5

Frames and the future of archival processing

Jenny Bunn

Introduction

In 1989, David Bearman published an essay entitled 'Archival Methods' in which he set out a comparison between the magnitude of the tasks archivists had set themselves and the magnitude of their capabilities, revealing 'substantial discrepancies' (Bearman, 1989). Revisiting this work in 1995, he noted that: 'No one has directly disputed the claim that our methods are out of sync with the problems by more than an order of magnitude' (Bearman, 1995, 381–2). Indeed, in 2005, a very similar claim was made by Mark Greene and Dennis Meissner who stated that: 'Processing backlogs continue to be a problem for archivists, and yet the problem is exacerbated by many of the traditional approaches to processing collections that archivists continue to practice' (Greene and Meissner, 2005, 208). They too, as Bearman had done, called for archivists to think again about their methods and made suggestions for new approaches.

The suggestion, in 1989, 1995 and 2005, was that traditional archival methods were inadequate; they could not hope to succeed given the scale of the task at hand and needed to change. It is now over 12 years later and yet it would be very easy for this chapter to become just another instance of the same: a stark reminder of the ever-increasing magnitude and difficulty of the archival task, a call to rethink and some suggestions for how to proceed. In striving to avoid this easy path, this chapter will instead seek to cover the ground from a slightly

different angle by exploring some of the solutions that have been proposed by others and using them as a starting point for a discussion about the future of archival processing.

In the above, the terms archival methods and archival processing have been used interchangeably and there is a broad sense in which both can be taken to mean all the things that archivists do: their activities, their processes, their methods. Archivists undertake many different processes, however, and the Society of American Archivists' glossary narrows its definition of processing down to the 'arrangement, description, and housing of archival materials for storage and use by patrons', although it also notes that 'Some archives include accessioning as part of processing' (Society of American Archivists, n.d.). The solutions which are being used as our starting point would seem to imply a mix of both senses, broad and narrow.

For example, one solution implying the more general sense is that offered by Greene and Meissner. They suggest that it would be 'a sign of professional maturity' if archivists were 'to own up to the limitations we work under and accept that the golden minimum recommended here . . . is all we can realistically accomplish' (Greene and Meissner, 2005, 255). Sensible advice to be sure, which could apply equally well to all forms of work, but in these times of increasingly constrained resources, this has become less a solution than a reality for the many archivists who are only too aware of the need to 'make each new situation argue for any additional investment of time and effort' (Greene and Meissner, 2005, 248).

On the other hand, a solution which seems to suggest processing in a narrower sense, of carrying out description, is that suggested by David Bearman in 1995 that archivists should 'Let Users Describe Records' (1995, 402). Much has been written about such user-generated description in recent years, particularly as new technologies have allowed it to morph into the field of crowdsourcing, and this is another solution that has become a reality, at least to some extent.

In this chapter we will be starting from the point of proposed solutions to the challenges faced by archival processing (in both the broad and narrow sense), but this is not where we will be ending up. We will not be providing detailed suggestions for how we can alter the processes of archivists to cope with the vast scale of the task they face. Instead, we will be considering solutions already proposed in

order to unpick the framings and meanings of archival processing implied by those solutions: the ways in which we characterise what archival processing involves. We shall do so in the hope that this will lead to a better understanding of the fact that we can frame archival processing and the problems it faces in many different ways. Perhaps understanding this will provide another type of solution, one which allows us to imagine different futures and framings and to avoid treading the same paths again and again?

Un-framing archival processing: moving information across systems

The solution we will use as our starting point is the suggestion made by David Bearman that 'archivists will need to identify that information which can be obtained from outside, and import it into their systems automatically' (1989). Archivists have always obtained and imported information from outside into their systems, but that process was not, until very recently, an automatic one. Previously, if an archivist wished to import, say, the information written on a file cover into one of their systems, such as the catalogue which detailed their holding of said file, it was generally a process of manual re-typing. Now the information on the file cover comes instead in the digital form of a standard binary code, such that a quick press of a button and a machine will have imported into archival systems a list of possibly hundreds of thousands of files, along with lots of other information about them, such as size and dates created and modified. Tools, such as DROID, have already been developed and are increasingly being adopted by archivists to carry out this task, with the added bonus of the automatic generation and importing of new information, such as a checksum to fix each file's integrity.

