Chapter 9

Crowding Out the Archivist?
Locating Crowdsourcing within the Broader
Landscape of Participatory Archives

Alexandra Eveleigh

A working understanding of crowdsourcing has evolved amongst archives professionals from a combination of practical experimentation with participatory web tools and platforms,[1] and rather more theoretical speculation about the transformative, democratising potential of such technologies.[2] The term crowdsourcing may then be loosely, and is often retrospectively, applied to almost any initiative in the field which seeks to engage users to contribute to archives or to comment upon archival practice through the medium of the Internet. For example, crowdsourcing has been used by archivists to describe a public consultation exercise regarding archives policy,[3] collaborative appraisal and collection development work,[4] and a volunteer scanning programme.[5] However, the word has perhaps come to be particularly associated with user involvement in archival description, transcription and metadata enhancement. In the field of archives, as in related information and cultural heritage domains, there has been a specific experimental focus upon the potential for users' contributions to be employed to help address acknowledged problems in the description and representation of collections.[6] These connections are the particular focus of this chapter on crowdsourcing in the archives domain.

Access to, and use of, archival source materials depends to a large extent upon the availability of appropriate and effective access routes. Traditionally, these access systems have taken the form of textual descriptions, typically created by a single professional archivist. Commentators have noted that, while archivists

---

1   For case studies and examples, see Theimer, *A Different Kind of Web*; also Theimer, *Web 2.0 Tools*.

2   For instance, Flinn, '"An Attack on Professionalism and Scholarship"?'; Yakel, 'Who Represents the Past?'.

3   Theimer, 'NARA Crowdsourcing Classification Reform'.

4   Pennock, 'Twittervane'.

5   Theimer, 'Fantastic Volunteer Scanning Project'.

6   For example, for this issue discussed in a museums context see Karp and Lavine, *Exhibiting Cultures*.

agree that archival description is important, there is considerable disagreement over what description is *for* (broadly dividing into those who focus on the *record*, and those who focus on *users*), the best descriptive method to implement and the point(s) in time at which description should take place.[7] Thus understanding of the professionalised process known as 'arrangement and description' is seen to be a 'fraught terrain',[8] containing the tension inherent between a custodial instinct to control context and authenticity, and a desire to share access and promote usage. This fault line is deepened by the prospect of user participation in the descriptive process, since involving others in description seems inevitably to weaken the archivist's control over the process but at the same time seeks to magnify the accessibility of the descriptive product.

However, given that most archive organisations struggle with significant cataloguing backlogs, the idea that users might create, or supplement descriptions, has a clear, practical appeal. Currently, descriptive 'finding aids' or catalogues to archives are often incomplete, or realised at an insufficient level of detail to satisfy the information-seeker's needs, or in some cases are simply non-existent. Even where finding aids are available, they are rarely updated to reflect new information, interests or perspectives. As such, catalogues may as easily prove a barrier as an enabler of access. This is a consequence both of practical collections management issues – cataloguing failing to keep up with the pace of new accessions[9] – but also of a professional compromise; a:

> resolution that [the products of description], in order to be useful for every kind of research, had to serve none in particular. Thus, all documents had to be described in equal depth, independently of their 'importance' for one kind of research or another, and the descriptions had to emphasize context and function rather than content.[10]

Crowdsourced description promises a solution to these dilemmas, enabling description – even transcription – of content to take place at a detailed level of granularity across a broad range of subjects and collections.

The term crowdsourcing makes an early appearance in the archival professional literature in 2008 from Isto Huvila who explicitly conceives of his concept of a radical participatory archive as something which goes *beyond* crowdsourcing: 'Even though a participatory archive is about crowdsourcing, it focuses on deeper involvement and more complex semantics rather than on larger crowds

---

7   Yeo, 'Debates about Description'; Hurley, 'Parallel Provenance'; Duff and Harris, 'Stories and Names'; Duranti, 'Origin and Development of the Concept of Archival Description'.

8   Duff and Harris, 'Stories and Names'.

9   Greene and Meissner, 'More Product, Less Process'.

10   Duranti, 'Origin and Development of the Concept of Archival Description', 52.

and simple annotations'.[11] This enhanced sense of user participation in archives being necessarily associated with 'a comprehensive shift in archival thinking and practice'[12] remains a strong current in the archival literature, in contrast to a shallower engagement in crowdsourcing comprehended as 'letting some others to (*sic*) play with (some of) *my* toys in *my* sandbox'.[13] The rhetoric here is strongly informed by constructivist philosophies, in particular perhaps by a relatively late flourishing of postmodernist critique as a creative influence upon the development of archival theory.[14] Participation in this vein is promoted as a means to address troubling issues of marginalisation and representation, professional passivity and power. Web 2.0 technology is then harnessed to this argument as a facilitating mechanism for achieving the vision of the archive as a community in a continual process of becoming, celebrating a multiplicity of different perspectives, meanings and contexts, and castigating the singularity of the authoritative, professional voice:

> Web 2.0 is about connection, collaboration, community. For archives, Web 2.0 connects communities with collections or, maybe even more conceptually, communities with their history and identity. What is more, it invites collaboration about that history: what it means, how it should be presented, and what we know. Shared authority and distributed curation are the point. Yet Web 2.0 technologies can be implemented and the community can be integrated in many different ways. This raises questions about how much authority we as archivists are willing to share and how to manage the voices of all those distributed curators.[15]

A growing number of essays by archivists take up this line and extol the transformative potential of the social web for 'opening up' both archival content and the archive profession 'for the people'.[16] But the actual word 'crowdsourcing'[17] still appears only infrequently in the formally published archival literature, although it has gained greater traction in the professional blogosphere as a convenient shorthand. Instead, a host of alternative terms have been coined to describe internet-based user involvement and participation projects in the archives domain, including the Participatory Archive, Archives 2.0, Citizen Archivists, the Archival Commons and Citizen-Led Sourcing. Most archival commentators too have preferred a notion of community (translated online) over 'the crowd' as the conceptual model for online user participation – supporting a greater

---

11   Huvila, 'Participatory Archive', 27.

12   Theimer, 'What Is the Meaning of Archives 2.0?', 58.

13   Huvila, 'What Is a Participatory Archive?', emphases in original.

14   Kaplan, '"Many Paths to Partial Truths"'.

15   Yakel, 'Who Represents the Past?', 258.

16   Anderson and Allen, 'Envisioning the Archival Commons'; Evans, 'Archives of the People'.

17   Or 'crowd sourcing' or 'crowd-sourcing' – the editorial confusion perhaps bears witness to archivists' uneasiness about the use of this term.

