Food Research and Digital Scholarship 2020; Archive Survey Results Summary

Results compiled by Libby Oaken and Christian Reynolds.

Original survey design by Christian Reynolds and Erin Spinney

DOI: 10.15131/shef.data.14039567

Background

Food has become an increasingly popular subject of study due to its inherently multidisciplinary nature. However due to this wide appeal, there is not one specific group of users who use one specific set of texts. Many cultural institutions have large collections relating to food, some of which, now fully or partially digitised, are accessible to the global research community. Prior to our 2019 community survey (10.15131/shef.data.13948154), we did not know if researchers are currently using digitised collections, let alone which collections should be given digitization priority. This project is a start to filling in these knowledge gaps by asking what (and how) scholars are currently using analogue and digital material, and how libraries and archives can better support food researchers through digitisation and activities

2019 the AHRC US-UK Food Digital Scholarship Network ran a "Community Survey" that asked what (and how) food scholars are currently using analogue and digital material, and how US and UK libraries and archives can better support food researchers through digitisation and activities (to improve accessibility). (See 10.15131/shef.data.13948154)

This ‘Archive” survey was a follow up to the 2019 community survey and was directed at curators and digitisation teams in 122+ cultural institutions. Cultural organisations are those that store and collect information for research use for example museums, libraries, and archives. We received responses from 40 archives.

The study was Funded AHRC US-UK Food Digital Scholarship network AH/S012591/1

Survey aims:

To find out more about:

1. Archive size and scope.
2. Specialised food related holdings.
3. Level of digitisation of archives generally including digitisation that enables optical character recognition (OCR) and about digitisation of food specific holdings.
4. Barriers (or perceived barriers) to digitisation.
5. Factors that may enable greater digitisation, including OCR enabled digitisation.
6. The communication channels used to engage users.

Method

The survey was carried out in early 2020 using the Qualtrics survey platform. It was distributed through email lists, twitter and Facebook. The survey was live from the (11) of February 2020 and closed on the (13) of March 2020.

University of Sheffield is the data controller. The project was ethically approved via the University of Sheffield's Ethics Review Procedure, as administered by the Geography department (Application number 030674)

Target respondents were curators and digitisation teams in cultural institutions. Specific archives were targets via twitter and direct email using the list of 122 archives and cultural organisations, these were identified through responses collected in the 2019 Food Research and Digital Scholarship Community survey – (see doi:10.15131/shef.data.13948154)

Simple statistical analysis of the data was carried out using Microsoft Excel software. The study was Funded AHRC US-UK Food Digital Scholarship network AH/S012591/1

Results

1. Characteristics of respondents and their associated archive

Respondents

There were 45 respondents to the survey and more than one respondent linked to a particular archive or institution.

A total of five respondents duplicated some data from five of the 40 archives. For example, there were four respondents from three distinct Harvard University archives, thus some data from the duplicate was consolidated to avoid double counting and to make sure that all pertinent information about that archive was captured.

The majority of respondents were the from the USA or UK which were the focus of interest for this survey however, response from other English-speaking countries Canada or Australia or New Zealand and in one case Sweden (where English is widely spoken), fig.1 were received.

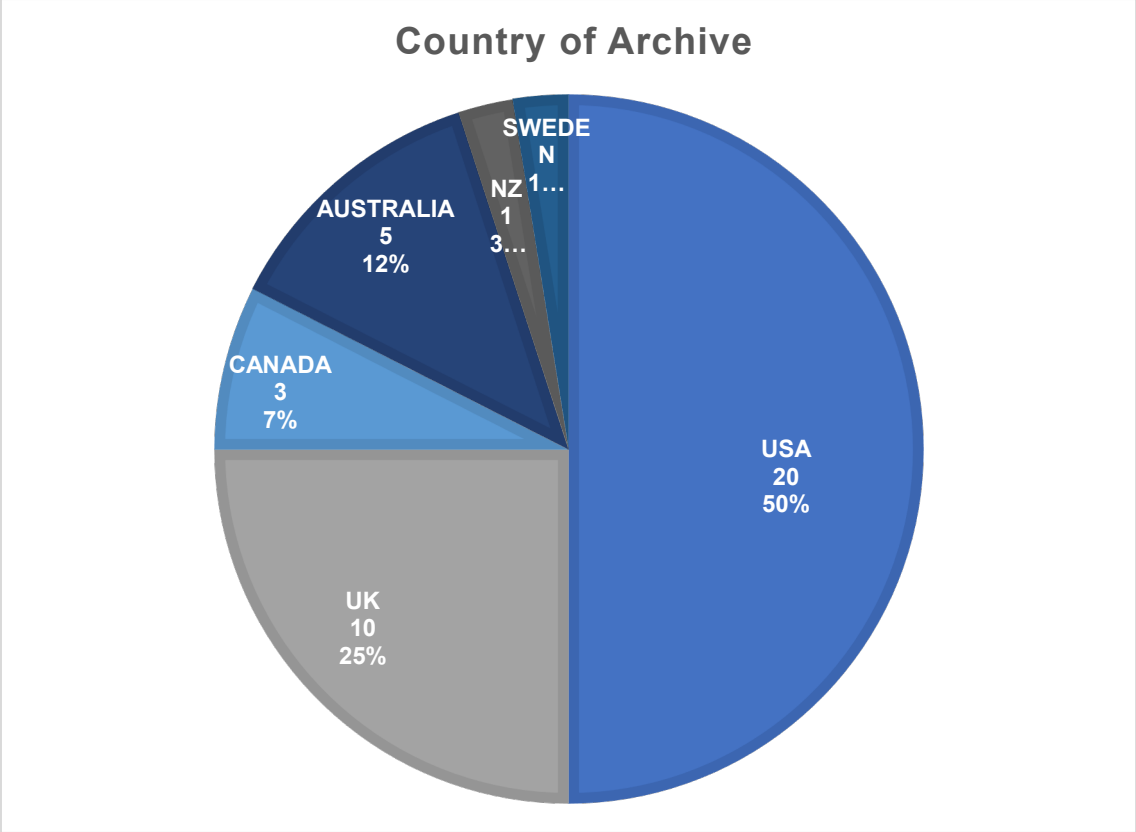

Figure 1. Archive by country (n=40) number and percentage of total archives excluding duplicates.

