Peer Community In Archaeology

Open data on beads, pendants, blanks from the Ceramic Age Caribbean

Ben Marwick based on peer reviews by **Robert Bischoff**, **Stefano Costa**, **Clarissa Belardelli** and **Li-Ying Wang**

Alain Queffelec, Pierrick Fouéré, Jean-Baptiste Caverne (2021) A database of lapidary artifacts in the Caribbean for the Ceramic Age. SocArXiv, ver. 4, peer-reviewed and recommended by Peer Community in Archaeology.
https://doi.org/10.31235/osf.io/7dq3b

Submitted: 13 November 2020, Recommended: 28 February 2021

Cite this recommendation as:
Marwick, B. (2021) Open data on beads, pendants, blanks from the Ceramic Age Caribbean. *Peer Community in Archaeology*, 100009. 10.24072/pci.archaeo.100009

Published: 28 February 2021

The paper 'A database of lapidary artifacts in the Caribbean for the Ceramic Age' by Queffelec *et al.* [1] presents a description of a dataset of nearly 5000 lapidary artefacts from over 100 sites. The data are dominated by beads and pendants, which are mostly made from Diorite, Turquoise, Carnelian, Amethyst, and Serpentine. The raw material data is especially valuable as many of these are not locally available on the island. This holds great potential for exchange network analysis. The data may be especially useful for investigating one of the fundamental questions of this region: whether the Cedrosan and Huecan are separate, little related developments, with different origins, or variants or a single tradition [2]. In addition to metric and technological details about the artefacts, the data include a variety of locational details, including coordinates, distance to coast, and altitude. This enables many opportunities for future spatial analysis and geostatistical modelling to understand human behaviours relating to ornament production, use, and discard. I recommend the authors make a minor revision to Table 1 (spatial coverage of the dataset) to make the column with the citations conform to the same citation style used in the rest of the text. I warmly commend the authors for making transparency and reproducibility a priority when preparing their manuscript. Their use of the R Markdown format for writing reproducible, dynamic documents [3] is highly impressive. This is an excellent example for others in the international archaeological science community to follow. The paper is especially useful for researchers who are new to R and R Markdown because of the elegant and accessible way the authors document their research here. [1] Queffelec, A., Fouéré, P. and Caverne, J.-B. 2021. A database of lapidary artifacts in the Caribbean for the Ceramic Age. SocArXiv, 7dq3b, ver. 4 Peer-reviewed and recommended by PCI Archaeology. https://doi.org/10.31235/osf.io/7dq3b [2] Reed, J. A. and Petersen, J. B. 2001. A comparison of Huecan and Cedrosan Saladoid ceramics at the Trants site, Montserrat. In Proceedings of the

XVIIIth International Congress for Caribbean Archaeology (pp. 253-267). [3] Marwick, B. 2017. Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. Journal of Archaeological Method and Theory 24, 424–450. https://doi.org/10.1007/s10816-015-9272-9

Reviews

Evaluation round #1

DOI or URL of the preprint: **10.31235/osf.io/7dq3b**

Authors' reply, 07 January 2021

Thank you Mr. Marwick for handling the review process of our manuscript and underlining the efforts for transparency and reproducibility. We have now read and answered to the comments of the four reviewers. Our answers are integrated in the document attached, in blue and italic. We have modified our manuscript accordingly and uploaded the new version on the preprint server, as well as the new versions of the different files for the database and the R markdown document.

Download author's reply

Download tracked changes file

Decision by Ben Marwick, posted 29 December 2020

Open data on beads, pendants, blanks from the Ceramic Age Caribbean

This manuscript is a most welcome announcement of a database of lapidary artifacts from the Ceramic period of the Caribbean islands. Open data such as presented here is essential for the sustainability of archaeology as a discipline and a community. I recognise the authors' impressive efforts to make this manuscript transparent and reproducible by writing in R Markdown. This has enabled the reviewers to identify some specific corrections necessary before we can recommend the article. The reviewers have also made many suggestions for improving the English text to make the text easier to understand. I request that the authors study the reviews and make all of the revisions requested (the reviewers often spot the same things, so this will not be a big task). Please also look into the suggestions to improve the text and make those changes (especially those made by Bischoff on the PDF he attached to his review). I believe that once these revisions are complete, we can recommend the article to the scholarly community.