Crowdsourcing our Cultural Heritage

degree of self-regulation and project ownership amongst participants, who, it is acknowledged, may well be experts in their own particular field.[18] And responding to the disappointment of some early experiments in online participation in archives which succeeded in attracting only nominal levels of engagement, the tendency has been to chide archivists for clinging to an archive-centred worldview; for their reluctance to share control and build equitable partnerships with these user communities – pushing further still at this agenda of anticipated professional revolution assisted by Internet technology.[19] Crowdsourcing in contrast has been characterised as merely a technologically enhanced version of volunteering, limited in its scope and anticipated impact upon professional practice.[20]

Trevor Owens (Chapter 12) chooses to retain the crowdsourcing buzzword, but also acknowledges crowdsourcing's debt, in libraries, archives and museums, to long-standing traditions of volunteering. He contends that most crowdsourcing projects in cultural heritage contexts 'have not involved massive crowds and they have very little to do with outsourcing labour' (p. 269). But rather than paint the small numbers of engaged participants as indicative of a professional failure to embrace change and cede control to the community, he suggests instead that the key to success lies in 'inviting participation from engaged members of the public' (p. 269), passionate amateurs who already identify with a particular professional mind-set. Owens' conception of crowdsourcing then, like Huvila's participatory archive, offers an opportunity for citizens to engage deeply with cultural collections and to contribute in meaningful ways to the 'public memory'. But this engagement is instead portrayed as a recognition and reinforcement of established identities through the incorporation of additional user knowledge into the existing professional domain of practice, rather than necessarily a source of innovation and creativity via an encountered heterogeneity of external opinion.[21]

Yet the hope of tapping into diverse new audiences, on a scale unimaginable without the Internet, making niche areas of history and research accessible and capturing the interest of a wider public – these remain the inspiration, if not perhaps always the reality, for many crowdsourcing-type ventures in archives and heritage practice contexts.[22] Existing models of crowdsourcing in archives (and the humanities more generally) that posit close-knit communities continuing the volunteering tradition online seem inadequately flexible to incorporate this larger-scale ambition, or those instances where the participants' online interactions may only be fleeting or serendipitous (for example, comments added to digital images or catalogue entries encountered during research or in browsing the web).

---

18 Flinn, '"An Attack on Professionalism and Scholarship"?'; Palmer, 'Archives 2.0'.

19 Yakel, 'Who Represents the Past?'.

20 For example, the Archivist of the United States, David S. Ferriero, 'Crowdsourcing and Citizen Archivist Program', speaks of 'crowdsourcing or microvolunteering'.

21 Owens, 'Digital Cultural Heritage and the Crowd'.

22 Dunning, 'Innovative Use of Crowdsourcing Technology'.

Furthermore, crowdsourcing in archives at least *is* sometimes viewed as a wholly pragmatic solution to a lack of financial and labour resource: 'crowdsourcing may help institutions faced with dwindling budgets address resource constraints by involving interested participants in the process of contributing metadata. […] If the experience engages participants and they value it, the "labor" involved in the exchange can be considered a voluntary, in-kind contribution'.[23] Again, this circumstance is poorly served by existing conceptualisations of crowdsourcing and online participation, since it involves neither innovation in, nor in-depth engagement with, existing professional practice, but rather the reallocation of some of the more tedious or repetitive parts of the descriptive process to unpaid human-computational effort.

Mapping the Participatory Landscape

Despite the growing popularity of crowdsourcing and participatory practice in archives, there have been few attempts systematically to map and evaluate this landscape. The results of practical initiatives in the sector have been mixed, in spite of what one commentator calls the 'triumphal rhetoric'[24] of participatory archives culture. Whilst some projects report apparently runaway success (the *Old Weather* project, discussed in Chapter 2, in which participants transcribe meteorological observations from historic ships' logs, reported 685,000 log pages transcribed in one year, for example), others have struggled to attract the anticipated rich seams of user knowledge, and several have quietly closed or transferred their content onto 'read only' websites.[25] The neologism 'crowdsourcing' also promotes an impression of transience, a passing fad; of participatory practice as merely an exercise in wanting to be *seen* as working at the cutting edge. This in itself may be limiting the potential of some participatory projects, contributing to an organisational reluctance to support and sustain their development beyond the pilot testing, experimental stage, and later, to a tendency to move on quickly to the next technological platform without a proper review of what worked or what did not. Crowdsourcing initiatives in archives, as in related professional fields, are also haunted and constrained by the fear that a contributor might be wrong, or that descriptive data might be pulled out of archival context, and that researchers using collaboratively authored resources might somehow swallow all of this without question or substantiation, in what has been described as a 'fundamental change in

---

23   Flanagan and Carini, 'How Games Can Help Us Access and Understand Archival Images', 536.

24   Palmer, 'Archives 2.0'.

25   For instance, the planned closure of the National Archives' (UK) Wiki site, Your Archives, went ahead during 2012. http://webarchive.nationalarchives.gov.uk/20121030162709/http://yourarchives.nationalarchives.gov.uk/index.php?title=home_page.