What sort of framing and meaning of archival processing is implied by this solution? Firstly, as the above discussion makes clear, one of the framings it implies is a distinction between manual and automatic processing and an assumption that the automation of processes will make those processes more efficient. This assumption may well be correct, but that is an argument that can be had in relation to all types of processes, not just archival ones. Secondly then, and in a way that speaks more to the definition of the sort of processes that are involved

in archival processing, it frames the activity it describes as importing information from external systems into internal systems.

Moving information across different systems is not always as seamless and as easy a process as we think it should be. As Duff and Harris remind us: 'Early twenty-first-century technological realities make it impossible to build a complex collective project without standards. For example, every e-mail message relies on over 200 internet standards for its successful transmission' (Duff and Harris, 2002, 283). This complexity, the legs paddling furiously beneath the swan gliding across the water, can most easily be seen through a focus on interoperability. A detailed look at this issue, from a recordkeeping metadata perspective, has been undertaken by Joanne Evans (2007). One of the lenses that Evans employs in her consideration of interoperability is provided by the Layers of Interoperability Model. This model sees an abstract layer supported by a representation layer supported by a transport and exchange layer.

As Evans writes, 'underpinning this abstract layer is a conceptual data model' (2007, 50). The need for such underpinnings highlights the way in which what is ultimately being transported and exchanged is not just data, or in this case metadata, but the meaning and understanding of that (meta)data. Such meaning and understanding depends on also understanding 'the perspective or context of metadata' and it is this that the conceptual data model sets out (Evans, 2007, 50).

With the development of the new Records in Contexts standard, archivists are attempting to create such a conceptual data model; a setting out of the important entities in the archival world view (e.g. records, agents, functions, dates) and the ways in which they are or can be related. The intention is that this model will also be encoded in the form of an ontology, such that it will be understandable to computers as well as human beings (International Council on Archives Experts Group on Archival Description, 2016). This last point raises again the distinction between manual and automatic processing, albeit in the form of human versus computer processing.

The framing of moving information across different systems also raises the question of the framing of those systems. Do we see just them in terms of different software packages, or do we take a wider view? What if you see them less as pieces of software and more as 'interdependent components organized to achieve an end . . . organized

collections of hardware, software, supplies, people, policies and procedures, and all the maintenance and training that are required to keep these components working together' (Bearman, 1993, 17). In this light, we might conceive of a system as an entire society, era or epoch, and the information we are moving across those systems as an understanding of those past societies, eras and epochs, such that they inter-operate with our own. Our view of archival processing would again be a wider one, seeing it more in terms of everything that archivists do, their very purpose and point, not just (meta)data automatically imported or not.

Perhaps then, another aspect of archival processing that this un-framing suggests is that the seemingly easy slippage between its broad and narrow senses is not just a question of whether accessioning is or is not a part of archival processing, but rather an indication of a deep connection between that processing and the purpose it serves?

Un-framing archival processing: documenting records-creating

So far, I have suggested that the frame for archival processing implied in the suggested solution that 'archivists will need to identify that information which can be obtained from outside, and import it into their systems automatically' is that archival processing can be conceived of in terms of moving information across different systems (Bearman, 1989).

I have surfaced the way in which such a conception can work in terms of both wider and narrower senses of archival processing, as everything that archivists do or just the management of metadata, and I have highlighted the distinction between manual and automatic processing that is implicit within this frame. I shall now turn to another suggested solution, that made by David Bearman, that 'archives should not describe records, but, rather, document records-creating activity' (1995, 401).

Within this solution there is clearly another distinction being used as a frame, this time it is not that between manual and automatic processing but rather that between description and documentation. This distinction is also discussed by Bearman in the article, 'Documenting Documentation' (1992). Here he explains it in terms of

focus, stating that whereas 'Description is focused on records . . . Documentation is focused on activity in the records-generating institution' (Bearman, 1992, 34). Later in the article he expands on his definition of this last activity as 'the activity that generated the records, the organizations and individuals who used the records, and the purposes to which the records were put' (1992, 45).