Reviewed by Clarissa Belardelli, 13 December 2020

I am going to articulate my review into three parts: 1. Scientific content 2. Bibliography, figure captions 3. Translation

1. Scientific content The paper is very good and the argument amazing. method is well conscructed; the overview maybe is not too vaste but there is a good bibliography. However something is not clear: P. 4: "Presently, the dataset of lapidary artifacts contains 0 entries, originating from 87 sites": so, how many entries does the dataset contain? P. 5: "We will thus describe only the English tables Islands, Sites and Beads. Each table also exists in French 3.0.1 Islands table (ISLANDS and ILES)" but it is shown also in french. So, what do you mean actually? It is not clear. P. 5: 3.02, Sites Table. "Nb_beads is the calculated number of artifacts related to this site in the BEADS table" That give place to confusion! A. Artifacts or only beads? B. In each site or in the structure linked to the site (see below in your paper)? It would be

pretty useful to add something like presence of other lapidary artifacts from the SITE: Y/N and then specify which type of, and the number of each type in each structure. P. 6: “Perforation is the number of perforation”. But if you found that object in bibliography, and you cannot explore the object, how can you decide whether the hole is only one or more than one? It would be much helpful PERFORATION: Y/N for doubtful cases, so you would have at least a generic information.

2. Bibliography, Figures captions There are some problems in the rendering of bibliography, but perhaps it depends on my pc. Cody’s quote is different in the three titles of the bibliography: why? In the text, figures are referred to with “Fig.” but each caption has “Figure”: why? Fig. 4: Figure 4: A. Screenshot of the ArkeoGIS application, a simplified GIS online system. B. Zoom on Guadeloupe, sowhing the potential of the ArkeoGIS visualization tool. (showing) To be corrected!

3. Translation In general, your english could be better. You should ask for a native speaker who helps you.

There are some minor mistakes. I noticed two of them but there could be more. Here, one: - P. 6: State specifies is (IF) the object is complete or broken To be corrected!

Finally, the paper is good but needs some more attention.

Reviewed by **Robert Bischoff**, 30 November 2020

This paper describes a database of lapidary artifacts from the Ceramic period of the Caribbean islands. A brief background of the setting and cultural history is provided, as well as a discussion on prior lapidary research. The bulk of the paper is a description of the data collection methods, database structure, and repository locations. The description contains the necessary metadata to interpret and access the database (URLs are provided), including caveats regarding data quality and limitations. Many archaeologists and researchers in other disciplines have called for more open-access databases with appropriate metadata. This article and database answer that call and provide a large database with location data that should be useful for researchers in the Caribbean as well as researchers and educators looking for high-quality datasets. I would prefer to see a short case study demonstrating the potential this database has for providing new archaeological interpretations; however, the article as written merits recommendation with the addition of some minor revisions to the article and by addressing a few problems with the database. The strength of this paper is its clear description of the different elements of the database and how they fit together with enough detail to confidently use the database to address research questions. The tables and figures are helpful and of good quality. The inclusion of the Rmarkdown document is an excellent example of open source science, and I had no trouble reproducing it. This will be useful for anyone looking to adopt a similar format for their own work. The database itself is of good quality compared to many archaeological databases and contains a number of useful fields. The location data does not have any obvious errors. There are a few important issues that need to be addressed prior to a recommendation. The Islands table discussed in the text is not included in the OSF repository and has not been reviewed. I used the tables in the OSF repository to examine the database. I was not able to join all of the sites to the beads tables in English or French. I also tried using the Index*Site field to join to the first part of the Index*B field (I removed the three-digit numbers at the end of Index*B), but I found a similar problem to just using the Site field. This should be fixed prior to a recommendation. I suggest including the Index*Site field in the BEADS table as it will be easier to avoid problems with typos and will be a more stable key for joining tables. There are three rows in the BEADS and PERLES tables that are exact duplicates and should be removed (GR-01-044, GR-01-099, and GR-01-128). There are also several rows that are included in the BEADS table but missing in the French version called PERLES. Some of the terms in the English version have not been translated from French. For example, the same field has entries for “Bead-Pendant” and “Bead-pendentif.” This will cause problems aggregating the data. There is one small problem I identified in the Rmarkdown document. The size of the database is calculated by the length of the Index*P field, but the English version is loaded which is called Index*B and results in a value of zero. Perhaps using nrow(Data) would be simpler. Some minor changes to language and phrasing in the article are suggested.

