

As archaeology operates using knowledge through destruction it is imperative that our records both accurately record the archaeology, to the best of our ability, and are easily accessible and reusable to future researchers. To this end it is imperative that any data produced throughout an archaeological project is preserved in a manner intended for long-term storage (ADS, no date). In 2016 an article detailing four key points for long-term data stewardships, named FAIR: Findable, Accessible, Interoperable, and Reusable (Wilkinson *et al.*, 2016) was published and led to the creation of the GoFAIR website (*GO FAIR initiative: Make your data & services FAIR*, no date). These principles ensure that any data recorded is preserved in the long-run in a manner that makes the data available for reuse. These principles should also be applied to metadata, which provides context to any data(set) produced and / or submitted to archive, ensuring that the data archived meets the FAIR principles (Zoldoske, 2020).

The dataset referenced in this commentary and metadata statement is the results from the rapid assessment of architectural, roofing, and otherwise worked stone from Sidney Sussex College in Cambridge. The data was first recorded on printed templates provided by a specialist and later digitised onto a spreadsheet. Most of the stone from the site has been recommended for discard (Shaffrey, 2021) in line with the deposition guidelines for Cambridgeshire (Croft, Carroll and Wiles, 2020) so it is imperative that the data and metadata is preserved as it will be the only remaining record of the stone from the site. This essay aims to explain and justify why certain choices were made in the structure and categorisation / classification of the dataset using ADS guidelines, the FAIR principles, and practicalities in the workplace.

The Findability and Accessibility principles have largely not been met with the dataset alone, and will only partially be met in the metadata statement. The Archaeology Data Service will assign a unique identifier to the dataset upon deposition, along with all other project data, in addition with a checksum which is assigned to each individual file and changes if the data is corrupted or otherwise changed (ADS, 2021). In terms of accessibility, the ADS is a free-to-use and open access (*A1.1: The protocol is open, free and universally implementable*, no date), and also requires depositors to have an account with them which allows the owner of the dataset to be authenticated (*A1.2: The protocol allows for an authentication and authorisation where necessary*, no date). These two principles will not be the primary focus of the commentary below, as these will be provided for after deposition.

The dataset is a .xlsx format which is one of the preferred file formats for submission to the ADS (ADS, 2021) which is why it was chosen. This is one of the preferred formats as it is the most popular file formats globally meaning it is not likely to be discontinued making it ideal for long-term curation; it is also supported on various hardware and across operating systems meaning that they are more accessible to other/future researchers. With propriety file formats there is always the possibility that the developer will discontinue a certain

version in order to push people to upgrade, similarly with open formats certain formats may become unpopular and therefore also be discontinued (Digital Preservation Coalition, 2021). Microsoft Office enables the user to save a document as many different formats: older versions of Office formats, open documents, and most notably comma separated values (.csv). The latter format, .csv, is a plain text file where cell values are separated by a comma. The simple plain text file enables the data to be transferred between spreadsheet and database programmes easily making it ideal for digital preservation as should the file format be discontinued it can be easily transferred; .csv files do not contain any formatting so it is a good format for storage as the file size is smaller and can be read by any spreadsheet or database programme. In regards to the FAIR principles by using a Microsoft Office format that is supported by the primary digital archive for archaeology it is ensured that the dataset is interoperable and reusable.

In order for the data to be interoperable a standardised set of vocabularies must be used within the dataset so it can easily be compiled with data from other researchers (*I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*, no date). The stone recording spreadsheet uses standardised categories for the lithology of the stones as set by the Forum on Information Standards in Heritage (FISH) in their Building Materials Thesaurus (Forum on Information Standards in Heritage, 2021) – though the range of stones on the site was constrained to clunch, Lincolnshire limestone, oolitic limestone and sandstone. The FISH website states that their forum is guided by the FAIR principles (Forum on Information Standards in Heritage, no date), meaning that they are intended to be a way of standardising archaeological data. Standardisation of data makes it: easier to find as there is a common word to search for, interoperable as it can easily be combined with other datasets (providing that they have also applied the FISH vocabularies), and reusable as it meets a community standard (*R1.3: (Meta)data meet domain-relevant community standards*, no date).

Though unconventional, the stones from the site were assessed by four members of the post-excavation team at L – P: Archaeology as opposed to a specialist, although the post-excavation team had received a day's training from the specialist who would be reporting on the site. Given the large quantity of stone from the site (some 268 pieces of architectural stone, 176 roofing tiles, and 513 miscellaneous stones (Appleyard, 2020)), and the inexperienced nature of the assessors, the quality of the data recorded changed throughout the assessment, and would not necessarily be of the same standard had a specialist assessed the site themselves. This was unavoidable due to the size of the assemblage, the schedule of the specialist, and budgetary restraints. As a result of these practicalities the dataset is perhaps not as standardised as it could be leaving many columns with descriptions and comments where a standardised answer would have sufficed. This is particularly evident on the architectural and roofing stone tabs. For the recording of the roofing stone a lot of the important information was recorded in the 'notes' column when on reflection this could have been divided into separate columns – one for perforation, one

for evidence of heating / burning, and another for probable original shape. This would have made the data more easily accessible and reusable for other and future researchers. It is difficult to justify decisions made when, in hindsight, it is clear that there were better options, although in this case the inexperience of the people recording the stone played a part, the templates provided by the specialist were also perhaps unequipped to handle an assemblage of this size or to handle multiple people recording the same site.

