

The significance of archaeological source data

Pat Tanner

Abstract: All archaeology is contingent on the source data and everything stems from that. Achieving a detailed understanding of both the object and the processes is predicated on the quality of the underlying source data. Past reconstruction projects may have used the source data, but it is not always possible to understand or interrogate how that source data has been used or interpreted, due to the lack of published paradata. This paper examines the development of methods employed in the surveying of an archaeological excavation site, in essence the capture and recording of the raw primary data from the site or artefact.

Keywords: Source-data, Metadata, Paradata, Digital Documentation

Introduction

The main theme of ISBSA 16, 'Reading the Past – Imagining the Future', is to reflect on the results of traditional methods and tools while considering new possibilities provided through the implementation of modern technology. From experience gained on projects such as the Newport Medieval Ship, the Bremen Cog re-analysis, the Poole Log-boat analysis, and the digital reconstruction of the Anglo-Saxon Sutton Hoo ship burial, it is clear that all archaeology is contingent on the source data, and everything produced stems from that source data. Source data by its nature can be difficult to understand, and as noted by Hocker (2004: 2): "the study of ship remains begins with the recording of seemingly trivial details, the thickness of planks, the number and size of nail, the direction of an adze stroke".

The examining of the minutiae of trivial details to understand the workforce involved in building the *Vasa* (Hocker 2013) is a good example of the great level of detail it is possible to achieve. However, achieving this level of detailed understanding of both the object and the processes involved is predicated on the quality and detail of the underlying source data.

This paper examines the development of methods employed in the surveying of an archaeological excavation site, in essence the capture and recording of the raw primary data from the site or artefact. Reconstruction is not an end of its own, but a logical continuation based on appropriate source data in which we have confidence. Past reconstruction projects may have used the source data, but it is not always possible to understand or interrogate how that source data has been used or interpreted, due to the lack of published paradata (Denard 2009: 13), and the fact that the raw source data is rarely published in a form devoid of interpretation. George Bass points out in his introduction to the *Oxford Handbook of Maritime Archaeology* (2011: 10–11) that archaeologists publish only a fraction of the sites they investigate and noted that it can take years or decades to produce excavation reports that are more than simple catalogues. This practise inhibits the continuous flow of information and slows down the collective research efforts to reconstruct the past. Primary data is often kept away from peer review even after final publication of an excavation.

As noted by Crumlin-Pedersen (Crumlin-Pedersen 1977: 165), the traditional method of manual scale drawing is considered discontinuous as it is based on measuring a number of points on the object and completing the remaining drawing by interpolating between the measured points. In addition, the selection of what is actually represented, or omitted, on a drawing is a further interpretation by the draftsman. Hocker (2000; 2003) noted that representing complex three-dimensional shapes on two-dimensional paper led to the development of a set of drawing conventions, in which the object is represented as a series of intersecting views, traditional 'top down' or plan, side, and front orthographic views typically perpendicular to each other. These two-dimensional paper drawings are a less than ideal medium for the archival storage of complex three-dimensional data sets. Table 1 illustrates a sample of how the process of documentation, reassembly, reconstruction, and validation of the resulting hypothesis has developed over the past two centuries.

Clearly, the technology used to record site data is not a secondary concern but is central to the activity of site archaeology. We have an obligation to bridge the gap between the exclusive knowledge of the excavator and the published record, utilising a mode of data capture and record that is devoid of interpretation, or where interpretation is inevitable, paradata is used to explain the human processes of interpretation and understanding of the data objects.