the relationship between researchers and archivists [and] between the records and researchers that leaves out archivists'.[26]

The ideological impetus which fuels much of the theoretical debate about participation in archives, both off- and online, can be equally unhelpful when it comes to evaluating practical initiatives and planning future ventures in this area. If crowdsourcing is an opportunity to democratise professional archival practice and to promote the active participation of the general public in co-creating historical meaning, how should a project be judged which fails to attract large numbers of new users, or where the interaction is brief or ephemeral? How and when should participants' contributions be integrated with professionally authored resources such as the archive catalogue, particularly when the contributions received are emotive or illustrative in contrast to the traditions of archival description? And what role does the professional archivist have to play in this democratised archive, and who is responsible for supporting or verifying the accuracy and reliability of contributed information?

A User Participation Matrix

This chapter seeks to analyse the variety of online participation practice in archives through four frames. These four frames, or quadrants, come together to form a proposed matrix of user participation (Figure 9.1). The borders between the frames are fluid, but together the four quadrants provide a conceptual map to help make sense of the ambiguities and contradictions, ideological inclinations and diversity of configurations observed in contemporary crowdsourcing and other online participation initiatives in archives. The aim of the matrix then is not to provide a definition of crowdsourcing in the archives domain, but rather to set out a framework through which existing practical initiatives can be assessed, particularly in terms of their influence on archival professionalism. Should success in archival crowdsourcing be gauged in the same terms as online outreach or volunteering? Or if crowd initiatives can be distinguished within a broader concept of participatory practice, how might this insight affect the design of future ventures which seek to reach out to the widest range of participants?

Nor is the matrix proposed as a strict classification of current programmes, for any single project may exhibit characteristics from more than one frame simultaneously. For instance, many transcription projects combine a directed primary task structure (the Transcription Machine) with a participant-managed forum to boost intrinsic engagement, and to facilitate learning amongst the members of the participant group (a Collaborative Community). But, since ultimately, achieving the objective of any participatory project is entirely dependent upon the response the project receives from participants, matching the intended outcomes to contributors' likely motivations and exhibited patterns of behaviour is vital

---

26   Yakel, 'Balancing Archival Authority', 77.

for the success of any project. The matrix then provides a tool through which to examine these various points of interaction between the project organisation (the upper and lower halves of the framework) and the participants (the left and right hand sides of the matrix).

The upper and lower halves of the framework represent contrasting approaches to the structure and management of online participation projects, adopting Burns and Stalker's classic distinction between 'mechanistic' and 'organic' styles of organisation.[27] Here, this spectrum pertains to the structural coordination of online participation, whether focused upon specific goals and objectives (mechanistic) or taking a more flexible and open-ended approach (organic).
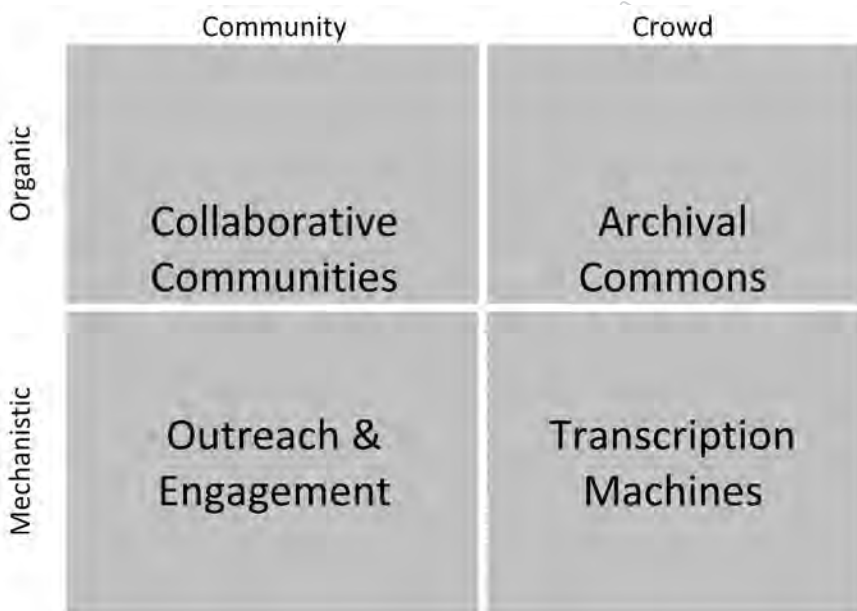

Figure 9.1   A user participation matrix

It is important to note that whilst a mechanistic structure might be assumed to represent directed management by a formal institution, there are examples in participatory practice where a mechanistic approach has been designed and implemented by the participants themselves. Many programmes with a genealogical focus, for example, are entirely volunteer-initiated and run, but many of these are coordinated under a formalised command and control structure (for instance, the long-running *FreeBMD* project uses a syndicate system, with appointed local coordinators and separate data teams for checking accuracy). Other

---

27   Burns and Stalker, *The Management of Innovation*.

projects operate as consortia led by external professionals and subject specialists, but have similar tiers of responsibility for separate parts of the processing of data. In the *Old Weather* project, for example, a specialist in naval history coordinates the extraction of chronological ships' histories from the log data transcribed. The role of the custodial institution in such partnerships varies greatly: in some instances, the organisation is a formal project partner, and archivists and other heritage professionals are involved in the design or testing of the participation interface or in supplying expert contextual knowledge on archival sources; in other projects, the archives organisation may be merely a supplier of source material or digitised content. Clearly the impact of crowdsourcing upon professional practice will vary according to the specifics of such partnership arrangements, and also with the individual employer's tolerance for professional autonomy. A risk-averse organisation may act as a restraint or deflection upon the potential for any transformation in the role of the professional archivist, perhaps unintentionally endorsing an overly cautious approach, in order to maintain the organisation's reputation or traditional position of authority.