Types of Archive Surveyed

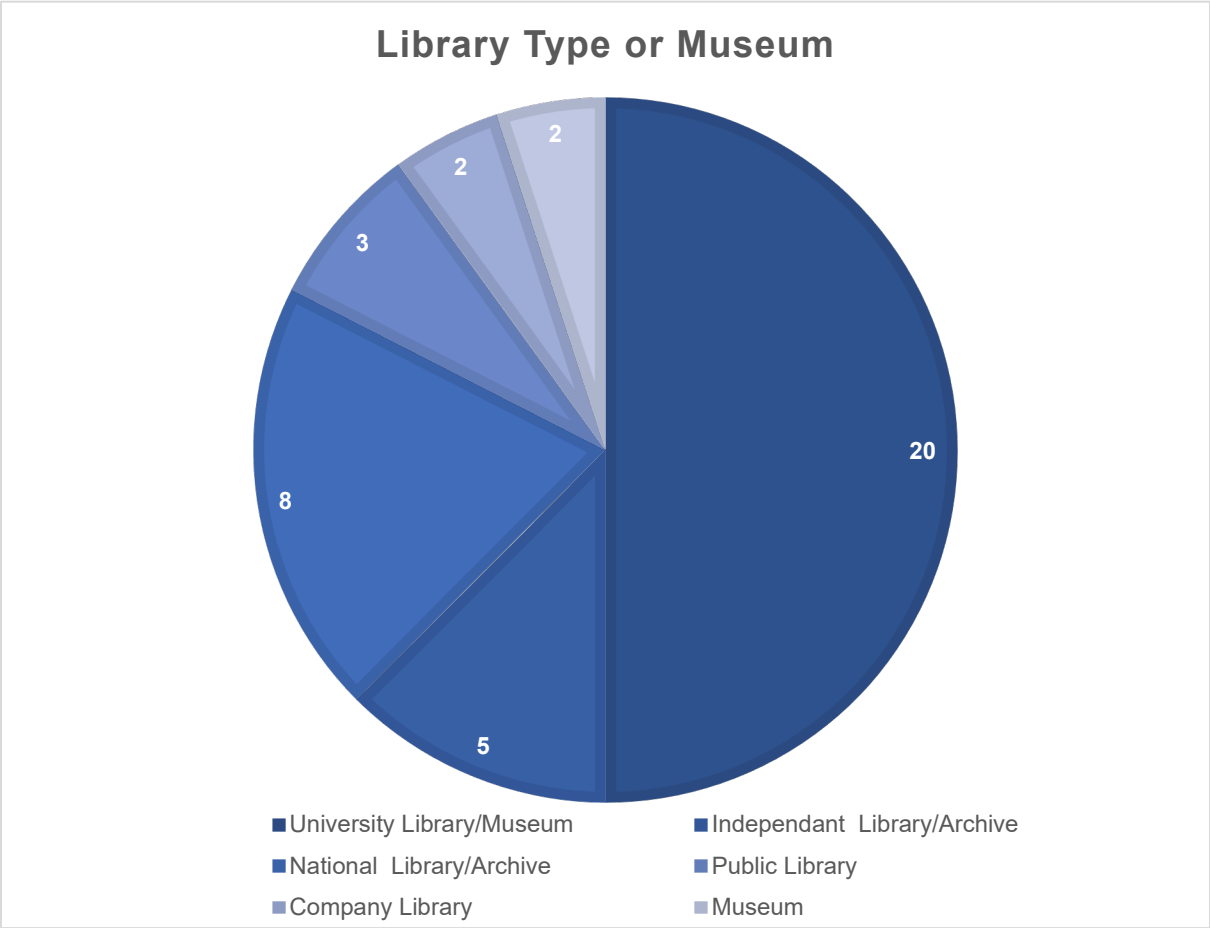

Figure 2. Type of archive (n=40) number and percentage of total.

Archives were categorised by type after the survey. The majority of archives were affiliated to University libraries, fig.2. Two of these, both semi-independent specialist archives had additional affiliations to NGO. Public libraries national archives and independent collections were most likely to be affiliated to government, or their parent NGO if independent.

2. The Archives

Materials held

There were 40 distinct/individual archives, the most frequent types of materials held were printed materials such as books magazines advertising and ephemera (37) held by 92% of archives that submitted data, fig. 3. Recipe books (27), cook books (26), photographs (26) and menus (e.g. from hotels, steam ships restaurants etc, (23) were also frequently in holdings (in 68%-58% of archives). Half (20) of the archives held

items relating to food culture and fewer held newspapers and magazines (19) or health and wellness related materials or food related objects(17). Audio visual film or video (15) was more common than sound recordings (13). Food guides were not common (13) (found in just under a third of collections) and Art was the least often present, in a quarter of archives (10) only, fig. 3.