Project	Date	Site Survey	Timbers Recorded	As Found	Initial basis for Hull Form	Development of Reconstruction	Validation of Reconstruction
Rother Barge	1822	Detailed site sketch	Some scantlings recorded		Reassembly for Display	Details of artefacts recovered	
Nydam	1859	Traditional 2D offsets survey	Detailed sketches and scale dwgs.		Reassembly for Display	Detailed sketches of artefacts recovered Subsequently re-excavated	full-scale Replica
Gokstad	1880	Traditional 2D offsets survey	2D Offsets		Reassembly for Display	2D scale dwgs. Full-scale Replica	Sinking of the full-scale Replica
Oseberg	1904	Traditional 2D offsets survey	2D Offsets		Reassembly for Display	Full Scale Replica based on measurements from the Museum display	
Woolwich	1912	Traditional 2D offsets survey		Survey Drawing	Name Identification	Name Identification	
Kalmar	1932	Traditional 2D offsets survey	2D Offsets	Survey Drawing	Scale Drawings	Scale Model	Scale Model
Ferriby	1937	Traditional 2D offsets survey	2D Offsets	Interpreted excavation dwg.	Scale Drawings	Scale Drawings and 5 Scale Models	Form Coefficients + displacement and scale replica
Yassi Ada 7th C	1961	Underwater Photography		Corrected and scaled from photographs	Reassembly of scaled model strakes	Scale Models	Tonnage
Skuldelev	1962	Photogrammetry Then scale dwg.	Full-scale Elevated plane tracing	Interpreted Torso dwg.	Reassembly of 2D scaled model strakes	Scale Model	Form Coefficients + displacement and full-scale replica
Kyrenia	1968	Underwater Photography	Drawn full-scale Photographed	Corrected and scaled from photographs	Initial hull form model	18 models – some full-scale	Tonnage and 2 full-scale replicas
Graveney	1970	Traditional 2D offsets survey and full-scale plaster cast	Full-scale Contact tracing	Reduced scale dwgs.	Reduced scale drawings	Scale Drawings and 4 Scale Models	Form Coefficients + displacement and scale replica
Serçe Limani	1977	Underwater Photography	Full-scale Elevated plane tracing	Site Diorama model	Site Diorama model and mould + batten model	Various models	Tonnage
Ma'agan Mikhael	1985	Underwater 2D offsets survey and photographs	2D Offsets	Survey Drawing	Surviving frame shape and reassembly of hull remains	reassembly of hull Remains and 3 scale models	Form Coefficients + displacement causing hull redesign and full-scale replica
Barland's Farm	1993	Traditional 2D offsets survey	2D Offsets	Interpreted 'as-found' dwg.	Reassembly of scaled model strakes	Scale Model	Form Coefficients + displacement
Roskilde	2002	Photogrammetry Then scale dwg.	3D full-scale digitising	Interpreted Torso dwg.	Flatten data to 2D and assemble scaled model strakes	Scale Model	Form Coefficients + displacement and full-scale replica
Newport	2002	Traditional 2D offsets survey and Photogrammetry	3D full-scale digitising	As surveyed and post-deposition model	3D Post deposition model	Full-scale 3D digital model	Full Orca3D hydrostatic and hydrodynamic analysis
Surenga 7	2007	Traditional 2D offsets survey	3D full-scale digitising	Traced from on-site digital scan	Surviving frame shape 3D printed and reassembly of hull remains	Scale Model	Orca3D hydrostatic analysis

Table 1 Summary of documentation and reconstruction approach for archaeological reconstructions

Recording source data

Traditional Surveying

The quality and detail of the source data is directly proportional to the accuracy of the measurements recorded, as well as the level of interpretation employed in that recording and the completeness of the elements recorded. As can be seen from Table 1, the recording of source data has developed from the 19th century two-dimensional site sketches, which were interpretive by their nature, to the 20th century traditional ‘baseline and offset’ survey technique, which involved a degree of interpretation in the reading of measurements, and more significantly, interpretation in the selection of which elements were recorded, often due to either time or economic constraints. The manual surveying of complex sites or objects can be a very time-consuming process and presents challenges in terms of subjectivity and accuracy (Holt 2003), and as noted by Baltsavias (1999: 84), measurement without interpretation is sometimes very difficult or impossible.

Tacheometric surveying

Tacheometric surveying, such as with a theodolite or total station, involves the measurement of individual three-dimensional points relative to one another, using a combination of angular and distance measurements. Both the theodolite and total station have a long history in surveying. One of the first uses of a total station to record large ship structures in Denmark was pioneered by Christian Lemée in 1996 during the excavation of the Renaissance ships at the B&W site in Christianshavn (Lemée 2006).

The concept developed was based on identifying the different structural parts of the ship and carrying out a complete survey of the ship as a whole, through the ‘selective field recording of individual elements’, combined with small sections removed for later detailed documentation. Differences occurred when ‘others’ documented portions of the wreck, not in the shape of the lines created, but in the omissions of features recorded by not knowing exactly what was important to document (Lemée 2006: 87).

This clearly highlights some of the issues with interpretive or subjective recording techniques, as pointed out by Lemée when he acknowledged that the process requires the operator to possess specific knowledge and understanding of ship structures when determining which elements to document. In addition, a decision was taken during recording at Christianshavn not to document the ‘Z’ or height measurement as the equipment used was capable of logging the ‘X’ and ‘Y’ coordinates of a point in two to three seconds, while the addition of logging the ‘Z’ coordinate added another seven to eight seconds. With typical logging of up to 1,200 points per day, and a total of circa 32,000 points logged, the additional ‘Z’ coordinate logging would have added days to the process (Lemée 2006: 82).

However, a major problem arose in that, with few exceptions, the shape of all planks and compass timbers were only known in the horizontal projection or plan view, due to the absence of height measurements from the total station data. Lemée (2006: 87) concluded that the total station is an important tool for recording large coherent structures, however, in most cases it needs to be supplemented with other methods.

