The Digital Archive of Huhugam Archaeology: Crowd Sourcing User Needs
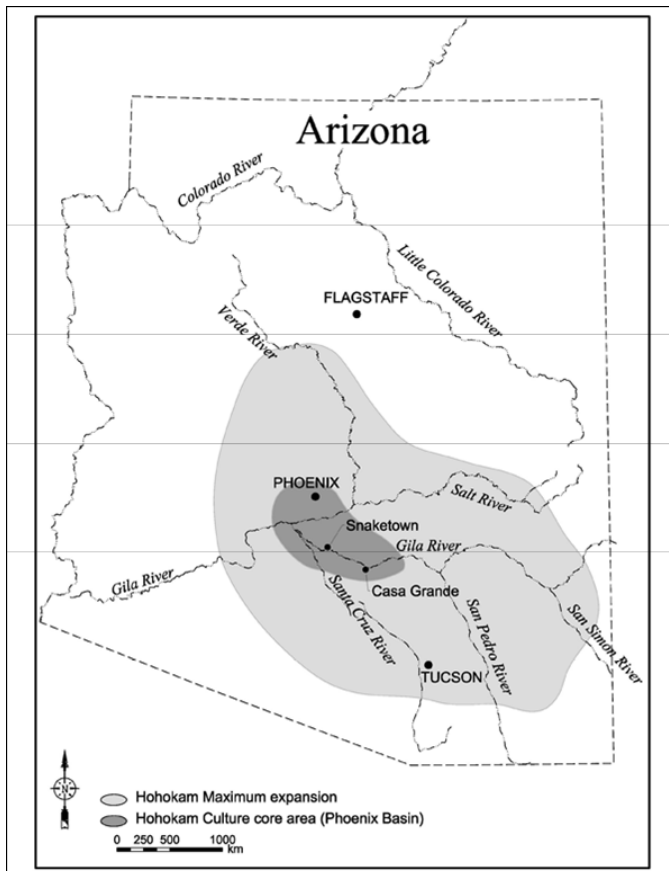
Keith Kintigh

Introduction

In July, 2017, Arizona State University's Center for Digital Antiquity, in partnership with the Amerind Foundation, was awarded a two year National Endowment for the Humanities grant (award #PW-253799-17) to build the Digital Archive of Huhugam Archaeology (DAHA) within tDAR, the Digital Archaeological Record. DAHA will be a comprehensive archive of reports on Huhugam (Hohokam) archaeology. When finished, we expect the archive to include the full text of more than 1,600 substantial documents from archaeologically significant locations in Arizona and northern Sonora and Chihuahua, Mexico. The grant also supports the development of natural language processing tools that can be used to stimulate and enable innovative uses of the Archive that advance scholarship. For additional information about the project, please visit the DAHA Project website: https://daha.tdar.org/

In year one of the grant, the DAHA team designed a survey to assess the relevant information-related needs and desires of the Archive's key user communities: archaeologists and others working in cultural heritage management who are concerned with Huhugam archaeology. The survey questions were focused on two areas: what research questions are of most interest to the user communities, and what IT tools and technological support would enhance and expand the user experience with the DAHA digital library in tDAR.

The survey confirmed our beliefs that there is a perceived need for DAHA and that the archive will be heavily used by Huhugam archaeologists. The survey's responses on how archaeologists use reports and what features they want to see in DAHA indicate that we should focus development on features that facilitate efficient discovery of the desired documents and that allow users to find or extract specific types of information they are looking for within reports. The survey results indicate that we should continue to pursue the software development as described in the proposal.

Project Overview

In July, 2017 Arizona State University's Center for Digital Antiquity, in partnership with the Amerind Foundation, was awarded a two year National Endowment for the Humanities grant (award #PW-253799-17) to build the Digital Archive of Huhugam Archaeology (DAHA) within tDAR, the Digital Archaeological Record. DAHA will be a comprehensive archive of reports on Huhugam (Hohokam) archaeology. When finished, we expect the archive to include the full text of more than 1,600 substantial documents from archaeologically significant locations in Arizona and northern Mexico.