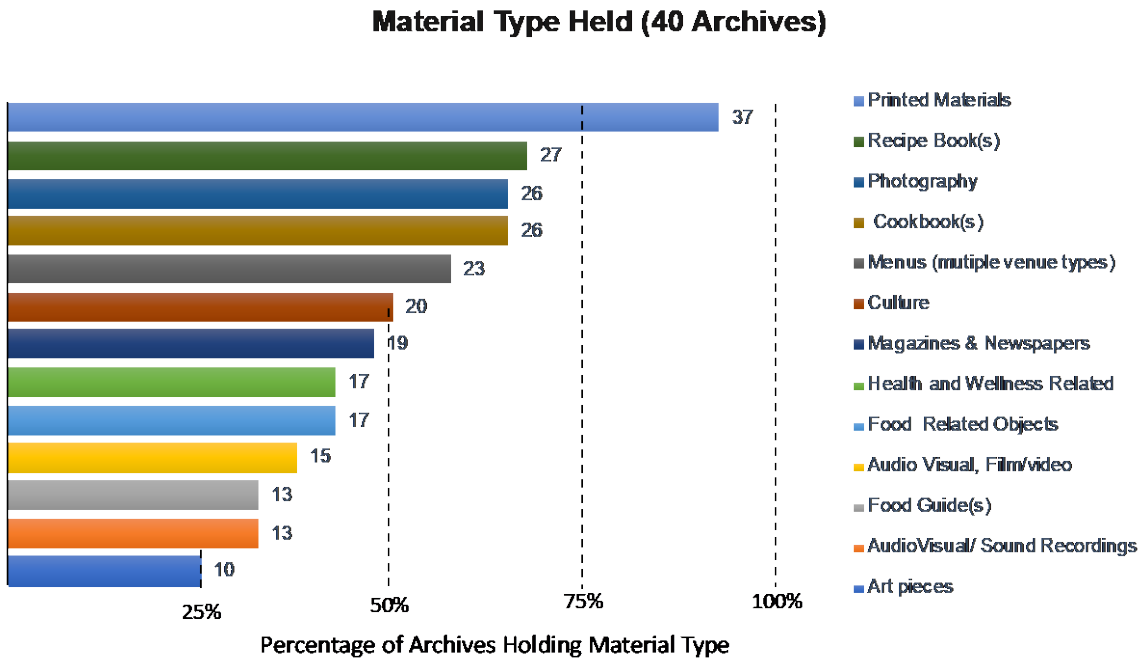

Figure 3. Frequency types of material held in archive collections (n=40). Respondents could choose multiple options.

Other kinds of specialist materials not in the predefined list that were held in unique collections included: ship plans, letters, publicity material; archaeological items, invoices & receipts for the purchase of food, archived websites for dining services, a beer collection and lastly eResources or databases.

Historical period of holdings

The forty archives, most frequently had holdings that included materials from the twentieth century. Materials from 1800-1900 were also common as were those from after the millennium (2000) whereas older materials from before 1800 less so. Only around half of the archives (21) held materials from the 1700-1800 period, fig. 4. Frequency of older holdings was lower, with less than half including materials 16000-1700 and only around a quarter from the 1400-1600 period. Five archives contained

holdings from the historical period prior to 1400 and only one contained prehistorical items.

Frequency of Food Related Holdings by Historical period

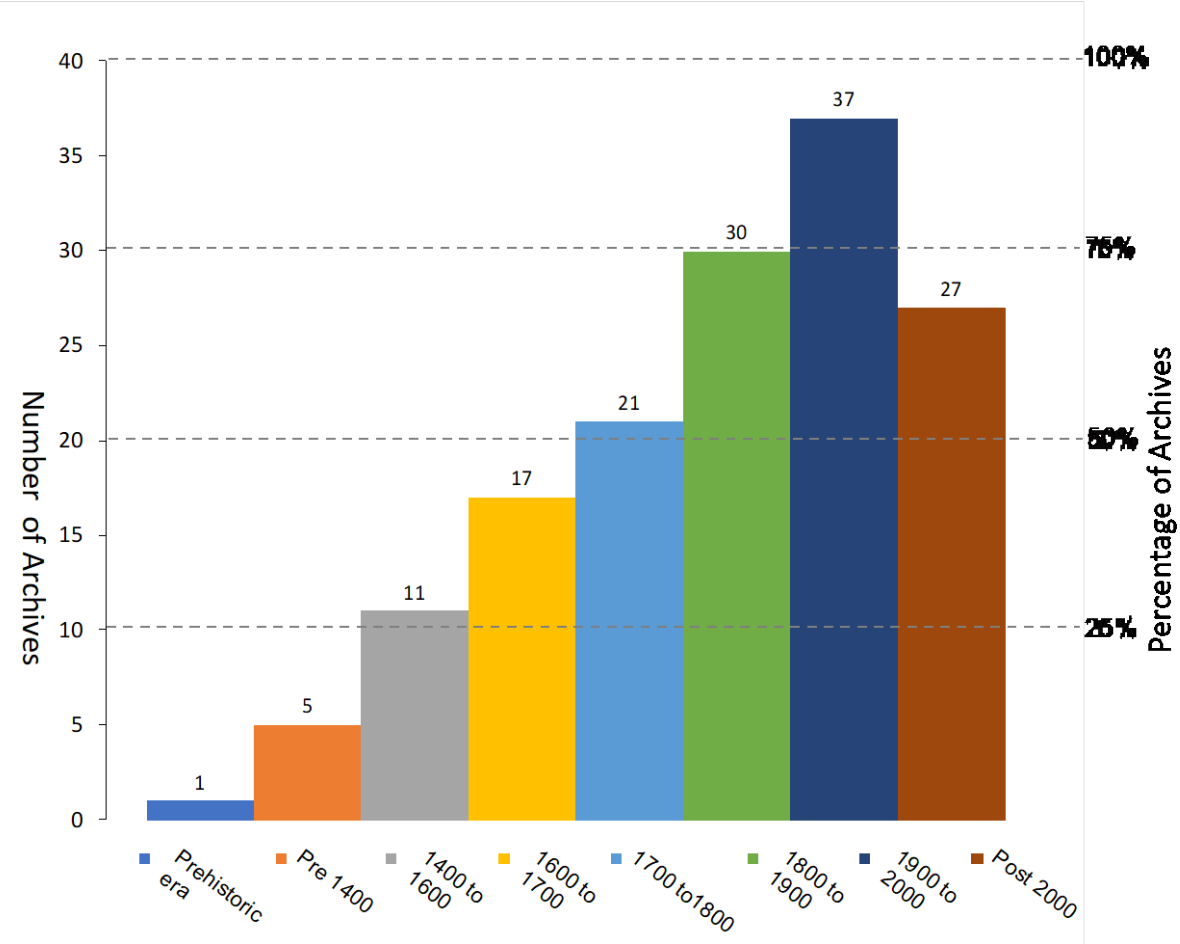

Figure 4. Frequency of food related holdings by historical period n=40

Geographic scope

The survey respondents shared information about the geographic origin of holdings from twenty archives, fig.5 (there were no duplicate respondents from same archive). There were 8 archives with only domestic collections 6 US, 1 UK and 1 Australian. Other collections (12) were global or multinational.