The recording of multiple wrecks in Yenikapi, Turkey, also employed a total station with a recorded point density of between 4,000 and 10,000 points per wreck depending on wreck size (Kocabaş 2008: 51). The Yenikapi excavation also insisted on the necessity of including photos, sketches and visual remarks to reach the best possible result (Kocabaş 2008: 53).

3D recording and digital modelling

Photogrammetric surveying

Stereoscopy or photogrammetry to produce a 3D view is not new; it was first developed in the 19th century and more fully developed during both World Wars. For underwater surveying of archaeological sites its potential was recognised as early as the 1960’s when George Bass used paired cameras mounted on a mini submarine to document a late Roman wreck. Over the last two decades stereo paired photogrammetry has been almost completely replaced by structure from motion (SfM) photogrammetry with the advent of multi-image photogrammetric software such as Agisoft Photoscan. Photogrammetric survey as a low-cost and rapid tool for maritime archaeological surveying is well reported on (Canciani *et al.* 2002; Skarlatos *et al.* 2012; Henderson *et al.* 2013; McCarthy, Benjamin 2014; Costa *et al.* 2016; Yamafune *et al.* 2017; McCarthy *et al.* 2019).

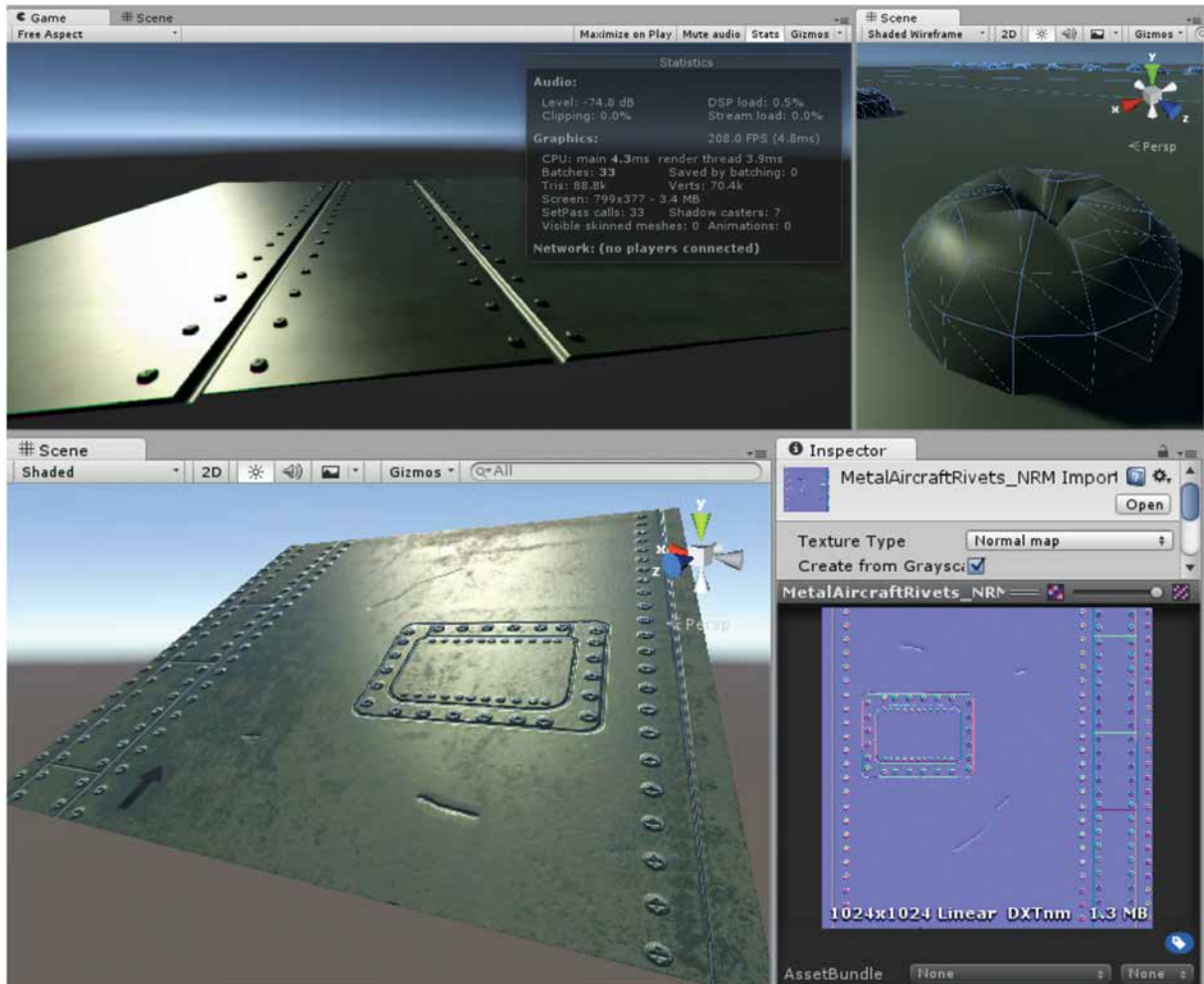


Fig. 1 Computer representation of 3D digital model (images courtesy of Unity Manual 2019.3)