The Huhugam were inhabitants of the Sonoran Desert region in central and southern Arizona and northern Sonora, Mexico, from approximately 1 A.D. to 1450 A.D. They built sophisticated irrigation canals, developed specialized craft production, utilized extensive trade networks, and inhabited large, long-lived towns. The DAHA archive will provide scholars with crucial, long-term data for comparative studies, indigenous communities with access to a wealth of research on ancestral populations, and the general public with a reliable, valued resource to learn about this fascinating ancient culture.

The Survey

The User Needs Survey was conducted to better understand the research needs of our principal users: archaeologists and others working in cultural heritage management who are concerned with Huhugam archaeology. The 16 question survey instrument was developed using Qualtrics software and tested in house before distribution (see Appendix 1 for survey questions and answers). The survey was reviewed by Arizona State University's Institutional Review Board (IRB) and was formally exempted from human subjects concerns on 3 October 2017.

We sought to distribute the survey to all Huhugam archaeologists for whom we could easily obtain email addresses. We distributed the survey to the 177 members of the Arizona Archaeological Council (the AAC is the professional organization of archaeologists working in Arizona) with email addresses listed as of May 2017, understanding that a substantial fraction of

the membership does not work in the Huhugam area. We distributed the survey link to 28 other Huhugam archaeologists not on the AAC list, for a total distribution of 205 people. Most individuals received an initial request to participate and at least one reminder email.

Between 18 October and 30 November 2017, we received 49 anonymous responses. We were encouraged by the 24% response rate, especially given that the distribution included a substantial number of individuals for whom the survey was not relevant. The survey was designed to elicit a reasonable number of responses from the community of Huhugam archaeologists and it was successful in that regard; it was not designed to obtain a statistically representative sample.

Appendix 1 contains a list of all survey questions and responses. In cases where long and/or diverse text responses were elicited, the response table shows summary responses. The survey results are available in full in tDAR. Those documents contain the detailed responses, including the full text of the free-text responses.

Crowd-Sourced Survey Documents available in tDAR's DAHA collection:

- Survey instrument: https://core.tdar.org/document/446372/daha-user-needs-survey-instrument
- Raw Data: https://core.tdar.org/dataset/446371/daha-user-needs-survey-raw-data
- Summary Data: https://core.tdar.org/document/446370/daha-user-needs-survey-summary-data
- Report: https://core.tdar.org/document/446227/daha-user-needs-survey-report

Analysis of the Result

Survey questions and responses are listed in Appendix 1, below. The bulk of the survey (Q2-Q13) focused on two areas: what research questions are of most interest, and what IT tools and technological support would enhance and expand the user experience with the DAHA digital library in tDAR. Finally, we collected some basic demographic information using 5 questions (Q14 – Q18).

Demographic Information. All respondents identified themselves as producers or consumers of the Huhugam literature, with 95% identifying as both producers and consumers. The respondents are an experienced group, reporting an average of 30 years as an archaeologist.

About half of the respondents (49%) were private sector Cultural Resource Management (CRM) archaeologists, 16% professionals in colleges or universities, 8% government, 8% graduate students, 4% nonprofits, 4% museums, with the remaining 12% in other categories.

Research and IT Information. The first 3 questions (Q2 – Q4) asked respondents to identify what they see as the first, second and third most important questions in Huhugam archaeology. No

options were offered; instead, people described their research priorities in their own terms. Consequently, responses were rich, though diverse. We have archived the full text of each response in the tDAR documents mentioned above. Here we present summarized results for Q2 – Q4. Because many responses contained similar themes, Kintigh was able to classify and tabulate the answers, as shown in Table 1.

Table 1. *What do you see as the three most important questions in Huhugam Archaeology?*

Count	Subject
21	Understanding the End of Classic/Huhugam Collapse
16	Huhugam Connections to Descendent Communities
14	Huhugam Organization
11	Preclassic/Classic Transition
10	Internal Hohokam Interaction
9	Adaptation to Environment
7	Identity/Ethnicity/Ideology
7	Modeling/Refinement of Population
6	Methods Issues
5	Water Management/Irrigation/River Flow
5	Relevance to today
5	Subsistence & Production
4	Chronology Refinement
4	Early Agricultural to Pioneer Period
4	External Interaction – Including with Mesoamerica & Pueblo areas
3	Nature of Classic Period
3	Huhugam Origins
3	Resilience of Huhugam

The remainder of the survey, Q5 – Q13, focused on how researchers use archaeological reports and what they would like to be able to do with them. These questions were designed to elicit useful information as we develop natural language processing routines and advanced search tools in tDAR for the DAHA archive.