Whereas Bearman's phrasing seems to suggest that the focus should not be on the records at all, current recordkeeping standards suggest that the focus should be on both the records and on certain aspects of records-creating activity. For example, ISO 23081-2:2009 suggests that entities of particular importance are:

> the records themselves, whether an individual document or aggregations of records . . .
> the people or organizing structures in the business environment . . .
> the business transacted . . .
> the rules governing the transaction and documentation of business . . .
>
> (ISO 23081-2:2009, 7)

The recent consultation draft of the Records in Contexts conceptual model brings into focus even more entities, providing a list of: record, record component, record set, agent, occupation, position, function, function (abstract), activity, mandate, documentary form, date, place and concept/thing (International Council on Archives Experts Group on Archival Description, 2016).

Viewed in these terms, what becomes most clear is the scale and magnitude of the archival processing task. For in these terms, archival processing becomes framed as the process of keeping track of and recording details about all these kinds of entities and how they are related. Worse still, if archival processing in these terms is set against the frame of needing to do this in a way that makes the resulting information available for automatic processing by computers (across different systems), that scale and magnitude leaps to a whole new level. In the past, human processing by archivists led to the creation of a reasonably efficient summary of this sort of information in the form of narrative description or text which could easily be processed by other humans (assuming that they spoke the same language). Even so Greene and Meissner were concerned that:

> An unfortunate tendency on the part of processing archivists is to use the preparation of these text notes as an excuse to demonstrate their own knowledge (of both collection and historical context) and writing ability. Perhaps this is an attempt to demonstrate professionalism but, if so, it is a misguided one that further reduces processing productivity. The goal should always be to convey such narrative content and contextual information as briefly as possible and with as little recourse to outside sources as possible.   (Greene and Meissner, 2005, 247)

And yet, despite Greene and Meissner's misgivings, I would suggest that this process of summary (dealing with all the different entities as part of one narrative) text production (and occasionally translation into other human languages) is still less labour intensive than what seems to now be required: creating separate descriptions for all the different sorts of entities involved, linking them together, encoding them all in machine readable form, developing the computer systems to store, process and display the resulting data and agreeing on the standards, ontologies and conceptual data models by which it will all be interoperable and portable as one short-lived digital technology is replaced by another short-lived digital technology.

Un-framing archival processing: information processing

In the previous section, I suggested that the frame for archival processing implied in the suggested solution that 'archives should not describe records, but, rather, document records-creating activity' (Bearman, 1995, 401) is that archival processing can be conceived of in terms of keeping track of and recording information on a large number of different sorts of entities and the relationships between them. I have surfaced how this task is complicated when it can no longer be completed solely by summary text production but needs instead to be undertaken in a way that also makes the resulting information available for automatic processing by computers. It is not my intention, however, to give the impression that computers are therefore the problem. Rather, the previous discussion is intended to lead into one more un-framing of archival processing, this time one which sees it in terms of information processing.

Information processing is defined by the Oxford English Dictionary

(2018) as 'the processing (by a machine or by an organism) so as to yield new or more useful information'. As the definition makes clear, it is something that is done by both living and non-living things, by both humans and computers. However, the ways in which they undertake that processing is different. To indicate this difference, I will refer to information processing in respect of computer information processing, but sense-making in respect of human information processing. This is not to suggest that sense-making and information processing are necessarily the same thing, but that I believe human information processing to be qualitatively different to that undertaken by computers.

In Bearman's initial framing of the problem which started this article, certain 'costly and inefficient steps' in archival methods were identified by name, these being those 'of analysing the records themselves in order to find clues to context, and then entering this data into our own descriptive systems by keyboard' (Bearman, 1989). Any archivist who has processed material, looking through the detritus of other people's lives in order to find and read those clues and make sense of what they are seeing for themselves and others, will tell you that it is time-consuming and difficult. But does that necessarily make it costly and inefficient? Depending on how much the archivist is being paid it could be costly, but is it inefficient? Is this specific kind of information processing, of sense-making, something that a computer, an automatic rather than a manual processor, could do better, faster or more efficiently?