Before discussing photogrammetry in more detail, it is important to understand how computers illustrate or represent 3D digital models. In Figure 1 (top) the digital model of an aircraft panel is a complex three-dimensional mesh model of almost 89,000 polygons (triangles), requiring 33 individual calculations by the computer's graphics sub-system. Model optimization for the video game industry has led to another method of displaying a similarly detailed model Figure 1 (bottom), without the need for complex (time and resource heavy) calculations in the computers graphics sub-system. This approach utilises albedo (colour) map and normal (surface orientation for light reflections) maps to generate the illusion of surface detail on the digital model.

If the model from Figure 1 (top) were 3D printed, it would have all of the detail and geometry as shown in the image. Each of the rivets would be a physical part of the printed model and could be measured etc. However, the digital model shown in Figure 1 (bottom) is a flat square surface consisting of just two polygons. All of the rivets and surface imperfections are an illusion generated using light and shadows, and if 3D printed, the result would be a flat featureless surface.

In the case of photogrammetry, it is not sufficient to simply generate a 3D model with a very high polygon count if there is an insufficient quantity of photographs for the software to calculate the underlying geometry. Figure 2 (top) shows a photogrammetry model of the Bremen Cog created with just 200 photographs using Agisoft Photoscan software. It would appear to have captured the internal structural details of the Bremen Cog, and the model appears to be high resolution, with a total of over 22 million polygons. Figure 2 (bottom) is a close-up view of the same 3D model with the albedo (colour) and normal map removed to illustrate the lack of geometric detail.