Question 5 asked "*What impediments do you face in getting the information you need in your work*?" and Hohokam researchers consistently reported several. Respondents indicated that they experience "significant" or "major problems" in learning of the existence of relevant reports (65%), finding copies of them (82%), with the time it takes them to obtain reports (65%), and in covering the cost of acquiring reports (46%).

We asked:

Q6 – *Once you have a copy of a report, how often do you do each of the following to find the specific information that you need?*

- Read cover to cover
- Skim or flip through the document
- Read the abstract
- View Table of Contents or list of Figures or Tables
- Use the Bibliography or References Cited
- Text Search of a Digital Document (If available)
- Other

Reports are "never" or "rarely" read cover to cover (82% of responses). Instead, 96% of respondents report "often" or "always" skimming or flipping through a report. Researchers also heavily use available ways of finding specific information in reports. They "often" or "always", use abstracts (94%), tables of contents, lists figures, and list of tables (100%), bibliography (94%), an index if available (82%), and text search of a digital document if available (82%). The patterns of usage revealed by this question indicate to us that enhanced, digital means of finding information within reports will be most useful.

Question 7 was a text note about using tDAR, including a short explanation about how to download a file.

Question 8 read "*Do you expect that you will use the Digital Archive of Huhugam Archaeology (DAHA) in tDAR?*" and the follow up in Question 11 asked in what context they would use DAHA. All respondents indicated that they would use DAHA for their job, for research outside their job, or both.

Since the response to Q8 was 100% yes, the next two questions (Q9: "*If no to Q8, why don't you expect to use it (check all that apply)?*" and Q10: *If no to Q8, is there anything that would make you more likely to use DAHA in tDAR?*") had no responses.

We wondered how people thought they would use DAHA reports, and provided these options in Q12:

- See what reports are available.
- Download copies of specific reports I am looking for.
- Search for particular information when I don't have a specific report reference.
- Extract bibliographic information.
- Literature reviews/class 1 surveys.

- Find tabular or numerical data in reports.
- View maps or figures.

Of all respondents to Q12, 80% or more indicated that they would use DAHA to see what reports are available, download copies of reports, search for particular information across reports, extract bibliographic information, and view figures or maps. Somewhat smaller percentages (59% and 71%, respectively) said they would use DAHA for Class 1 surveys and finding numerical or tabular data.

Question 13 focused on what IT tools and/or technological support would enhance the user experience with the DAHA. We classified and tabulated the free-text answers to this question in Table 2. The table also indicates which of these features are already implemented in tDAR.

Table 2. Q13: *What features would help you in using grey literature reports to advance knowledge of Huhugam society?*

Count	Feature
13	Keyword Search/Index to reports [already implemented]
12	Full Text Search [already implemented]
4	Good Abstracts/Summaries of Scope & Results
3	Master (Annotated) Bibliography of Huhugam Reports
3	Spatial Search [already implemented]
2	Extract Tables as Spreadsheets
2	Organization Search Output to Facilitate Selection
2	Topic Search
1	List of Analysis Types Reported
1	Indication if Full Text is Available in Search Result
1	Indication if Report is Peer Reviewed or Agency Approved
1	Abstract preview before download [already implemented]
1	Quick Response Time
1	Partial Download
1	Connect Tabulated Data with Associated Text
1	Integrate with AZSite
1	Voice Search

Implications for DAHA Project Outcomes

The goal of the crowd-sourced user survey was to provide feedback from the Huhugam archaeological and heritage management communities for use in developing the DAHA archive.

The results are helpful in both prioritizing the kinds of resources to add to DAHA, and for the development of natural language processing (NLP) tools. We expect to use the survey results in workshops focused on developing NLP tools, as well as in conference presentations and publications.

