Cuneiform Analysis that is enabled by a dataset such as the one provided here. If possible, the creation history of the given data should be reflected in the metadata itself, as outlined in [5]. When creating the data we followed the steps outlined in the following:

STEPS
At first, the original cuneiform tablets were scanned using two different structured light scanners and their accompanying scanning software.

The cuneiform tablets of Haft Tappeh have been scanned with a Range Vision 3D Scanner without the ability to capture colour in 2018 in the context of a campaign in the museum funded by the Johannes Gutenberg-University Mainz. The scans were conducted by Ali Zalaghi.

The 3D scans for the Hilprecht Collection have been acquired by a smartSCAN 3D-HE structured light scanner with the option to capture colour. A detailed report on how the measurements were conducted is referenced in [6].

Secondly, the 3D models were saved in temporary storage format. For the Haft Tappeh scans, this format was STL [7]. For the Hilprecht Collection scans, this format
has been PLY [8]. PLY, in contrast to STL, may preserve colour information.

The 3D models were then processed in the following steps:

- Automatic Mesh Polishing
- Cleaning (Erosion)
- Filling (Dilation)
- Manual orientation for the establishment of an object coordinate system by experts (front side parallel to X-Y axis)
- Export in the PLY format (only Haft Tappeh)

These exported PLY files are the first data product we deliver in this data publication.

In addition, metadata about the meshes are exported in linked data formats such as Resource Description Framework (RDF) [9] Files in the Terse RDF Triple Language (TTL) [10] and TXT files.

In the third step, renderings have been calculated from the 3D models in a process adapted and/or followed from [11] and [12] and with the use of the GigaMesh Software Framework [13], which is described as a video tutorial in [14].

This rendering process results in six images per 3D scan that describe the front, back, top, bottom, left, and right sides of the cuneiform tablet.

In a fourth step (cf. Figure 1), 2D annotations have been created using the JavaScript library Annotorious on the front renderings of the cuneiform tablets. The annotations describe cuneiform characters with their respective position on the cuneiform tablet (i.e., line, character index) and its transliteration and in the case of tablet HS 1174 also single wedges according to the Gottstein system [15, 16] (cf. Figure 2). The Gottstein system describes cuneiform signs according to their types and amount of included cuneiform wedges, whereas wedge type a represents the vertical wedge, b the horizontal wedge and c and d the diagonal wedges. In addition we have described the special Winkelhaken wedge as wedge type w. This wedge type is usually classified as wedge type c in the traditional Gottstein encoding. The annotations in 2D have been saved as JSON-LD [17] using the W3C Web Annotation Data Model [18] and have also been cropped as 2D JPG images [19], with appropriate metadata in XMP [20].

The fifth step involves a transformation of annotations in 2D to annotations in 3D (cf. Figure 2, 3 and 4) in the coordinate system of the 3D mesh. These annotations are represented in three distinct ways:

- As a sidecar file to a 3D model (JSON-LD) using an extended web annotation data model with Well-Known Text [21] annotation selectors
- As single 3D models, one 3D model per annotation (PLY)
- Within other 3D model formats (COLLADA [22], X3D [23]) or as label components within PLY (only displayable within GigaMesh)

Figure 1 2D annotations of cuneiform signs and cuneiform wedges on the front rendering of cuneiform tablet HS 1174.
Figure 2 Transformed cuneiform wedge annotations to 3D with a machine-detected color code according to the Gottstein system [15, 16]. The color coding may not in all cases be correct.

Figure 3 3D sign annotations on cuneiform tablet HT 07-31-95 as labelings of the 3D model.
Finally, metadata about the 3D model, the 3D renderings, and the annotations have been exported.

- Metadata of the 3D scan as TTL
- Metadata of renderings as TTL
- Metadata about the representation of annotations as part of the image

**SAMPLING STRATEGY**

We chose two 3D scans of cuneiform tablets, whereas one 3D scan has been scanned for the project “Digital Edition of Cuneiform Tablets in Haft Tappeh (Iran)” funded by the German Research Foundation (DFG), representing texts from the Middle Assyrian time period written in the Akkadian language. The second cuneiform tablet belongs to the Hilprecht Collection, is written in the Sumerian language, and represents the Ur III period. With this selection, we show two practical examples of 3D scans for two time periods and a proof of concept for annotations on both of them.

**QUALITY CONTROL**

We established a quality control process for the publication of 3D models in the Haft Tappeh project, which included the manual curation of data by experts in the project. The alignment of 3D meshes has been checked once when the mesh was created and in the following annotation process. The annotation of cuneiform signs has been verified with a comparison to the respective transliteration and with consultation by experts. The annotation of wedges has been verified by referring to the respective line art of the cuneiform tablet and a manual check by experts. In addition, we followed the quality assurance process for mesh cleanings exemplified in [12] on page 120, which is also described in the section Steps previously. This process guarantees that no zero-areas are present in the 3D models and that holes in the 3D models are filled.

The cuneiform tablets published in the HeiCuBeDa Hilprecht collection have been quality-assured in a process described in [11], that is, they followed the same mesh cleaning process and the same image rendering process that was applied to the Haft Tappeh tablets.