The left and right sides of the participation matrix are created from Caroline Haythornthwaite's 'crowds and communities' peer-production spectrum, representing the motivations and behaviours of online participants.[28] 'Crowd' in this model does not necessarily refer to large numbers of people, but instead relates to the relative strength of the social ties between participants, being strong in a community, but becoming weaker as the continuum line moves towards the crowd. This conception of a crowd retains the all-embracing sense of scale which underpins the ambition to reach out to infinite new audiences, but also allows for comments and encounters of a more serendipitous nature from participants with no previous connection to a particular organisation or set of archive documents, or indeed to each other.

The Archival Commons

The Archival Commons metaphor characterises user participation with a strong ideological bent. This is an understanding shaped significantly by the popular rhetoric which promises a relentlessly positive social transformation on a dispersed, global scale through engagement with Web 2.0 technologies. Consequently, it is sometimes dismissed as a utopian and romantic vision.[29] Nevertheless, it is a vision that continues to be a dominant influence in shaping and understanding crowdsourcing in archives, and within the cultural heritage sector more generally.[30]

---

28   Haythornthwaite, 'Crowds and Communities'.

29   Schafer, *Bastard Culture!*.

30   For example, Smith-Yoshimura and Shein, *Social Metadata for Libraries, Archives, and Museums*; Zarro and Allen, 'User-Contributed Descriptive Metadata for Libraries and Cultural Institutions'.

A specifically archival vision of the Commons idea has been put forward in some detail in an eponymous 2009 article in the *American Archivist*. In essence, the concept could be summarised as an all-encompassing, postmodern, archival ecology. The article's authors envisage 'a decentralized market-based approach to archival representation' (elsewhere referred to as a 'democratic culture'). They anticipate a shift away from a professionally regulated, 'singular arrangement' of archives towards a malleable, continually evolving descriptive practice reflecting the 'constantly changing views and meanings' of archives. This is a vision of archives for a global, interactive society; an emergent, organic orientation representing a 'sea change in how users engage' with archives online.[31]

As its 'crowd' position in the user participation matrix indicates, this 'distributed but integrated' mode of production is envisaged to operate at a cross-repository, cross-domain, 'web of connectivity'[32] magnitude, reaching 'thousands of potential volunteers'.[33] This sense of infinite scale is seen in the promotional texts used for archives' participation initiatives: an inclusive, welcoming vocabulary – explore, share, collaborate, contribute; an ambition to reach as many people as possible, particularly the elusive 'new users'; and an awareness of archives' wider contexts. Recent developments with Linked Data (structured data that can more easily be linked to other data sets) are also beginning to put in place a plausible technological underpinning to the Commons concept, providing the elasticity required to serve the varied and unpredictable demands of a technologically astute Interactive User Community[34] and extending the shareable, extensible, flexible principles of the Commons into the realm of open data reuse.

But for the most part, whilst the Archival Commons remains a source of inspiration, it seems current practice is more constrained. The Commons concept relies substantially upon users being willing and able to participate *and* on archivists being prepared to accept their contributions. This would appear to imply that the envisaged contributors to the Commons (researchers, historical society members and students are specifically mentioned) have, like Owens' crowdsourcing volunteers, some prior experience of the archival domain, and feel motivated to impart their knowledge in a public space. Unfortunately, the dispersed, global nature of the Commons could militate against both of these characteristics by increasing the likelihood of incidental participation from individuals who have encountered archives by chance online, who may not contribute what archives organisations expect to hear or express themselves in ways in which archivists want to hear it. Archivists become caught between the conflicting trajectories of an imagined radical professional transformation and the defence of their employer's reputation. This is not merely a matter of inappropriate language or the pedalling of historical myths and falsehoods, as much as a misalignment between a professional

---

31   Anderson and Allen, 'Envisioning the Archival Commons', 384–90.

32   Ibid., 389.

33   Evans, 'Archives of the People, by the People, for the People', 395.

34   Anderson, 'Necessary but Not Sufficient'.

understanding of description and users' more often emotive and personal response to archives. It appears too that the very openness of the Commons – the weakness of social connections amongst the crowd and the consequent fragility of trust between them – may act as a barrier to participation by precisely those more expert users whom archivists had most hoped to attract. Furthermore, for all the avowed intent to create an open, inclusive space, the Archival Commons concept still reserves a particular centrality for the archives repository as a memory institution. Certainly practical experience of crowdsourcing initiatives inspired by the Commons ideal is already suggesting that 'if we build it', we cannot assume that 'they' will come.[35]

Outreach and Engagement

As an alternative then to building bespoke platforms, some archivists have turned to external social media services, such as Flickr and HistoryPin, to provide a space for user participation, and to furnish access to communities beyond individual archives' immediate, local audience boundaries. Elizabeth Yakel notes how engaging with these third-party services shows 'the initiation [of archives] into and understanding of social norms in these peer production systems'.[36] Such ventures are rationalised as an exercise in taking archive material out to a place where an interested audience already exists, and have much in common with traditional audience engagement and marketing initiatives, extended in reach and ambition by means of the internet. Sometimes, in-person outreach events may be incorporated into online projects as a means of building community around the archival content.

Yet for all their avowed intent of dipping into spaces inhabited by users, these outreach-type projects take a strong strategic steer from the archives organisation, with participation taking place according to a planned timetable and organisationally defined remit. This mechanistic notion of project planning and management also necessitates clearly defined objectives (rather than outcomes emerging according to the participants' interests), and the impact of the participation is bounded in terms of time-scales, carefully selected archive content and target user communities.