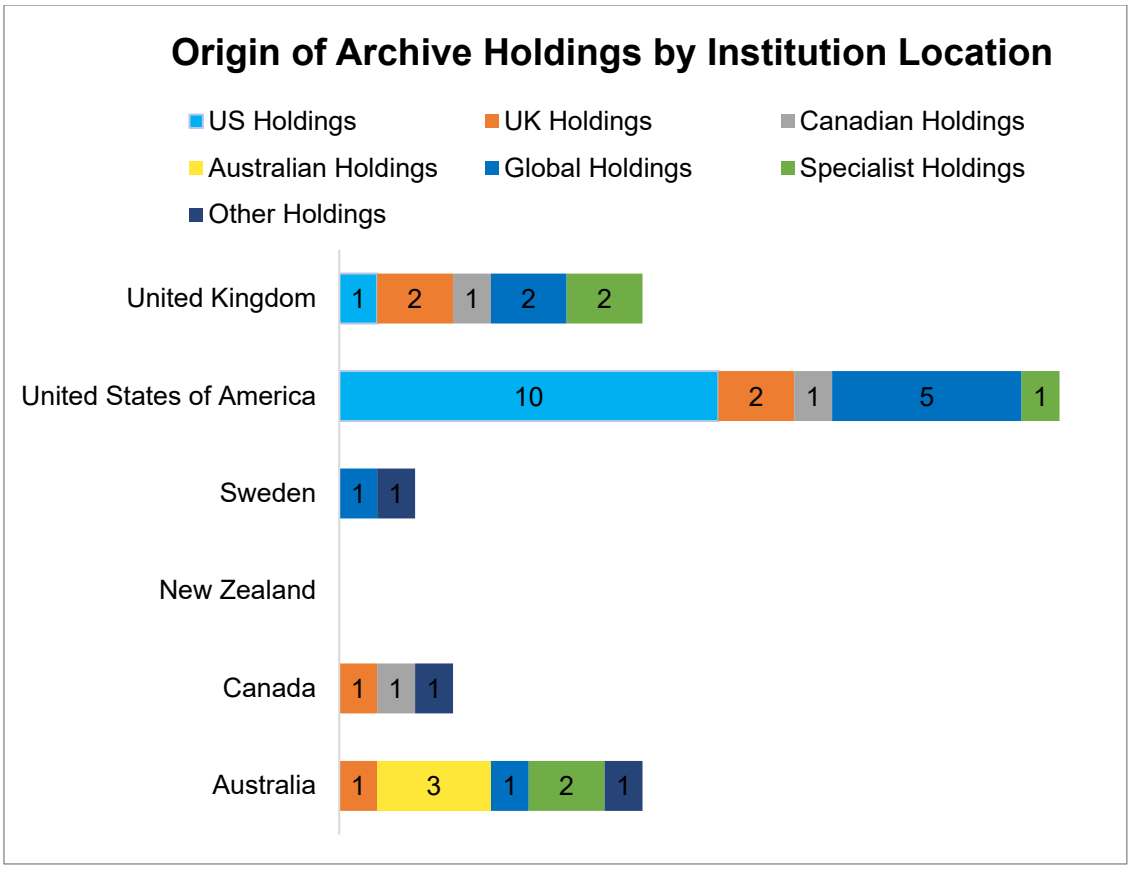

Figure 5. Geographic origin of holdings categorized by location of archive (n-20). Respondents choose one or more categories for each respondents; archive locations 10 USA, 4 UK, 3 Australia, 2 Canada,1 Sweden).

Half of those who gave geographic information (10) specified some detail about global, specialist or other holdings. Six archives that had global holdings did not specify what these global holdings were.

Specified global holdings (3/9) were described as follows:

> From one US institution-
>
> ***"Mostly America, but some European and a small amount of African and Asian materials."***
>
> By a second - *from* (sic)
>
> ***"The Netherlands".***
>
> From one Australian institution, ***"Ocean voyaging - particularity in the Pacific Ocean."***

Where survey participants shared information about **specialist holdings** (n=5) they were described as follows:

> One UK based organisation held materials described as being from the ***"Classical world"*** another on the subject of ***"Asian cookery".***
> One US institution held specialist items from ***"Mississippi, American South"***
> From two Australian organisations specialist holdings that were described as domestic in origin ***"Victorian/Australian"*** and from ***"Australia".***

Other holdings (n= 3) were described as follows:

> ***"Swedish"*** which were domestic holdings in a global archive located in Swedish university.
> A Canadian archive held ***"predominantly Canada and UK with a few early European items"*.**
> Lastly an Australian archive held ***"Australian Commonwealth Government Records"***

3. Digitisation of Archives

The question 'Have you digitized any of your holdings?' was answered yes by 40/45 respondents. Some five archives that had some digitisation had more than one respondent. Thus survey participants shared information about the degree of digitisation of their collections in 35/40 distinct archives.

Fewer respondents (21) answered some or all questions about the size and scope of digitisation of general and/or of their food related holdings. These were all responses about distinct archives, with no duplicate respondents for any one archive.

Two respondents for organisations reporting no digitisation of their collections, gave information about their (physical) holdings, one held 120 items (with less than 2% catalogued) and the other 5000 items (fully catalogued).

Of those archives with digitised holdings (n=16) the total size of collections ranged between 1400 and 40 million items (the latter being a sizable National Archive).