**CONSTRAINTS**

The constraints in data products can be seen in the very nature of its data acquisition:

- Structured light scanning produces shading effects at broken areas of the cuneiform tablet, which are usually mitigated in the cleaning process.
- Structured light scanning methods produce some of the following potential errors, which are mitigated before publication. For example: Holes in meshes, zero-area faces. Ideally all these problems should be corrected in the mitigation process elaborated on previously.
- Colour information is missing on the cuneiform 3D scan from Haft Tappeh due to scanner limitations, and the colours of the Hilprecht Collection scans are uncalibrated.
- Due to the nature of structured light scanning technologies, some surface properties like reflectance are not acquired in the scanning process.
After scanning, we may observe constraints in the annotation process.

As annotations are created manually with human effort, mistakes may happen in the precision of the annotations (annotations may not completely encompass the area they should encompass) and in the annotation content (e.g., wrong line or character indices may be annotated).

The mitigation of annotation errors has been attempted with manual checks by experts.

(3) DATASET DESCRIPTION

OBJECT NAME
Annotated 3D-Models of cuneiform tablets

DATA TYPE
Processed data (3d meshes) and interpretation of data (annotations)

FORMAT NAMES AND VERSIONS
Polygon File Format (PLY) 1.0 [8], Terse RDF Triple Language (TTL) 1.1 [10], Textfile in UTF-8 encoding (TXT), Extensible 3D Format (X3D) 3.3, Collaborative Design Activity Format (COLLADA) 1.5, JSON-LD 1.1 [17], JPEG (JPG) 1.0

CREATION DATES
Between 2014 and 2019: 3D Scans of the cuneiform tablet HS 1174 were created in Jena, Germany (Publication in 2019)
2018: 3D Scans of the cuneiform tablet HT 07-31-95 were created in Haft Tappeh Iran
2019: Creation of 2D renderings from HS 1174 in the course of the publication of HeiCuBeDa
Autumn 2021: Creation of 2D annotations on HT 07-31-95 and HS 1174 and creation of 2D renderings from HT 07-31-95
February 2022: Creation of 3D annotations and data exports for both cuneiform tablets

DATASET CREATORS
Ali Zalaghi – Created the 3D scans of the Haft Tappeh Project in STL format.
Tim Brandes, University Mainz – Created the orientation of the Haft Tappeh tablet 3D scan and, in this process, created aligned PLY files as an export.
Hubert Mara, University Halle – Created the 3D renderings and metadata of all cuneiform 3D scans.
Timo Homburg, Mainz University Of Applied Sciences – Created annotations on 2D renderings of the 3D scans
Robert Zwick, Mainz University Of Applied Sciences – Created 3D annotations and data exports in all available formats

Eva-Maria Huber, Marc Alexander Weber, Laura Krimmel, Jan Eric Tärnhuvud, Franziska Lutz, Mainz University assisted in the data curation process
Doris Prechel, University Mainz, Kai-Christian Bruhn, Mainz University Of Applied Sciences – Acquired funds for the 3D scanning project in Haft Tappeh
Kai-Christian Bruhn, Mainz University Of Applied Sciences – Coordinated research data management for this data publication

LANGUAGE
English in metadata, Akkadian, and Sumerian on the given image data.

LICENSE
CC-BY SA 4.0

REPOSITORY LOCATION
The data is published on Zenodo under the following DOI: https://doi.org/10.5281/zenodo.6506560
The cuneiform tablet scan HT 07-31-95 is published at the University of Heidelberg library using the following DOI: https://doi.org/10.11588/heidicon/1611840
The cuneiform tablet scan HS 1174 is published at the University of Heidelberg library using the following DOI: https://doi.org/10.11588/heidicon/1113605

PUBLICATION DATE

(4) REUSE POTENTIAL

We see the potential for this dataset to be reused in the domain of OCR, in particular cuneiform character recognition. The dataset provides the basis for annotating cuneiform characters in 3D and 2D in interoperable formats. It, therefore, shows how annotations on 3D scans of cuneiform tablets can be achieved with current technology. Further, we seek to contribute a good practice example for studies of Assyriology to engage researchers and students about the potential of usage of annotations in 2D and 3D, in research communication, and their use in online repositories. Only well-prepared annotation data allows for a convincing, comprehensive visual representation and, at the same time, accessible data and the possibility to query them according to linked data standards. The data can also be used to transfer the approach and results to other annotation projects on different objects of the digital cultural heritage domain. Finally, we hope that this dataset could encourage the support of annotations in 3D viewing software, contribute to the standardisation efforts of 3D web annotations, their vocabularies for cuneiform studies, and raise awareness for this technology and usage in the field.
NOTES
&task=showDetail&id=424957759.

ACKNOWLEDGEMENTS
We thank the Cuneiform Digital Library Initiative (CDLI)1 and the digital cuneiform community for fruitful discussions and feedback on our approach. In addition, we would like to thank the NFDI4Culture Plenary Session,10 which allowed our approach to be presented, and for the valuable feedback, we got in its aftermath.

FUNDING INFORMATION
Deutsche Forschungsgemeinschaft (DFG) Project Number: 424957759.

COMPETING INTERESTS
The authors have no competing interests to declare.

AUTHOR AFFILIATIONS
Timo Homburg orcid.org/0000-0002-9499-5840
Mainz University Of Applied Sciences, DE
Robert Zwick orcid.org/0000-0002-5839-8192
Mainz University Of Applied Sciences, DE
Hubert Mara orcid.org/0000-0002-2004-4153
Martin Luther University of Halle-Wittenberg, DE
Kai-Christian Bruhn orcid.org/0000-0001-8322-1260
Mainz University Of Applied Sciences, DE

REFERENCES