Approaching a tightly knit community of interest with such a mechanistic approach to project coordination can lead to some structural friction. Community engagement strategies may aspire to a bilateral exchange between professionals and participants, but the boundaries between 'us and them' remain substantially intact. Contributions are treated as supplemental rather than fundamental, since crowdsourcing in this vein seems to require a bedrock structure of professional description onto which participants are invited to add embellishments. And since participation in this outreach mode is generally staged in spaces apart from the main archive website, the results of such crowdsourcing projects are often poorly integrated with finding aids and other organisational web resources. This restricts

---

35   Palmer, 'Archives 2.0'.
36   Yakel, 'Balancing Archival Authority', 86.

the impact that the interaction with new users might otherwise have had upon either professional practice, or upon established researchers who have no reason to encounter the contributed information in the course of their habitual work routines.

Nevertheless, success in this style of participatory practice is still contingent upon professional sensitivity to the user environment, in order to be able to identify, and negotiate relationships of trust with suitably motivated participant communities. Participant 'energy' is sought to promote the sustainability of the archival enterprise by widening the pool of advocates for the activities of archives. A professional renewal then, if not a professional re-birth, this outreach form of participatory practice is a natural response to turbulence and complexity in the external operating environment, to economic pressures and to public policy and legislative shifts which challenge archivists to show strong leadership in shaping the future of the profession towards greater openness and flexibility.[37]

The role of the archivist here seems to be that of intermediary between organisation and target community, or a boundary gatekeeper maintaining the archivist's position of authority.[38] Editorial control in these outreach enterprises is usually reserved to a staff moderator, and contributions may even be rejected according to professional judgements about suitability and pertinence. But this role can only be performed successfully if archivists also operate within a new degree of empathy for the participants' points of view, and accept responsibility for a deeper involvement in interpreting and presenting the archival record. Yakel argues that the authority claimed here is a kind of cognitive influence – the archivist and archives institution acting in concert as a proxy for personal knowledge of the accuracy of archival finding aids and the authenticity of the records described therein.[39] It is authority which many users indeed may be willing to recognise in the archivist, since it implies no 'right to command', and also lessens the filtering and verification burden on research user. But it is also vulnerable to allegations of censorship, even where the archivist operates a relaxed moderation policy.

Collaborative Communities

More rarely, a more thoroughgoing remodelling of archival practice is sought which aims to break down, or at least redraw, the boundaries between archivists and participants. To achieve this shift, both archives organisations and the archival profession actively seek to embrace uncertainty in the environment, as a catalyst towards new ways of working, scanning the horizon for newly emergent

37 Morgan, *Images of Organization*. A useful summary of recent developments towards greater openness and accountability relating to the UK archives sector can be found in Dacre, *Review of the 30 Year Rule*.

38 Duff et al., 'Finding and Using Archival Resources'; Hedstrom, 'Archives, Memory, and Interfaces with the Past'.

39 Yakel, 'Balancing Archival Authority'.

opportunities. This frame sees archival practice in a state of evolution and flux, indicating a greater depth of change than the professional renewal brought about through outreach and engagement projects.

In particular, before archivists can participate equitably in a collaborative community with participants, they must first address any cultural issues over sharing knowledge within their own domain. The literature is critical of the high visibility of archivists, rather than users, on some public participation sites.[40] But an alternative reading might see this as an important staging post in the transformation of archival practice, in learning to share archival knowledge more openly, and in new and more adaptable ways. In this way, archivists can begin to identify areas where current professional processes and services fail to meet user needs. The next step is to entrust specific user communities to help resolve these issues, by reorienting participation opportunities around the intrinsic interests of the diverse communities of practice which already surround the archival record (for instance, family historians, geographers, economic historians might use the same source, but in different ways).[41]

When user communities are invited to input into the *processes* of participation in this way, as well as to contribute content, the results are no longer restricted by the established structures of acceptable professional archival practice. The professional role is reoriented away from a mechanistic focus on strong archival leadership and hierarchically determined goals towards a new emphasis upon facilitation, dispersed community coordination and emergent design. This may demand unaccustomed levels of professional humility: recognising that archivists too make errors, and welcoming dispute and debate around the contents of a catalogue in 'always beta'.[42] Yet it is by handing over some responsibility for the maintenance of community norms and standards, and for the direction and sustainability of the site of participation, that archivists seek to resolve the tension of cognitive authority encountered in outreach initiatives.[43] Participation can then begin to move beyond a channelled exchange of supplementary descriptive information towards a deeper understanding of historical sources as genuinely new knowledge and unanticipated discoveries emerge from the network of (redundant) connections. The greater freedom granted to participants within a collaborative community can also lead to the creation of new descriptive services, such as visual finding aids or 'mash-ups' using archival data. [44]

---

40   Ibid.; Palmer, 'Archives 2.0'.

41   Wenger, *Communities of Practice*.

42   Yeo, 'Debates about Description', 102.

43   Yakel, 'Balancing Archival Authority'; Wasko and Teigland, 'Public Goods or Virtual Commons?'.

44   See, for example, the *Guardian*'s visualisation of data from the *Old Weather* project: http://www.guardian.co.uk/news/datablog/interactive/2012/oct/01/first-world-war-royal-navy-ships-mapped.

Transcription Machines

If outreach-style participation is to defend the professional boundary, Collaborative Communities seek to redraw it, and the Archival Commons to dissolve it, a fourth option is to reinforce it. Rules and structure are imposed from above to ensure consistent, standardised input (and output). Quality control becomes a matter of consecutive processing up through a hierarchical chain of command, combined with double (or sometimes triple) entry, statistical sampling and automated error detection. The emphasis in a Transcription Machine is on bureaucratic or administrative control over user input, which is characterised by the reductive nature of both the participative task and of contributors' commonly fleeting commitment to that task and to each other. Participants may shun opportunities to contribute beyond the basic data input task (for instance, only a small proportion of registered members of projects like *Old Weather* are regular contributors to the project forum). The issues of attracting participants, and of motivating and rewarding performance may even be implemented as a competitive game.[45] This mechanical image of user participation can even be extended, metaphorically and also literally, into the ways in which archival metadata can be released for use through the structured delivery mechanisms of APIs (application programming interfaces) and Linked Data.