There were 16 archives where respondents reported food specific collections these comprised from below 1% in one case, up to 100% for some archives. One of these had exclusively food related items and no general holdings.

Archive Holdings, Food Specific Holdings and Digitisation

Table 4. Size of holdings, proportion food specific, proportion catalogued and the proportion of general and food specific collection digitised or digitised and OCR enabled. First row shows number of responses to question.

n=18	n=16	n=14	n=12	n=17	n=13	n=13	n=12
Total Objects	**Uncatalogued**	**Digitised**	**OCR Digitised**	**Food Specific**	**PropUnCatFood**	**Prop Food Specific Digitised**	**Food Specific OCR**
4000000	65,0%	4,5%	4,5%	0,08%	0,0%	10,0%	0,00%
2500000	—	0,0%	—	1,40%	—	22,9%	—
226470 9	4,4%	Yes	—	—	—	—	—
1232558	0,6%	0,0%	0,0%	0,97%	5,0%	1,3%	0,33%
1200000	16,7%	33,3%	0,0%	25,00%	16,7%	66,7%	0,00%
300000	3,3%	7,7%	7,6%	1,00%	33,3%	5,0%	4.00%
263948	1,9%	15,2%	15,2%	3,09%	24,5%	7,1%	7,10%
150000	0,1%	0,2%	0,0%	0,50%	0,0%	—	0,00%
140000	0,0%	51,4%	7,1%	5,71%	0,0%	50,0%	37,50%
82956	48,5%	26,8%	0,8%	19,62%	8,7%	0,0%	0,00%
58338	2,7%	98,2%	—	1,63%	—	48,4%	—
38738	25,8%	3,9%	0,0%	11,35%	0,0%	3,4%	0,00%
35000	85,7%	14,3%	0,0%	0,57%	37,5%	10,0%	0,00%
30000	40,0%	100,0%	0,0%	63,33%	—	100,0 %	0,00%
3000	—	Yes	—	100,00 %	100,0 %	—	—
1401	0,0%	100,0%	100,0%	100,00 %	0,0%	100,0 %	100,00%
5000	100%	No	N/A	98%	2%	N/A	N/A
120	1,7%	No	N/A	0%	N/A	N/A	N/A

The average food specific content in archives that was reported was 22%. Food related content in 8/15 was less than 5% of total archive holdings and in 4 cases between 6% and 25% of content. Only three digitised collections had significant proportion of food

related items 60-100%. A separate archive with predominantly food related items (98% of holdings) was neither catalogued and nor digitised.

The average proportion of archive digitisation was 33% for food related materials. A smaller proportion was digitised with optical character recognition (OCR) capability this averaged 12% of an archives specific food related items. However, without recognising the detail about all food holdings this average level of moderate OCR enabled digitisation is misleading. Two archives were 100% and 37.5% digitised with OCR and these skew the average, obscuring low levels of OCR enabled digitisation in archives. Of the remaining archives where digitisation of food related materials was reported 7/12 were not OCR enabled and four had low proportion (0.3 to 7.1%) of OCR enabled digital content.

The average proportion of food related items digitised 33% or digitised with OCR 12% overall was comparable/consistent with levels of digitisation for general archive holdings which were 32.5% digitised and 11% digitised with OCR respectively.

Metadata storage

Information about metadata was provided for 19 archives. Respondents were asked 'How do you store your metadata?' and to choose one or more from up to eight options.

Format of metadata relating to archive holdings

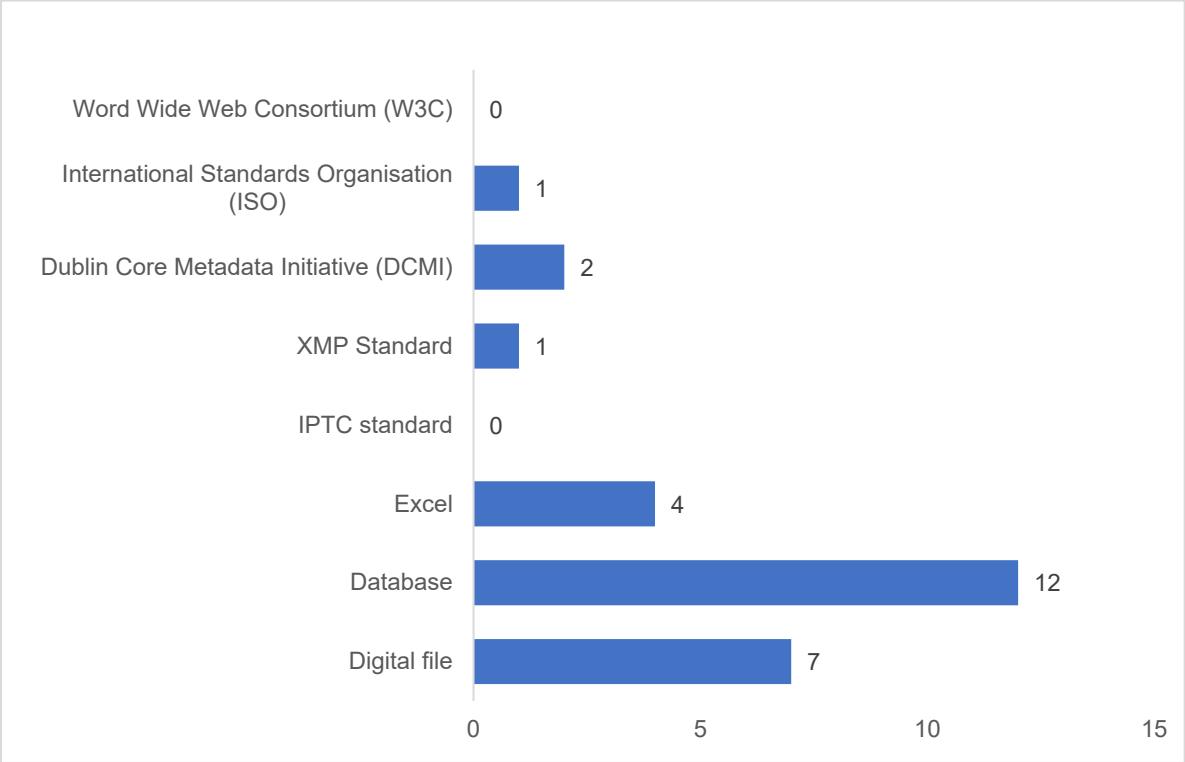

Figure 6. Incidence of metadata formats for archive holdings (n=19)