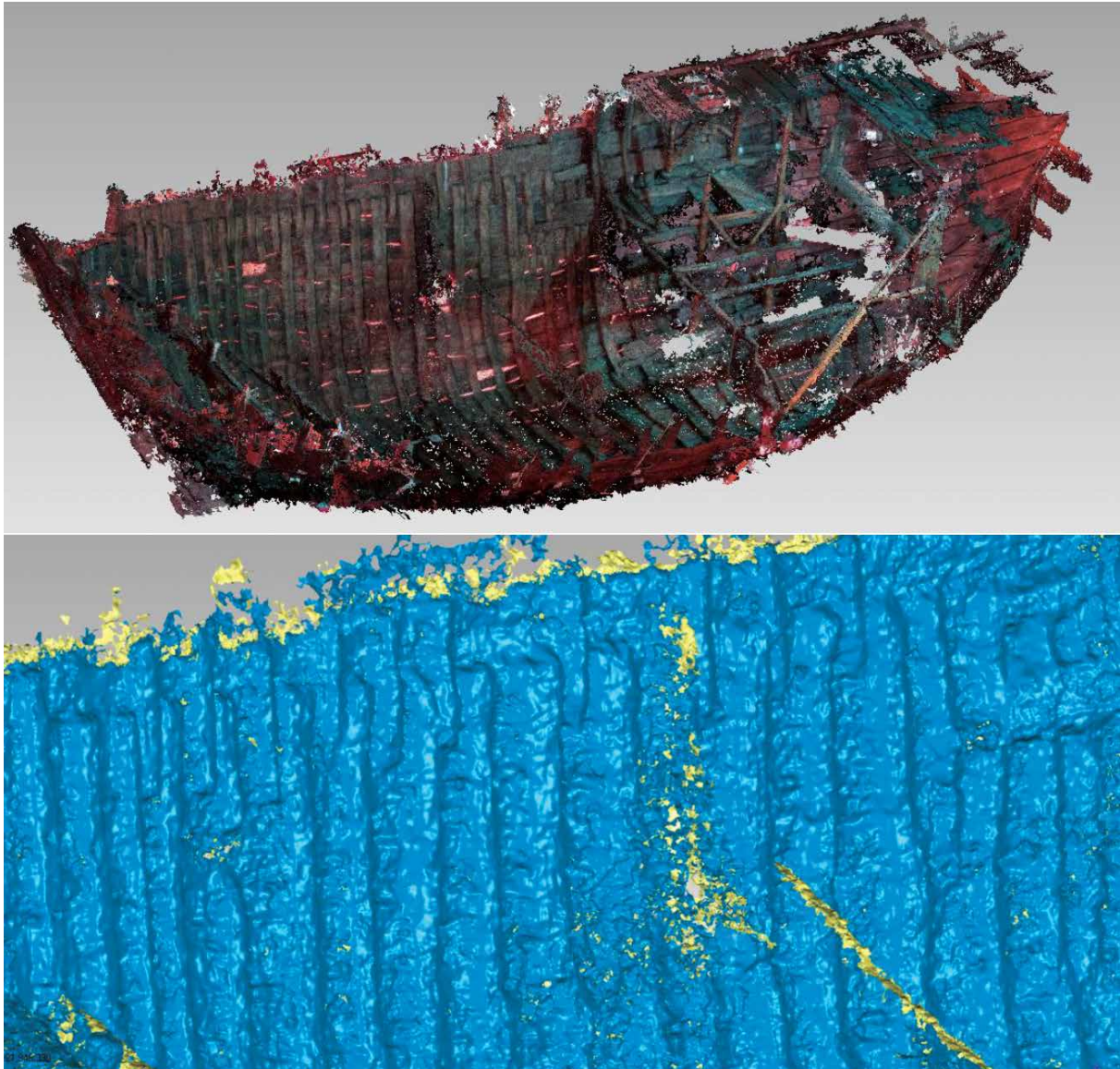


Fig. 2 Photogrammetry model of the Bremen Cog (author: Pat Tanner)

Photogrammetry is not a real time or automated process and most photogrammetric tasks are laborious and tedious. The surveyors can only be sure of the data collection consistency after they have successfully resolved image bundle adjustment during the post-processing phase. It is the software which produces a 3D model by extracting a dense point cloud from the bundle adjusted and stereo matched images. The resultant 3D point clouds created have arbitrary scale, orientation and position in 3D space. All 3D photogrammetric models will require the application of a scale factor correction, as well as a rotation and translation matrix to known correspondence points obtained from other surveying techniques, in order to obtain the correct scaling and georeferencing.

LiDAR or terrestrial laser scanning (TLS)

Photogrammetry and LiDAR are often juxtaposed and presented as being at odds with each other. Photogrammetry uses two-dimensional photographs to interpret measurements between objects and create a three-dimensional geometric representation of the objects themselves. LiDAR uses lasers to detect the position and geometric shape of an object by generating three-dimensional point clouds based on laser shots, and each individual point within the point cloud has its own measured X, Y and Z coordinates. Simply put, with photogrammetry it is each two-dimensional image which is the Raw Source Data, and the 3D pointcloud is a computer based interpretation of that source data, whereas, with laser scanning, the 3D pointcloud is the Raw Source Data.



Fig. 3 3D laser scanning a 9 m vessel (author: Pat Tanner)

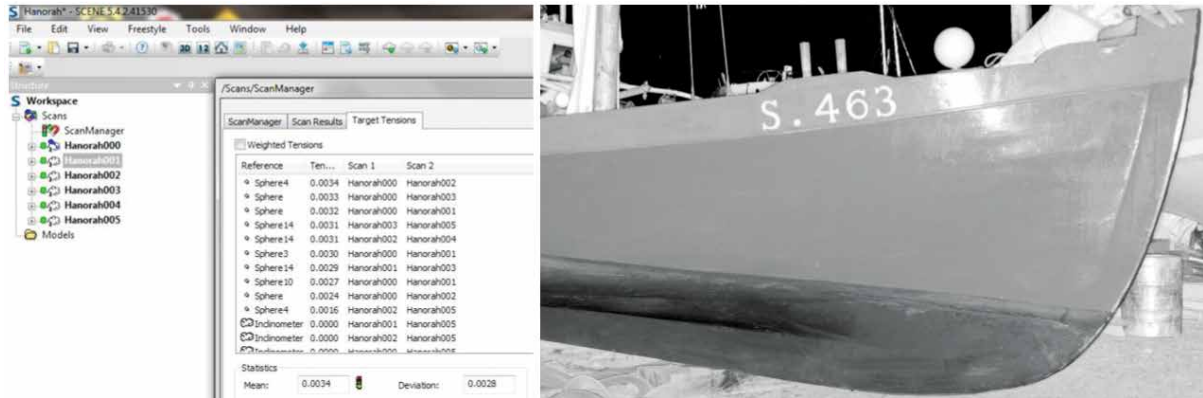


Fig. 4 Post processing scanned point cloud (author: Pat Tanner)

With 3D laser scanning the scanner records the physical characteristics of everything within its line of sight. For this reason, it is necessary to reposition the scanner (red circles in Fig. 3 left) to record the occluded elements not recorded from the initial scan position. Initial on-site registration (Fig. 3 right) between individual scans uses internal GPS sensors, accurate to within a few centimetres. Post processing software allows for further refinement in the registration using a target-based, or cloud to cloud registration (Fig. 4 left), typically resulting in 2–3 mm accuracy. The individual scans are not modified, but rather, a translation and rotational matrix, based on the optimised registration, is stored as metadata with each individual scan file. The result is a geometrically accurate digital three-dimensional pointcloud record of the source data (Fig. 4 right).