Questions 2 – 4 provided valuable information concerning the research questions of most interest to the user communities. Those results will likely be of interest to many Hohokam archaeologists, and will help structure the organization of the final DAHA archive, as well as provide guidelines for decisions about the most important documents to include in the archive. The data will be used internally to allocate project efforts in ways that best serve the user community.

Question 13 was the most useful for the DAHA project in terms of directing the development of natural language processing tools and adding or enhancing tDAR search and access features. The two most common requests, keyword and full-text search are core features built into tDAR from its beginning. The report abstracts are generally extracted and made available on the metadata pages as the document summary. It is not feasible for us to manually produce additional summaries. Because the report metadata includes all of the information necessary to produce a bibliographic reference, it is possible that we could produce search result-dictated bibliographies (based for example on a keyword or spatial search). tDAR already supplies full references in the report metadata and with all downloads. Like full text and keyword search, spatial search is a core feature of tDAR available from the beginning. Being able to extract document tables as spreadsheets is a challenging request that we are considering. In addition, several suggestions were offered to facilitate the process of human processing of the DATA search result.

Appendix 1: Survey Questions and Responses

Q1 – Introductory text describing the survey, then click to agree to continue – not a question

Q2 – What do you see as THE most important question in Huhugam Archaeology?

Q3 – What do you see as the SECOND most important question in Huhugam Archaeology?

Q4 – What do you see as the THIRD most important question in Huhugam Archaeology?

Count	Summary of Q2, Q3, and Q4 text responses
21	Understanding the End of Classic/Collapse
16	Huhugam Connections to Descendent Communities
14	Huhugam Organization
11	Preclassic / Classic Transition
10	Interaction - Internal Hohokam
9	Adaptation to Environment
7	Identity/Ethnicity/Ideology
7	Modeling/Refinement of Population
6	Methods Issues
5	Water Management/Irrigation/River Flow

5	Relevance to today
3	Nature of Classic Period
4	Early Agricultural to Pioneer
4	Interaction - External, Incl. Mesoamerica, Pueblo
5	Subsistence & Production
4	Chronology Refinement
3	Huhugam Origins
3	Resilience of Huhugam

Q5 – What impediments do you face in getting the information you need in your work?
- Difficulty of learning of the existence of relevant reports.
- Difficulty of finding copies of the relevant reports.
- Time it takes to get relevant reports.
- Cost of obtaining relevant reports.
- Not enough time to do the comparative or synthetic work that should be done.

#	Question	No Problem		Minor Problem		Significant Problem		Major Problem		Total
1	Difficulty of learning of the existence of relevant reports.	0.00%	0	34.69%	17	61.22%	30	4.08%	2	49
2	Difficulty of finding copies of the relevant reports.	0.00%	0	18.37%	9	55.10%	27	26.53%	13	49
3	Time it takes to get relevant reports.	0.00%	0	34.69%	17	51.02%	25	14.29%	7	49
4	Cost of obtaining relevant reports.	12.24%	6	40.82%	20	42.86%	21	4.08%	2	49
5	Not enough time to do the comparative or synthetic work that should be done.	4.08%	2	24.49%	12	34.69%	17	36.73%	18	49

Q6 – Once you have a copy of a report, how often do you do each of the following to find the specific information that you need?
- Read cover to cover
- Skim or flip through the document
- Read the abstract
- View Table of Contents or list of Figures or Tables
- Use the Bibliography or References Cited
- Text Search of a Digital Document (If available)
- Other

#	Question	Never		Rarely		Often		Always		Total
1	Read cover to cover	6.12%	3	75.51%	37	14.29%	7	4.08%	2	49
2	Skim or flip through the document	4.08%	2	0.00%	0	57.14%	28	38.78%	19	49
3	Read the abstract	0.00%	0	6.12%	3	34.69%	17	59.18%	29	49
4	View Table of Contents or list of Figures or Tables	0.00%	0	0.00%	0	40.82%	20	59.18%	29	49
5	Use the Bibliography or References Cited	0.00%	0	6.12%	3	61.22%	30	32.65%	16	49
6	Use the Index (if available)	6.12%	3	12.24%	6	55.10%	27	26.53%	13	49
7	Text Search of a Digital Document (If available)	2.04%	1	16.33%	8	42.86%	21	38.78%	19	49
8	Other	0.00%	0	0.00%	0	50.00%	3	50.00%	3	6

Other – Text Q6
Use data tables, maps and appendices
summary data tables
For the gray lit, the (SHPO) Compliance Summary
Ignore the cultural histories and environmental portions because they are almost always wrong or of poor quality, often unreferenced or poorly referenced.
check photos and tables
Read conclusions/discussion

Q7 – Text about tDAR as a repository and how to download a file – not a question

Q8 – Do you expect that you will use the Digital Archive of Huhugam Archaeology (DAHA) in tDAR?