Other metadata storage formats/approaches were described in free text:

- Some metadata is pulled from the catalogue to a database (manuscripts) and other data is ingested into the Internet archive from the Catalogue
- EAD and MARC
- Library catalogue
- Preservica
- Stored in our library management system/catalogue
- Library catalog, ContentDM, Archon
- Australian Series System for Descriptive Metadata
- METS

In all 7/19 respondents indicated from the list and/or in the free text that metadata was stored according to some standardized format.

4. Barriers to and enablers of digitisation

There were 16 responses to the question is your digitised text OCR capable. There was 1 respondent who answered ‘no’, 8 ‘ answered maybe’ and 7 answered ‘yes’. There were no duplicate answers for any one archive for this question.

Of those who answered ‘yes’ or ‘maybe’, eight commented on OCR capability. Comments illustrated several possible reasons for low level of OCR capability in archives. These can be put into 3 categories:

1) Digitisation with OCR capability in some cases was dependent on technical resources:

   “yes, when we do it through the internet archive. Our in-house digitization for manuscripts does not have OCR capability”

   “Previously we were manually enabling OCR, now most readers (Adobe, Foxit) are automatically enabled so we don't manually enable as many files for productivity reasons”

2) Handwritten materials are less easy to make OCR capable:

   “If it is typescript yes but not if it is handwritten”

   “Partially (some is handwritten)”

   “Excluding handwritten manuscripts”

3) Making digitised records OCR capable was a was a more recent advance and incomplete:

   “Most is, but this has not yet been done”

   “Some of it is. Only started OCR in the last four years or so.”

One archive had OCR capable digital files, because of inclusion in a large repository managed in a collaboration between several north American Universities. https://www.hathitrust.org/ and https://www.hathitrust.org/access_use

"Digitized published material is added to Hathi Trust"

Respondents were asked to rate barriers to digitisation on a five-point scale, from small to large. Respondents from 20 separate archives answered. They were able to select more than one perceived barrier and rate each one. These answers were analysed and a mean average rating for each barrier on a scale between 1 and 5 was calculated. Table 1 & Fig.7.

Barrier	Mean rating	Rating Range	St'd Dev.	Variance	No of Responders choosing
Time	4.30	3-5	0.84	0.71	20
Staffing #	4.05	1-5	1.15	1.31	19
Cost	3.95	1-5	1.28	1.65	20
Equipment	2.22	1.5	1.07	1.15	19
Lack of Expertise	2.00	1-5	1.17	1.37	19
Complexity of materials	1.95	1-5	1.16	1.35	20
Space to digitise	1.89	1-4	0.97	0.94	19

Table 1. Rating of Perceived Size of Barriers to Archive Digitisation (n=20) but not all barriers were chosen and rated by every respondent.

Eight of the respondents cited other barriers not in the predefined list. The average rating for these other barriers was 2.75 on the five-point scale (variance 2.69) Table 1.

The highest ranked barrier was time. Time was rated as more problematic than staffing and staffing was rated a greater barrier than cost. Less highly ranked barriers, were; lack of equipment, expertise, complexity of materials and lack of space, Table , see also Fig. 7. Other barriers respondents commented on were copyright (4 respondents.) and agreement of collection trustees to digitise (1 respondent.). Priority of time and resources to digitise archived collections versus other objects was identified as a barrier by a single respondent.

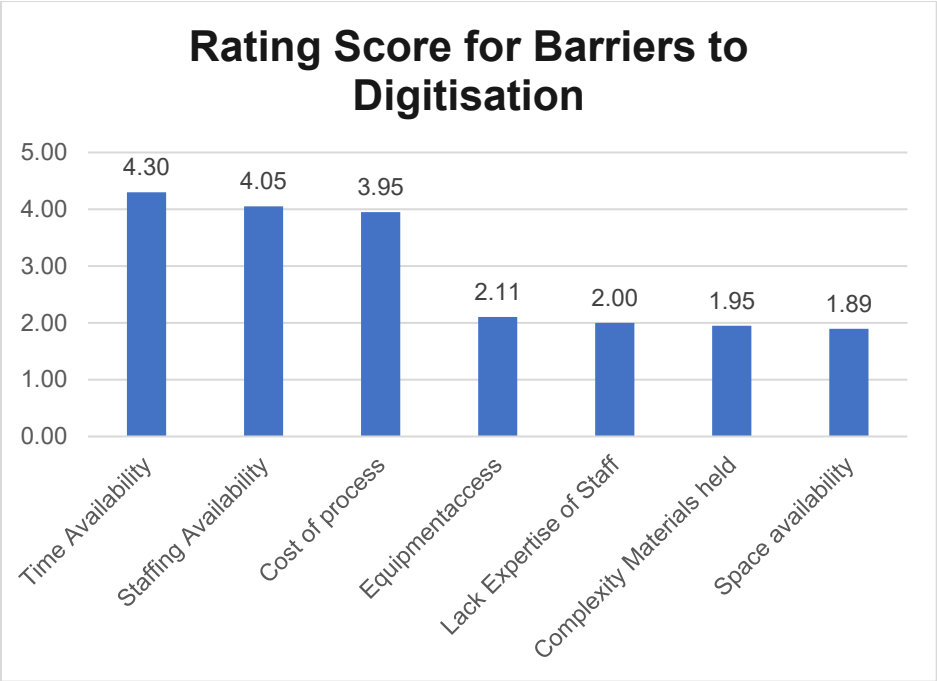

Figure 7. Rating of barriers to digitisation.