Hybrid or Combined techniques

Multi-image photogrammetry software interprets pixel information from digital images to create 3D point clouds, however control points are required to correctly orientate and scale the resultant model. Lidar or terrestrial laser scanning, on the other hand, produces geometrically accurate 3D point clouds, but tends to have a lower surface colour resolution. Recent software developments such as RealityCapture® means the two techniques can now be used together. This software can use the geometrically accurate laser scanning as a control network for scale and relative positioning, while using the high-resolution photogrammetry for colour and surface texture. In addition, any occluded areas which are not recorded from one source can be augmented by the other and vice versa.

In a proof-of-concept case study, 22 individual laser scans were recorded of the Mary Rose, taking a total of two hours scanning time without colour, 11 locations at ground floor level and 11 locations on the third-floor gallery. The limitations on scanner positioning caused occluded areas, resulting in missing data clearly visible on the intermediate decks (Fig. 5 top). In addition to the laser scans, 374 high resolution colour photographs were recorded. With the dimensionally accurate (within 2 mm) 3D laser scan model being used as the control network, RealityCapture® then aligned the photographs, which enabled the creation of the missing geometry in the occluded areas (Fig. 5 middle), and the application of high-resolution photographic quality colour texturing (Fig. 5 bottom). This resulted in a very high

resolution, geometrically accurate, digital research model, as well as lower resolution models for easier dissemination such as on sketchfab (<https://sketchfab.com/3d-models/mary-rose-316db8d7099b42b28f889aedddc86e9d>), which has had over 17,300 views.