The impact upon the professional role here is not transformation so much as extension or translation of function. The enforced consistency can be viewed as a continuation of the international standardisation of archival descriptive practice, extending control over input standards down to a micro-level which enables this type of crowdsourcing to operate across consortia of different organisations and subject interests. Meanwhile, responsibility for the actual process of data input transfers to the users, the archivist taking on more of a coordinating role, released from the drudgery of the routine and freed up to concentrate on tasks demanding a greater degree of professional skill.

But this shift might also harbour a hidden threat to professionalism. The bureaucratic nature of the authority wielded here is not relative to a particular sphere of interest or expertise,[46] and hence does not of necessity have to be exercised by professional *archivists*, and the reductive nature of the transcription task is easily dismissed as beneath the professional dignity of an archivist. Archives organisations have outsourced responsibility for many such projects to a range of external delivery partners, ranging from the entirely volunteer-led and managed, to subject specialist consortia, to commercial enterprises such as Ancestry. So whereas one benefit of the more community-focused forms of user participation is the advocacy role that such projects can play in raising the profile of professionalised archive services, participants in outsourced transcription machines may be disinterested or

---

45   For example, Flanagan and Carini, 'How Games Can Help Us Access and Understand Archival Images'.

46   Wilson, *Second-Hand Knowledge*.

simply unaware of any link to a formal repository or of any input of expertise made to the project by professional archivists. Furthermore, restrictive licensing deals or simply a lack of foresight over data rights can also lead to a loss of *archival* control over the extensive quantities of descriptive metadata generated by such projects.

This is a particular issue in the context of the increasing prominence of open data and the potential for 'big data' computational analysis to transform historical research using archives. As the role of the archivist shifts away from the sole authorship of description, a new opportunity or imperative opens up in respect of descriptive information retrieval: linking together the multiple representations and contexts of each archival asset, and devising new tools for filtering, searching and understanding the historical world: 'Gatekeeping of information resources *shifts from contribution to retrieval*. When "anyone" can post to the web, the value is in being retrieved'.[47]

Conclusion

Writing about the relationship between museum computing practice and the emergent theory of digital heritage, Ross Parry has observed that 'commentators … have too easily adopted the posture of either advocate or sceptic'. He identifies a need to find 'more nuanced ways of thinking and communicating' which resolves this polarity between the advocate's enthusiasm for professional revolution on the one hand, and the sceptic's fear of professional demise on the other.[48] This chapter has proposed a framework to support such an analysis of crowdsourcing in the archives domain, given that crowdsourcing initiatives sit within a broader landscape of participatory practice similarly moulded by the intersection of theory and practical experimentation. Using the matrix to reflect on the strengths and weaknesses of current practice, it becomes evident that existing community-focused theories of crowdsourcing as a deeply engaging, collaborative participant experience, are not sufficiently elastic to accommodate the task-driven, individual involvement of the majority of contributors to a Transcription Machine, for instance, just as innovation in professional practice has not inevitably resulted from the targeting of specific expert communities in outreach and engagement programmes.

Yet the encounter between profession and participants might still be productive even where it is not transformative. Susanne Justesen distinguishes between learning, which she defines as 'more knowledge about an existing domain', and innovation, which 'is about the exploration and creation of new domains'.[49] Hence an outreach and engagement project might prompt the diffusion of existing, but perhaps latent or particularly specialist, knowledge, whilst a basic Transcription

---

47   Haythornthwaite, 'Crowds and Communities', 8, emphasis in original.

48   Parry, 'Digital Heritage and the Rise of Theory in Museum Computing', 343.

49   Justesen, 'Innoversity in Communities of Practice', 84, citing Katz and Lazer, 'Building Effective Intra-Organizational Networks'.

Machine is designed to augment professional knowledge with additional layers of detail which cannot (currently at least) be extracted from manuscript source material algorithmically by computer.[50] These are both examples of learning, but may leave professional practice itself relatively unscathed, since participation is channelled towards a pre-determined output or an outcome designed to complement rather than change established professional work methods. In the context of citizen science, Bonney et al. have termed such initiatives *contributory* projects, as contrasted with collaborative or co-created projects, where participants have more input into the design of the work they undertake.[51] In some instances, particularly in Transcription Machine projects, outsourcing a task to the crowd may even substitute for paid labour, but this is generally justified as freeing up professional time to focus on more expert phases of the overall process. In the *Transcribe Bentham* project, for instance, participants carry out the initial transcription work which might previously have been undertaken by an editorial assistant, enabling staff to concentrate on the more detailed work of preparing the transcriptions for formal publication.

Such contributory projects are often conceived and understood as enriching a pre-existing barebones informational structure about a particular set of historical sources, whereby supplementary knowledge is 'pulled' from the participant and embedded within the professional domain of practice. Yet the transfer of knowledge described here as learning may equally operate in the opposite direction, and relate to participants' initiation into professional norms of practice – such as learning the expert language used to describe a medieval charter, perhaps, or the terminology of the nineteenth-century Poor Law. In other instances, the professional role might lie specifically in providing learning materials, for example supplying expert collections knowledge to external consortia.