Digitisation goals

Some respondents indicated their archive had digitisation goals 10/19. Of the list provided the majority had goals relating to quality and format of digitisation (6). Additional goals set were, timelines for digitisation (3) or the proportion or percentage to be digitised (2). One archive had goals relating to timeline alone.

Goal of Quality Format	Goal of Timeline for digitization	Goal of Percentage of collection	Other Goal
6	3	2	6

Table 2. Frequency types of goal for digitisation chosen n=12; respondents could choose multiple goals.

Six of the twelve archives with digitisation targets had ‘other’ goals not listed in the predefined list. These were:

> ***“Priority is Victorian material as we are deposit/reference library for state of Victoria”***
>
> ***“Want to have enough materials that it makes sense to create a digital collection***
>
> ***“Out of copyright items also not available in digital form elsewhere”***
>
> ***“Digitize cookbooks prior to 1970 published in Mississippi. Digitize food ephemera.”***
>
> ***“Accessibility of collections to researchers”***
>
> ***“goal is to digitize all of the manuscript cookbooks in time for an exhibition last year, and to digitize the printed materials as we come across them in other focused areas (Canadian imprints etc.)”***

Five of the six who listed other goals were specialist collections and the sixth was a local archive.

5. Funding

Funding for archives came from a variety of sources.

The most frequently reported funding sources were state or national funding (11) and internal institutional funding (11).

Rank: funding sources by occurrence

Funding Source	Count of responses	Rank
National/federal government	11	1
Internal institutional funding	11	1
Provincial/State/County Council government	6	3
Private donors	6	3
Academic societies	3	5
Local government	2	6
Cultural groups/societies	2	6

Table 3. Incidence of reporting sources of funding, one or more could be chosen from the predefined list (n=20 archives).

No respondents gave religious organisations as a funding source which was an option on the predefined list. Other funding sources cited but not on the list were: collection (by) trustees, commercial income and internal budget**.**

Digitisation specific funding (15 archives provided information about this) came mainly from university (7) or government (7) other sources were other institution (1) Leventis Foundation (1), Grant bodies for specific projects only (1) and through internal staffing resources (1).

6. Accessibility

The survey provided some accessibility information about 20 of the 35 archives that reported a level of digitisation. Eighteen collections were identified as searchable online and 13 were open access, with 7 of these searchable collections having search

engine interoperability. Fewer collections were not open access (4) and were sat behind paywalls.

One respondent said the archive they were linked to was both searchable and not searchable. Investigation after the survey suggests that some collections that particular archive are digitised and a proportion of these digitised collections are searchable on line, without institutional login requirements.

The interoperability of catalogues with search engines was not homogenous; no two archives cited interoperability with the same search engine. Compatibility with the following engines was reported:

1) Regular search engine
2) National Archives of UK (Discovery)
3) World Cat and Archives Hub
4) ArchiveGrid (OCLC), Google for some metadata
5) Digital Commonwealth and thus DPLA
6) Trove (National Library of Australia)

A seventh archive that claimed search engine interoperability did not specify the engine.

7. Networks and communication channels

In answer the question 'Do you belong to a research network/cluster?' only three out of seventeen of those who responded to the question had membership to a network. These were to: 1) the AHRC US-UK Food Digital Scholarship Network, 2)the School of Advanced Study, and 3) to ASERL (Association of Southeastern Research Libraries).

Communication with users or audience was via a website in 18/18 cases (where this question was answered). Additionally, communication route(s) was via one or more social media channels such as Twitter (13) Facebook (14) or Instagram (12) or via a blog (10).

Respondents provided information about what they wanted to communicate having been asked "What do you feel researchers don't know about your holdings and

services that you wish they knew?" Respondents chose one or more options from a predefined list, fig. 8.

What Respondents Wished Users Knew About Their Archive

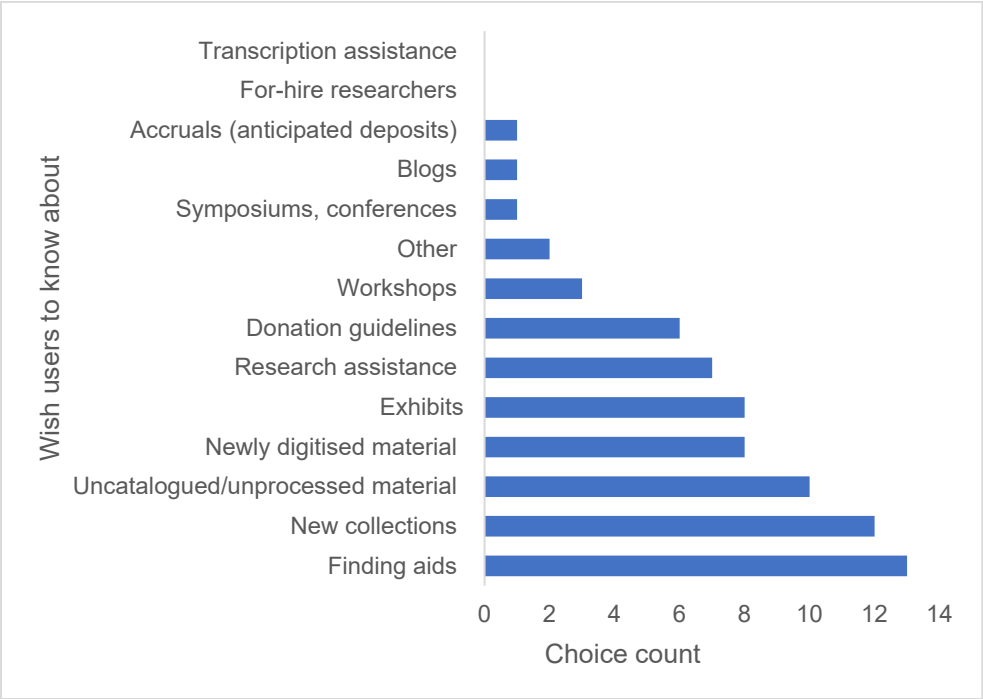

Figure 8. Frequency of answers chose to the question 'What do you feel researchers don't know about your holdings and services that you wish they knew?' Respondents could choose multiple answers, responses from 18 archives.