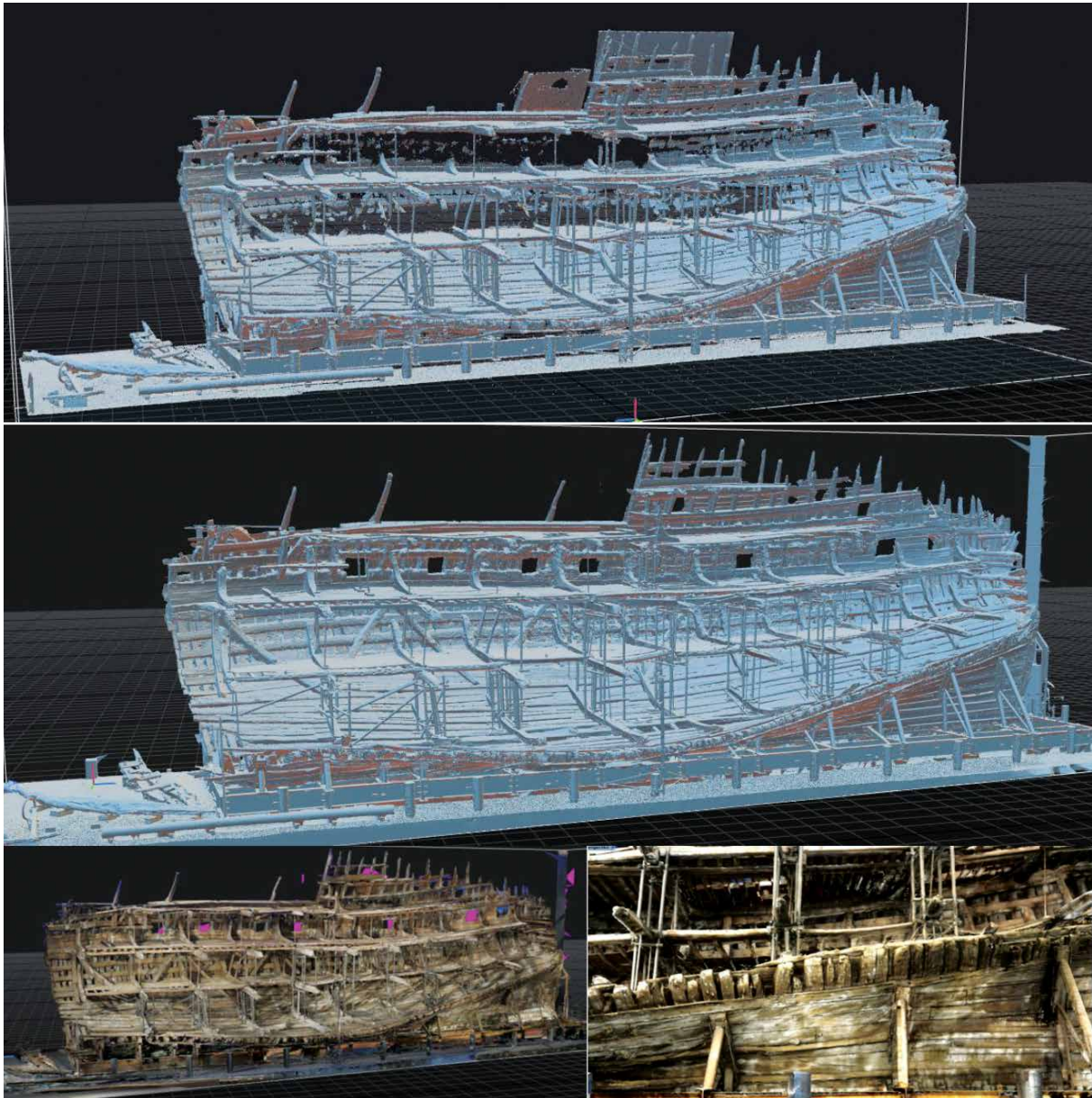


Fig. 5 Hybrid data capture combining 3D laser scanning and photogrammetry (author: Pat Tanner)

4D recording

As each stratigraphic layer is excavated, and the subsequent layer exposed, it is now possible, using photogrammetry, laser scanning or a combination of both, to accurately (and objectively) capture the site information as each subsequent stratigraphic layer is excavated and exposed (Pacheco-Ruiz *et al.* 2018). This would capture an accurate, three-dimensional, point-in-time snapshot, at various significant stages throughout an excavation, and as archaeological excavation is destructive by its very nature, each 3D survey would capture information otherwise impossible to re-examine at later stages.

Metadata / Paradata

Quite simply, metadata is data that describes other data. Metadata describes whatever piece of data it is connected to, whether that data is video, a photograph, spreadsheets, illustrations or reports. Various types of metadata include: Descriptive metadata – in its most simplified version, an identification of specific data; Structural metadata; Preservation metadata; Provenance metadata; Use metadata; and administrative metadata. Even this article has metadata such as the author, number of pages or words, creation date, modification date, file size etc.

Paradata on the other hand, is information about the human processes of understanding and interpretation of data objects. Examples of paradata include descriptions stored within a structured dataset, of how evidence was used to interpret an artefact, or a comment on methodological premises within a research publication. It is closely related, but somewhat different in emphasis, to ‘contextual metadata’, which tends to communicate interpretations of an artefact or collection, rather than the process through which one or more artefacts were processed or interpreted. Once we begin to sort, arrange or interpret the raw source data, paradata becomes critical in order to explain both how and why.

An example of metadata for a single floor timber from the Newport Medieval Ship would be the timber database file which identifies that timber as F44 and relates that to the Cow Tag CT797 attached to that timber on-site, CT797_Timbersheet.pdf – a digitised copy of the hand written timber record sheet describing that timber when excavated, digital photographs (CT797_1.tif, CT797_2.tif, CT797_3.tif etc), F44.3dm – the digital 3D wireframe data from the Faro Arm contact digitising, F44.stl – the digital file for 1:10 scale 3D printing, as well as conservation reports, dendrochronology reports and additional drawings illustrating that timbers find location as well as its location in the hypothetical reconstruction.

The paradata relating to that timber is contained within the specialist report – Reconstructing the Hull Shape (Tanner 2013). That report describes how the object was analysed and interpreted, the process and methodology employed, and the reassembly of individual components into a coherent structure. It also describes how that structure was analysed, how distortions were repaired, how the hypothetical reconstruction was created, and how that hypothesis was tested using hydrostatic analysis to establish seafaring, cargo capacity and performance characteristics.

Conclusion

The capturing of three-dimensional high volume, high quality raw data, using either photogrammetry, 3D laser scanning, or a combination of both, generates a superior archival record, which is stored as a full-sized three-dimensional object, rather than a reduced scale, two-dimensional paper-based interpretation. Advances in viewing technology allow for three-dimensional interactive models to be viewed on a computer monitor, allowing the user to view the object from any angle or viewpoint, or, with the aid of augmented reality or virtual reality headsets, to actually interact with the three-dimensional raw source data.

In the case of cultural heritage sites and historical artefacts, the physical context of the objects contained within a site are just as important as the artefacts themselves. Just as a site has stratigraphy, ships also have a stratigraphy in themselves, which must be carefully recorded. The rapid advances in both hardware and post-processing software means it is now possible to easily capture high volume, high-quality, 3D data. This form of digital data, combined with suitable metadata and, more importantly, paradata, should go a long way towards bridging the gap between the exclusive knowledge of the excavator and the published record, allowing lots of people to use the data in new and novel ways. This raises the question: As archaeologists, should we all be doing this?