#	Answer	%	Count
1	Yes	100.00%	49
2	No	0.00%	0
	Total	100%	49

Q9 – If no to Q8, why don't you expect to use it (check all that apply)?
- Hohokam archaeology is not my interest
- I don't use archaeological reports in my job
- I already have all the reports I need
- I don't like doing online research
- Other

Q10 – If no to Q8, is there anything that would make you more likely to use DAHA in tDAR?
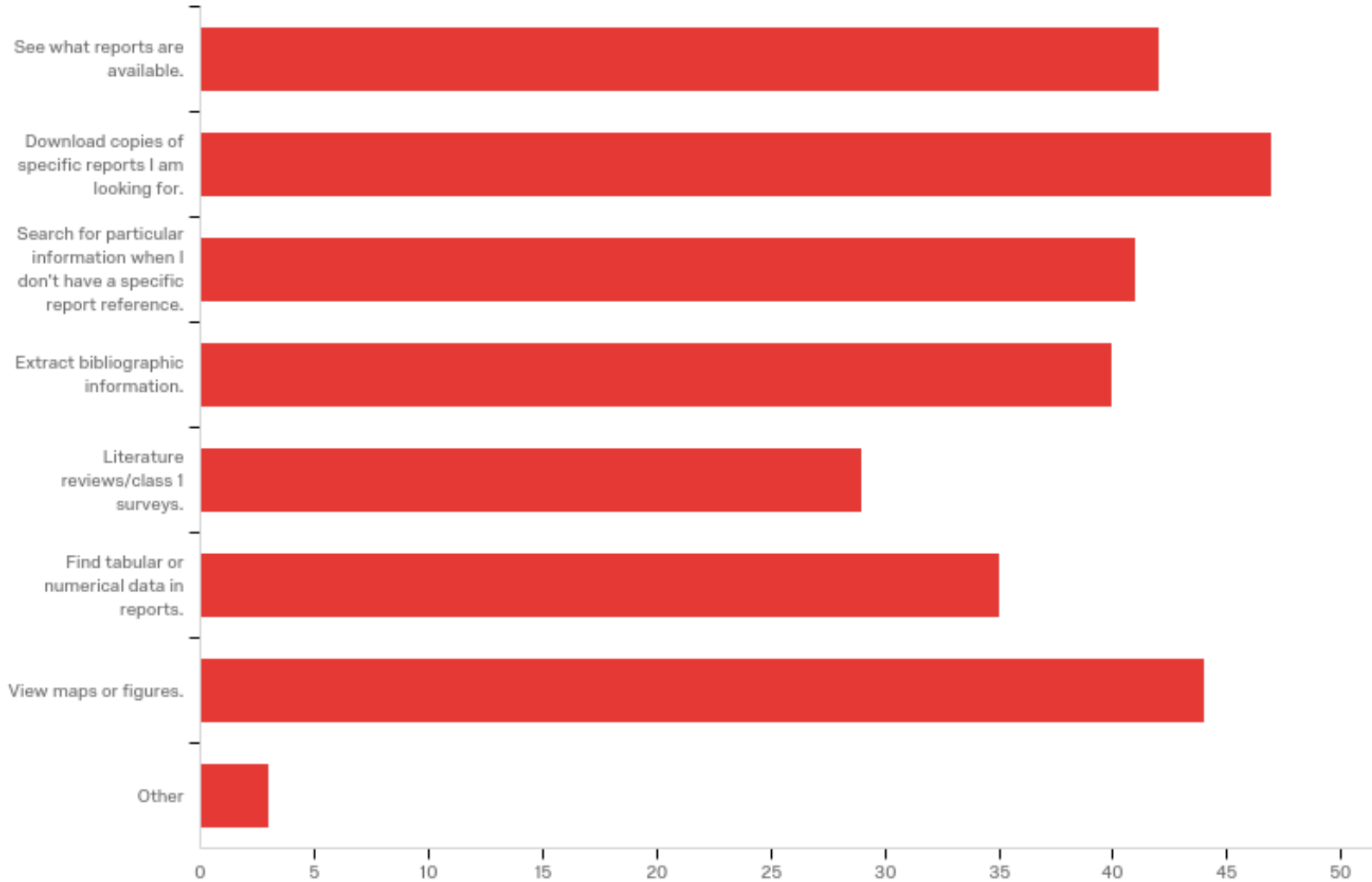
Q11 – I expect to use DAHA (check all that apply)?