Download the review

Reviewed by Stefano Costa, 18 December 2020

This data paper describes an interesting and very wide-ranging dataset about a specific artifact category found in archaeological contexts across the Caribbean region. I can recommend its publication with some revisions described in the following review comments. Section 1 Spatial coverage

Since the spatial distribution of sites is concentrated in the Eastern part of the Caribbean, with the exception of Jamaica and Bahamas, it seems useful to provide a more detailed description of the coverage beyond the geographic bounding box, perhaps including the names of smaller archipelago groups.

The authors mention that

> some islands did not yield any lapidary artifact, or at least none that we could find in the literature

and I think it is important to give a list of what islands were investigated but yielded no results.

The map in figure 1 is a little confusing because some of the site IDs do not match those in the table or the sites listed in the table are not labeled on the map. It could be useful to draw the bounding box in the map, perhaps differentiating islands that are included in the dataset versus those that are not included. Section 2 Steps

When the authors write that

> the dataset of lapidary artifacts contains 0 entries.

this **must** be corrected with the actual number, by changing the code snippet at line 86 of the PAAF-datapaper.Rmd source file from r length(Data$index_P) to r length(Data$index_B), which correctly gives 5011 entries. Section 3 Dataset description

The ISLANDS and ILES tables are not included in the archived dataset at SocArXiv / OSF. These **must** be added for the dataset to be complete.

The dataset description is very detailed and well organised. There is a lot of potential for reuse and further analysis with the radiocarbon data, as published JOAD data papers have shown: I may suggest to split the *Date_BP* field in two separate fields, to have better error checking (it becomes two integer fields instead of a text field) and make it easier to reuse the information. If the field contains more than one date, perhaps a separate table could be added.

It is unclear why the dataset is described as CSV but the reproducible paper is based on the XLSX version, this could create unnecessary complexity and/or mismatch between different versions of the dataset.

The Filemaker server is useful for quick interaction and most importantly contains photographs and drawings. A reason for not including those in the archived dataset should be given, since it represents a major source of information for the study.

Reviewed by Li-Ying Wang, 19 December 2020

This paper describes the dataset of lapidary artifacts during the Ceramic period in the Caribbean islands. My comment below is based on the criteria listed on JOAD website that focuses on the description of the data and best practices for data deposition.

For the method section, I think the paper provides good information for understanding how the dataset was created, including the source of data, the methodology for collecting data, sampling strategy, quality control, and constraints. But I would like to suggest to the authors to provide more details about how they searched the ornament-related words in literature. For example, they mentioned “the words…have been systematically searched for.” What is the exact method for systematically searching? I think it would be better to specify. Similar issues for the quality control section, how did they do for data cleaning? For constraints, how did they

indicate the missing values in the dataset? I can see most missing values are indicated by blank in csv files, but some are indicated by a dash. Are these the same? This needs to be corrected or clarified. This can be addressed in the main text since incompleteness is a constraint in their dataset as they recognized.

In addition to the above questions, there are some mistakes in the text:

1, Page 4, 2.1 steps: "the dataset of lapidary artifacts contains 0 entries, originating from 87 sites". Surely zero is not correct. There should be a number instead of 0.

2, Page 4, 2.4 constraints, "The quality of information has been problematic for several topic of the database" should this be 'topics' as a plural?

3, Also, this sentence "the quality of the reproduction of ancient photographs in the numeric documents now accessible for this literature" is unclear and would be better to rephrase.

For the openness of data, they deposited their dataset on OSF where can be easily accessed and examined. The data is actionable and mostly labeled nicely. However, the information in the manuscript and repository about the number of files is not consistent with each other. In the 3.7.1 Download section, there are six csv files listed in the main text, but only four csv files are uploaded to OSF. The "ILES" and "ISLANDS" files are missing. Also, in OSF, what are the xlsx files under the folder of Data paper JCA2020? The file structure and naming on OSF is confusing and needs to match exactly what is described in the paper. They should make it clearer in the paper and use more informative names instead of just "Table 01.xlsx" to guide readers or users.

This paper meets most of the requirements for publishing. It is well organized with detailed archaeological contexts at the beginning, followed by data information, but the current version has some incorrect descriptions that are required to be solved. I recommend acceptance after the corrections of the issues mentioned above.