When describing other aspects not in the list one respondent was specific about accessibility:

> ***"Our system is complex, and we aim through research guides to support inquirers in their research tasks. The research guide is the chief way into our food related collections, canned searches mean new material is discovered when a link is clicked on."***

Another gave the thought-provoking answer ***"That we exist.".***

These free text answers echoed the most frequently chosen options from the predefined list either that relating to ease of finding materials (finding aids) or to improving user knowledge about archive content ; newly digitised material, uncatalogued/unprocessed material, new collections.

Discussion/Commentary

Scope/limitations

The survey attracted respondents from 40 archives from the 122+ approached. Not all survey participants gave answers to the full question set many were answered by half survey participants or fewer. Specific information about content, including size of food specific holdings were not always known by respondents. Some non-participating organisations wrote to communicate (via email) reasons they did not participate including not knowing precise information about food related holdings in their collections, and therefore did not engage further.

This survey provided a convenience sample and was not designed to capture information about all organisations with food related items in their archives. It does provide pilot data and an indication of the scope of food related collections and some of the perceived barriers and enabling factors for increasing levels of digitisation and/accessibility in (predominantly) UK and US based cultural institutions.

The survey did attract a majority of participants located in the target countries USA and UK with three quarters of archives. The remaining quarter of the archives represented were from a wider selection of organisations located in English-speaking countries, mostly in former UK colonies. The one outlier was a response from a participant from a Swedish University.

Funding

The predominance of Government and institutional funding reflects large proportion of national cultural organisations and those associated with universities or independent foundations. Similarly, most funding specific to digitisation efforts (when described) came from governmental or internal institutional sources.

Holdings

The variety of content revealed in terms of geographic origin would appear to potentially reflect colonial history in the particular country the archive is located. The Holdings in the US cultural institutions in this sample appeared to be skewed toward materials with domestic origin. Half the US organisations and over half of these that

gave information about the geographic sources had US sourced holdings only (6/10). A larger study would be required to determine whether this domestic focus was really predominant in the majority of US based archives.

The numbers in other nations who reported geographic focus of collections was relatively low. Whilst there was some variety in geographic source of holdings of non-US institutions broader conclusions cannot be drawn.

This sample suggests that institutions are more likely to hold newer materials most frequently nineteenth and twentieth century. Older holdings are reported in fewer cultural institutions and there is a decreased likelihood of materials being held the older they are. This may be a function of the rarity of older materials and/or the specialist conservation requirements that may be needed to maintain them. Contemporary holdings post 2000 are held in a similar number of archives to the nineteenth century materials i.e. fewer than hold twentieth century collections.

Forty archives that participated, provided responses about the to the range of material types of their archival holdings. The more specialist holdings were present in fewer archives (e.g. Audio-Visual film/video). Most archives held a variety of kinds of food related documents, fewer holdings included: physical objects, food guides, AV materials both film and sound and fewest archives held art. This may be a reflection of the large proportion of libraries and document archives in the sample.

Digitisation

The level of digitisation of catalogues was variable. Similar levels of digitisation and OCR enabled digitisation were found for both general and food specific collections. This suggests that lack of digitisation in food related collections in archives is not related to their level of importance.

Survey participants provided information on what barrier they thought were important by rating them. The most highly rated barriers to digitisation were staffing time and cost which are interrelated capacity issues. Equipment expertise and space are were less highly rated barriers, but may be mitigated by allocating more resources to the problem. Other than allocation of resources one potentially rate limiting barrier may be the format of materials. Where these are handwritten high quality capable

digitisation OCR may be difficult to achieve. Some databases/services use human readers to transcribe and translated to overcome technical limitations of automatic methods.

Accessibility

Collections with interoperability to search engines were reported by few respondents 7 only from 40 archives. This suggests the availability of information about archives for users may be constrained.

In answering the question 'What do you feel researchers don't know about your holdings and services that you wish they knew?' The participants who responded (18/40 archives) identified that they wish users/audience knew more about content in their collections and how to find information about content. We speculate that that the seemingly low level of search engine interoperability and information gap between archives and their users are could be linked issues.

Communication

Communication with users/audience was most commonly via a websites (18) and some social media presence was reported by these archives (17). However, most organisations who participated in the survey (22/40) did not respond to the question about communication channels. Three archives only, gave affirmative answers about belonging to networks. There was no question in the survey about effectiveness or perceived effectiveness of particular communication channels. If we had a wider/larger sample, we could have performed a social media engagement analysis of the food community and the archives they use to understand how they interact. However the sample size is currently too small.

Together the answers to the accessibility questions and communication channels may suggest that there is great potential for an improvement in accessibility and communication with users/target audiences by archivists and their cultural institutions.

This survey provides an indication of the range of holdings and levels of digitisation and the balance of the constraints one making collections more accessible. A larger survey that targeted archivists at a larger number of cultural institutions would be

required to fully survey and compare the characteristics of UK European and US holdings.

The companion survey to this contacted food researchers and presents some data from the food researcher community that are complementary to what is presented here. See 10.15131/shef.data.13948154 for this community survey.

COVID-19

This survey was run at the start of 2020, and was not overly impacted by national lockdowns. However, it is worth highlighting the impact of the COVID-19 pandemic, and rapidly increasing need for digitised content for research as physical access to resources became restricted through social distancing measures, lockdowns etc. We hope that many archives have implemented better digital access as part of the response to COVID-19, however assessment of archive response and offering to COVID-19 was outside the scope of this survey.