#	Answer	%	Count
1	For my job	46.94%	23
2	For research outside my job	32.65%	16
3	Other (Probably Both)	20.41%	10
	Total	100%	49

Other – Text Q11
I would check both buttons but only one is accepted.
Both
Your survey does not allow multiple "checks". I would check both
The survey wouldn't let me check both job and research outside my job, but both apply.
All of the above; it won't let me select both
Both, but its only allowing me to check 1
Both
Both job and research outside my job
Probably for both my job and research outside my job
I am retired. From my point of view my "job" is research.

Q12 – How do you expect to use DAHA (check all that apply)?
- See what reports are available.
- Download copies of specific reports I am looking for.
- Search for particular information when I don't have a specific report reference.
- Extract bibliographic information.
- Literature reviews/class 1 surveys.
- Find tabular or numerical data in reports.
- View maps or figures.
- Other

#	Answer		Count
7	See what reports are available.		42
1	Download copies of specific reports I am looking for.		47
3	Search for particular information when I don't have a specific report reference.		41
2	Extract bibliographic information.		40
4	Literature reviews/class 1 surveys.		29
5	Find tabular or numerical data in reports.		35
6	View maps or figures.		44
8	Other		3
Other – Text Q12			
chronology and feature descriptions			
Information provided by tribes			

look for artifact photos

Q13 – What features would help you in using grey literature reports to advance knowledge of Huhugam society?

Count	Feature
12	Full Text Search
8	Keyword Search (existing)
5	Index to Reports
4	Good Abstracts/Descriptions
3	Bibliography of Reports
3	Spatial Search
2	Extract Tables
2	Search Output Organization
2	Other
2	Topic Search
1	Kinds of Analyses Performed
1	Integrate wth AZSite
1	Quick Response
1	Partial Download
1	Peer Reviewed
1	Voice Search

Demographic Questions:

Q 14 – Work Setting (Please check the one that best characterizes your current work setting).

#	Answer	%	Count
1	CRM Private Sector	49.02%	25
2	CRM Tribal	1.96%	1
3	CRM Other	0.00%	0
4	Government	7.84%	4
5	NGO/Nonprofit	3.92%	2
6	Museum	3.92%	2
7	College or University	15.69%	8
8	Graduate Student	7.84%	4
11	Avocational	1.96%	1
9	Primary or Secondary School Teacher	0.00%	0
10	Other	7.84%	4
	Total	100%	51

Q15 – Are you a producer or consumer of the Huhugam archaeology literature?

#	Answer	%	Count
1	Producer	0.00%	0
2	Consumer	6.12%	3
3	Both Producer and Consumer	93.88%	46
4	Neither - I'm not engaged in the Huhugam literature.	0.00%	0
	Total	100%	49

Q16 – For how many years have you considered yourself an archaeologist?

Minimum	Maximum	Mean	Std Deviation	Count
0.00	50.00	30.18	13.36	45

Q17 – Have you ever used tDAR?

#	Answer	%	Count
1	Never	20.41%	10
2	A few times	32.65%	16
3	Sometimes	26.53%	13
4	Frequently	20.41%	10
	Total	100%	49

Q18 – Have you ever contributed to tDAR by uploading a file?

#	Answer	%	Count
1	Yes	43.59%	17
2	No	56.41%	22
	Total	100%	39