References

- Baltsavias, E. P. 1999. A Comparison between Photogrammetry and Laser Scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* 54.2, July 1: 83–94.
- Bass, G. F. 2011. Introduction. In A. Catsambis, B. Ford, D. Hamilton (eds), *The Oxford Handbook of Maritime Archaeology*. New York, Oxford University Press: 3–22.
- Canciani, M., Gambogi P., Romano F.-G., Cannata G., Drap P. 2003. Low Cost Digital Photogrammetry for Underwater Archaeological Site Survey and Artifact Isertion. The Case Study of the Dolia Wreck in Secche Della Meloria-Livorno-Italia. In *ISPRS Workshop Vision Techniques for digital architectural and archaeological archives, Jul 2003, Ancona, Italy*: 95–100. Retrieved (05/2020) from <https://www.isprs.org/proceedings/XXXIV/5-W12/>
- Costa, E., Beltrame, C., Vernier, P., Balletti, C., Guerra, F. 2016. Digital survey techniques for the documentation of wooden shipwrecks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B5. https://www.academia.edu/28203887/DIGITAL_SURVEY_TECHNIQUES_FOR_THE_DOCUMENTATION_OF_WOODEN_SHIPWRECKS
- Crumlin-Pedersen, O. 1977. Some Principles for the Recording and Presentation of Ancient Boat Structures. In S. McGrail (ed.), *Sources and Techniques in Boat Archaeology*, [BAR Supplementary Series 29]. Greenwich, National Maritime Museum: 163–177.
- Denard, H. 2009. The London Charter for the Computer-Based Visualisation of Cultural Heritage. *Draft* 2.1: 1–13.
- Henderson, J., Pizarro, O., Johnson-Roberson, M., Mahon, I. 2013. Mapping Submerged Archaeological Sites Using Stereo-Vision Photogrammetry. *International Journal of Nautical Archaeology* 42.2: 243–256.
- Hocker, F. 2000. New Tools for Maritime Archaeology. *Maritime Archaeology Newsletter from Roskilde, Denmark* 14: 27–30. http://www.maritimearchaeology.dk/?page_id=1714
- Hocker, F. 2003. *Three-Dimensional Documentation of Ship Timbers Using the FaroArm*. Manual. Roskilde, Nationalmuseets Marinarkaeologiske Forskningscenter.
- Hocker, F. 2004. Shipbuilding: Philosophy, Practice and Research. In F. M. Hocker and C. A. Ward (eds), *The Philosophy of Shipbuilding: Conceptual Approaches to the Study of Wooden Ships*. College Station, Texas A&M University Press: 1–11.
- Hocker, F. 2013. In Details Remembered. In J. Adams (ed.), *Interpreting Shipwrecks: Maritime Archaeological Approaches*, [Södertörn academic studies 56]. Southampton, Highfield Press: 72–84.
- Holt, P. 2003. An Assessment of Quality in Underwater Archaeological Surveys Using Tape Measurements. *International Journal of Nautical Archaeology* 32.2: 246–251.
- Kocabaş, U. (ed.) 2008. *The 'Old Ships' of the 'New Gate', Vol. 1 Yenikapı'nın Eski Gemileri*. Istanbul, Istanbul University.
- Lemée, C. P. P. 2006. *The Renaissance Shipwrecks from Christianshavn: An Archaeological and Architectural Study of Large Carvel Vessels in Danish Waters, 1580-1640*. Roskilde, Viking Ship Museum.
- McCarthy, J., Benjamin, J. 2014. Multi-Image Photogrammetry for Underwater Archaeological Site Recording: An Accessible, Diver-Based Approach. *Journal of Maritime Archaeology* 9.1: 95–114.
- McCarthy, J. K., Benjamin, J., Winton, T., van Duivenvoorde, W. (eds) 2019. *3D Recording and Interpretation for Maritime Archaeology*, [Coastal Research Library]. Springer International Publishing.
- Pacheco-Ruiz, R., Adams, J., Pedrotti, F. 2018. 4D Modelling of Low Visibility Underwater Archaeological Excavations Using Multi-Source Photogrammetry in the Bulgarian Black Sea. *Journal of Archaeological Science* 100: 120–129.
- Skarlatos, D., Demestiha, S., Kiparissi, S. 2012. An 'Open' Method for 3D Modelling and Mapping in Underwater Archaeological Sites. *International Journal of Heritage in the Digital Era* 1.1: 1–24.
- Tanner, P. 2013. *Newport Ship Specialist Report: Digital Reconstruction and Analysis of the Newport Ship*, [Newport Medieval Ship Archive]. York, Archaeological Data Service. <https://doi.org/10.5284/1044659>
- Yamafune, K., Torres, R., Castro, F. 2017. Multi-Image Photogrammetry to Record and Reconstruct Underwater Shipwreck Sites. *Journal of Archaeological Method and Theory* 24.3: 703–725.