Justesen additionally connects the complexity of the knowledge exchange which takes place to the strength of the ties between participants. The homogeneity of a tightly bonded community is said to facilitate more complex learning or innovation[52] – in the case of crowdsourcing in cultural heritage, perhaps around sensitive topics or issues of some historical or technical intricacy – whereas the diversity of weakly connected individuals who make up the crowd are well placed to diffuse less specialised knowledge or more widely appealing content into a range of different external contexts or in ways novel to the archival and related professions. An example might be the PhotosNormandie Flickr project which deliberately replicated (described as 'liberated') out-of-copyright images

---

50   Quinn and Bederson, 'Human Computation', would site the task of transcribing handwritten data in the intersection between crowdsourcing and human computation, allowing for some work to be undertaken by a single person in isolation as well as in the context of a group.

51   Bonney et al., *Public Participation in Scientific Research*.

52   Justesen, 'Innoversity in Communities of Practice', citing Katz and Lazer, 'Building Effective Intra-Organizational Networks' and Hansen, 'The Search-Transfer Problem'.

outside of the professional custodial context in order to boost access and interaction with the photographs.[53]

In all cases of learning, the host sphere of practice ('competence regime') remains strong, absorbing new knowledge into the established domain rather than being challenged by it.[54] In contrast, more innovative outcomes are achieved when the participants' outsider perspective is able to influence an established specialist or professional domain to shift in new directions – although learning is also a pre-requisite for innovation, and vice versa, so this process is cyclical and iterative. One example of this might be the committed community moderators and super-contributors in a Transcription Machine who make suggestions for project enhancements and adaptations. Other participants may wish to use the transcribed information in their own research, or to explore and experiment with new ways of presenting and understanding the data. Many of these innovations will be in themselves just small-scale redefinitions of practice on the boundary between professional and participant communities, but together they can contribute towards some broader shifts of perspective – Justesen labels this process 'incremental innovation'.[55] Included here might be the growing appreciation amongst cultural heritage professionals of the sustained effort required to motivate and support contributors, moving well beyond a simple understanding of participatory practice as simply 'a means of designing a better and more user-friendly finding aid or of crowd-sourcing metadata in an era of diminishing resources'.[56]

More radical innovations, which occur when a completely new knowledge domain is created, are often the aspiration of participatory projects established with an adaptive, organic orientation, particularly of the Archival Commons' vision for the global networked environment 'where archives are not singular destinations for research and inquiry, but are integrated into the daily fabric of activities'.[57] Innovation on this scale is hard to pinpoint whilst in progress, but might it be detected in the convergence of traditional cultural heritage disciplines (archives, museums, special collections and so forth) and the emergence of the digital humanities as a creative influence upon the future development of the professional role within this much broader landscape? Crowdsourcing in cultural heritage is ultimately all about making connections – in its different guises these may be connections between traces of the past or between people in the present. Perhaps its enduring legacy will be in fostering the participants' perspective of the digitised cultural heritage realm, encouraging professionals in these fields too to transcend their own view of the world – to focus then not inwardly on narrowly defined disciplinary goals, but to look outwards, embracing complexity and uncertainty, but also opportunity.

---

53 Peccatte, 'Liberating Archival Images'.
54 Justesen, 'Innoversity in Communities of Practice', 83–4, 89.
55 Ibid.
56 Yakel, 'Who Represents the Past?', 258.
57 Anderson and Allen, 'Envisioning the Archival Commons', 400.

References

Anderson, I.G. 'Necessary but Not Sufficient'. *D-Lib Magazine* 14, no. 1/2 (2008). doi:10.1045/january2008-anderson.

Anderson, S.R. and R.B Allen. 'Envisioning the Archival Commons'. *American Archivist* 72, no. 2 (2009): 383–400.

Bonney, R., H. Ballard, R. Jordan, E. McCallie, T. Phillips, J. Shirk and C.C. Wilderman. *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Information Science Education. A CAISE Inquiry Group Report*. Washington, DC: Center for Advancement of Informal Science Education (CAISE), 2009. http://caise.insci.org/uploads/docs/PPSR%20 report%20FINAL.pdf.

Burns, Tom and G.M. Stalker. *The Management of Innovation*. London: Tavistock, 1961.

Dacre, Paul. *Review of the 30 Year Rule*. London: The Stationery Office, January 2009. http://www2.nationalarchives.gov.uk/30yrr/30-year-rule-report.pdf.

Duff, W.M., B. Craig and J. Cherry. 'Finding and Using Archival Resources: A Cross-Canada Survey of Historians Studying Canadian History'. *Archivaria* 58 (2004): 51–80.

Duff, W.M. and V. Harris. 'Stories and Names: Archival Description as Narrating Records and Constructing Meanings'. *Archival Science* 2, no. 3 (2002): 263–85.

Dunning, Alastair. 'Innovative Use of Crowdsourcing Technology Presents Novel Prospects for Research to Interact with Much Larger Audiences, and Much More Effectively Than Ever Before'. LSE Impact of Social Sciences blog, August 25, 2011. http://blogs.lse.ac.uk/impactofsocialsciences/2011/08/25/ innovative-use-of-crowdsourcing/.

Duranti, L. 'Origin and Development of the Concept of Archival Description'. *Archivaria* 35 (1993): 47–54.

Evans, M.J. 'Archives of the People, by the People, for the People'. *American Archivist* 70, no. 2 (2007): 387–400.

Ferriero, David S. 'Crowdsourcing and Citizen Archivist Program'. AOTUS: Collector in Chief blog, May 19, 2011. http://blogs.archives.gov/aotus/?p=2938.

Flanagan, Mary and Peter Carini. 'How Games Can Help Us Access and Understand Archival Images'. *American Archivist* 75, no. 2 (2012): 514–37.

Flinn, Andrew. '"An Attack on Professionalism and Scholarship"? Democratising Archives and the Production of Knowledge'. *Ariadne* 62, January 30, 2010. http://www.ariadne.ac.uk/issue62/flinn/.