Overall, the structure and classification / categorisation of the data meets the FAIR principles and the guidelines set out by the Archaeology Data Service. Regarding the FAIR principles the data meets the Interoperability and Reusability principles more than the Findability and Accessibility. The Findability and Accessibility principles will be achieved through the deposition of the dataset with the Archaeology Data Service where it will become part of a searchable database on their website with a unique and persistent identifier. The Interoperability and Reusability principles have been met through the use of standardised vocabulary and a standardised metadata statement template.

A1.1: The protocol is open, free and universally implementable (no date) *GO FAIR*. Available at: <https://www.go-fair.org/fair-principles/a1-1-protocol-open-free-universally-implementable/> (Accessed: 16 November 2021).

A1.2: The protocol allows for an authentication and authorisation where necessary (no date) *GO FAIR*. Available at: <https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/> (Accessed: 16 November 2021).

ADS (2021) 'Guidelines for Depositors: Preparing Datasets'. Archaeology Data Service. Available at: <https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#>.

ADS (no date) *Guides to Good Practice: DbSht_1, Section 1. Introduction to Databases and Spreadsheets*. Available at: https://guides.archaeologydataservice.ac.uk/g2gp/DbSht_1 (Accessed: 12 November 2021).

Appleyard, E. (2020) 'Stone Recording Data - ECB4266 - Sidney Sussex College'. Unpublished.

Croft, S., Carroll, Q. and Wiles, J. (2020) *Deposition of archaeological archives in Cambridgeshire*. Cambridgeshire County Council.

Digital Preservation Coalition (2021) *Fixity and checksums - Digital Preservation Handbook*. Available at: <https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums> (Accessed: 16 November 2021).

Forum on Information Standards in Heritage (2021) 'FISH Terminologies: Building Materials Thesaurus'. Available at: <http://www.heritage-standards.org.uk/fish-vocabularies/>.

Forum on Information Standards in Heritage (no date) *About, FISH*. Available at: <http://www.heritage-standards.org.uk/> (Accessed: 16 November 2021).

GO FAIR initiative: Make your data & services FAIR (no date) *GO FAIR*. Available at: <https://www.go-fair.org/> (Accessed: 15 November 2021).

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation (no date) *GO FAIR*. Available at: <https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/> (Accessed: 15 November 2021).

R1.3: (Meta)data meet domain-relevant community standards (no date) *GO FAIR*. Available at: <https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/> (Accessed: 16 November 2021).

Shaffrey, R. (2021) *Sidney Sussex Stone Report*. Unpublished. L-P: Archaeology, p. 3.

Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3(1), p. 160018. doi:10.1038/sdata.2016.18.

Zoldoske, T. (2020) 'The exciting world of Metadata | Sound Bytes from the ADS', 5 November. Available at: <https://archaeologydataservice.ac.uk/blog/2020/11/the-exciting-world-of-metadata/> (Accessed: 16 November 2021).

Word Count: 1810 (547 on metadata table)

File Name	Title	Description	Creator (if more than one individual/organisation, add on a new row)		
			First Name	Last Name	Organisation
ECB4266_Sidney_Sussex_College_stone_data.xlsx	Stone Recording Data from ECB4266	Results of the rapid assessment of the stone from Sidney Sussex	Aimee	Hudson Brown	L-P: Archaeology
			Ella	Appleyard	L-P: Archaeology
			Yasmine	de Gruchy	L-P: Archaeology
			Gosia	Krawczyk	L-P: Archaeology
Sheet Name	Sheet Description	Number of rows		Field name (Please start a new cell for each row)	Field description (inc. units of measurements, description of codes, etc. used within each field).
Building Stone	The data collected from the rapid assessment of the building stone collected from Sidney Sussex College	269		Context	The context the find is from
				Date Recorded	the date the stone was assessment

				Condition	whether the stone is complete or damaged, if possible providing a percentage completeness
				Category	is it an ashlar block, moulded stone, rubble, or other. Codes used: ashlar, moulded, rubble, other
				Original Faces Surviving	how many original faces are present and identifiable
				Weight (g)	how much does the stone weigh in grams
				Mortar	is mortar present, if so on how many faces

				Tool Marks	are tool marks present, and if so from which tool (giving measurements in mm)
				Description	a description of the stone, including anything worth noting such as engravings or evidence of reuse
				Length	how long the stone is in cm
				Width	how wide the stone is in cm
				Depth	how deep the stone is in cm
				Lithology	what type of stone is it, in line with the FISH Building Materials vocabulary
				Sketch	was the stone sketched

				Further Recommendations	does this stone require further analysis
				Discard?	can this be recommended for discard.
Roofing	the data collected from the rapid assessment of the stone thought to be roofing from Sidney Sussex College	177		Context	The context the find is from
				Date Recorded	The date the stone was recorded
				Lithology	what type of stone is it, in line with the FISH Building Materials vocabulary
				Weight (g)	The weight of the stone (g)
				Type	R = Roofing // PR = Probable Roofing

				Notes	size of perforation in mm, how many corners survive, original shape, any evidence of heating, burning, or fire.
				Depth	the thickness of the stone (mm)
				Width	the width of the stone (mm)
				Length	the length of the stone (mm)
				Further Recommendations	recommended for further analysis
				Discard?	recommended for discard
Small Fragments	the data collected from the rapid assessment of small stone fragments from Sidney Sussex College	186		Context	The context the find is from
				Date Recorded	the date the find was recorded

				Lithology	what type of stone is it, in line with the FISH Building Materials vocabulary
				No.	how many sherds
				Weight (g)	the total weight of the sherds
				Type	R = Roofing // PR = Probable Roofing // BS = Building Stone // OTH = Other
				Additional Notes	any evidence of heating or burning, any tool marks, evidence of reuse etc.
				Discard?	recommended for discard

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]