Greene, M.A. and D. Meissner. 'More Product, Less Process: Revamping Traditional Archival Processing'. *American Archivist* 68, no. 2 (2005): 208–63.

Hansen, Morten T. 'The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits'. *Administrative Science Quarterly* 44, no. 1 (1999): 82–111.

Haythornthwaite, Caroline. 'Crowds and Communities: Light and Heavyweight Models of Peer Production'. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*, 1–10. Los Alamitos, CA: IEEE Computer Society, 2009.

Hedstrom, M. 'Archives, Memory, and Interfaces with the Past'. *Archival Science* 2, no. 1 (2002): 21–43.

Hurley, C. 'Parallel Provenance: (1) What, If Anything, Is Archival Description?'. *Archives and Manuscripts: The Journal of the Archives Section, the Library Association of Australia* 33, no. 1 (2005): 110–45.

Huvila, I. 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management'. *Archival Science* 8, no. 1 (2008): 15–36.

Huvila, I. 'What Is a Participatory Archive?', August 24, 2010. http://istohuvila. eu/what-participatory-archive.

Justesen, Susanne. 'Innoversity in Communities of Practice'. In *Knowledge Networks: Innovation through Communities of Practice*, by P.M. Hildreth and C. Kimble, 79–95, Hershey, PA: Idea Group, 2004.

Kaplan, E. '"Many Paths to Partial Truths": Archives, Anthropology, and the Power of Representation'. *Archival Science* 2, no. 3 (2002): 209–20.

Karp, Ivan and Steven Lavine, eds. *Exhibiting Cultures: The Poetics and Politics of Museum Display*. Washington, DC: Smithsonian Institution Press, 1991.

Katz, Nancy and David Lazer. 'Building Effective Intra-Organizational Networks: The Role of Teams'. Working Paper, Kennedy School of Business, Harvard University, 2003.

Morgan, Gareth. *Images of Organization*. Updated Edn. London: SAGE, 2006.

Owens, Trevor. 'Digital Cultural Heritage and the Crowd'. *Curator: The Museum Journal* 56, no. 1 (2013): 121–30.

Palmer, Joy. 'Archives 2.0: If We Build It, Will They Come?'. *Ariadne* 60 (2009). http://www.ariadne.ac.uk/issue60/palmer.

Parry, R. 'Digital Heritage and the Rise of Theory in Museum Computing'. *Museum Management and Curatorship* 20, no. 4 (2005): 333–48.

Peccatte, Patrick. 'Liberating Archival Images: The PhotosNormandie Project on Flickr'. In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by K. Theimer, 148–57. Translated by Lynne Thomas. Chicago: Society of American Archivists, 2011.

Pennock, M. 'Twittervane: Crowdsourcing Selection'. UK Web Archive blog, December 2, 2011. http://britishlibrary.typepad.co.uk/webarchive/2011/12/ twittervane.html.

Quinn, A.J. and B.B. Bederson. 'Human Computation: A Survey and Taxonomy of a Growing Field'. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1403–12. New York: Association of Computing Machinery, 2011.

Schafer, Mirko Tobias. *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam: Amsterdam University Press, 2011.

Smith-Yoshimura, K. and C. Shein. *Social Metadata for Libraries, Archives, and Museums. Part 1: Site Reviews*. Dublin, OH: OCLC Research, 2011. http://www.oclc.org/research/publications/library/2012/2012-01r.html?urlm=161049.

Theimer, K. *Web 2.0 Tools and Strategies for Archives and Local History Collections*. London: Facet, 2010.

Theimer, K. 'Fantastic Volunteer Scanning Project with National Archives – Great Example of Crowdsourcing (and Interesting Remarks from Mr. Ferriero)'. Archivesnext blog, February 11, 2010. http://www.archivesnext.com/?p=1018.

Theimer, K. 'NARA Crowdsourcing Classification Reform'. Archivesnext blog, March 18, 2011. http://www.archivesnext.com/?p=1823.

Theimer, K. 'What Is the Meaning of Archives 2.0?'. *American Archivist* 74, no. 1 (2011): 58–68.

Theimer, K., ed. *A Different Kind of Web – New Connections between Archives and Our Users*. Chicago: Society of American Archivists, 2011.

Wasko, M.M.L. and R. Teigland. 'Public Goods or Virtual Commons? Applying Theories of Public Goods, Social Dilemmas, and Collective Action to Electronic Networks of Practice'. *Journal of Information Technology Theory and Application (JITTA)* 6, no. 1 (2004). http://aisel.aisnet.org/jitta/vol6/iss1/4.

Wenger, Etienne. *Communities of Practice: Learning, Meaning, and Identity*. Learning in Doing. Cambridge: Cambridge University Press, 1998.

Wilson, P. *Second-Hand Knowledge: An Inquiry into Cognitive Authority*. Westport, CT: Greenwood Press, 1983.

Yakel, E. 'Balancing Archival Authority with Encouraging Authentic Voices to Engage with Records'. In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 75–101. Chicago: Society of American Archivists, 2011.

Yakel, E. 'Who Represents the Past? Archives, Records, and the Social Web'. In *Controlling the Past: Documenting Society and Institutions (Essays in Honor of Helen Willa Samuels)*, edited by Terry Cook, 257–78. Chicago: Society of American Archivists, 2011.

Yeo, G. 'Debates about Description'. In *Currents of Archival Thinking*, edited by T. Eastwood and H. MacNeil, 89–114. Westport, CT: Libraries Unlimited, 2010.

Zarro, M. and R. Allen. 'User-Contributed Descriptive Metadata for Libraries and Cultural Institutions'. In *Research and Advanced Technology for Digital Libraries Lecture Notes in Computer Science 6273*, 46–54. Presented at the 14th European Conference ECDL Glasgow, UK, September. Berlin and Heidelberg: Springer, 2010.