Sense-making

In highlighting this process, I do not assert that it is one solely undertaken by archivists and records managers. Historians and researchers, groups from which archivists and records managers sometimes individually, and increasingly as a professional collective, feel the need to make a definitional difference, also undertake such readings of archival material (albeit mostly readings of material that has already been processed by archivists to some degree). For example, Liz Stanley, coming from a research perspective, talks in terms of archigraphics, and the development of 'an archival sensibility', as well as processes of questioning, mapping, documentary analysis and interpretation (Stanley, 2017). Then again, Jennifer Meehan, coming

from the archival perspective, talks in terms of 'an analytical process' involving 'acts of interpretation and representation' with reference to the idea of inference (Meehan, 2009, 72–3). This is the process to which I am referring and the terms in which I see archival processing as a form of sense-making.

Returning then to the question of whether this process can be carried out more efficiently by a computer than a human being, one variable that is hugely important is the nature of the material on which the process is being carried out. For, if it is not digital in nature, the computer will not be able to process it at all and the human being is the only option. Increasingly though, the nature of the material we are working with is changing as it is born-digital. With this change, the sort of clues that archivists had become expert at reading, such as changes in format or patterns of organization, may no longer exist in the same way and may no longer be discernible without help. Is one reason why we tend not to surface this process of sense-making within archival processing very often because we know we have lost it? If so, do we want to prevaricate until this loss leaves us in danger of becoming useless or do we want to relearn it? How can we combine the computer's information processing, admittedly faster and more accurate as it is over large quantities of digital material, with our own sense-making?

This question is very much an open one, but there have been projects that have considered it. For example, innovative research carried out by Kathryn Chandler as part of a Masters of Archival Studies at the University of British Columbia, has explored the potential of using new tools for exactly the sort of archival sense-making that has been discussed above (Chandler, 2016). These tools allowed for the extraction of certain pieces of data from a data set and then the running of this data through several community detection algorithms to produce visualisations of putative communities, or social networks, within a single department at the University. These visualisations were not the end-point of the research, but were rather the starting point for a process of sense matching in which the senses of the (net)working within the department produced by the algorithms in the form of visualisations were matched (or not) against the senses of the same as they were experienced by those working within that department (Chandler, 2017, 271).

Members of staff within the department were asked which of the visualisations 'best reflected their pattern of record modification' and the manager was asked which 'best represented the department' (Chandler, 2017, 271). In this way, Chandler was able to explore a number of things including: (a) in the new digital environment, the effectiveness of clues, such as the names of agents and the dates of actions, traditionally relied on by archivists for building up a sense of things in the old paper environment; and (b) the effectiveness of different kinds of computer assisted processing (for example the use of different algorithms) for automating the generation of a sense of things that could be taken as authentic, at least in the sense of matching the way those things had been experienced by those who were a part of them (Chandler, 2016).

Research of this kind is also starting to take place in a practice rather than an academic context. For example, The National Archives have recently experimented with the tools and techniques brought together and marketed as e-discovery software and have started to learn how and to what degree they can (and cannot) assist in, amongst other things, 'understanding born digital collections at a high level' and 'extracting meaning' (The National Archives, 2016, 14, 17). Experiments of this kind allow archivists to relearn how to work with the material remnants of an organization and the available processing power (now encompassing the use of algorithms, text mining and so on) to not just generate any sense, but a specifically archival sense. What this specifically archival sense entails is not very well defined, but it can perhaps be tentatively seen as being one which acts as an authentic or at least reasonably accurate representation of the activity from which it arose; a sense which would make at least some sense to those, like the staff members in Chandler's study, involved in that original activity.

Conclusion

This chapter started by recalling the argument made on several occasions that archival 'methods are out of sync with the problems by more than an order of magnitude' (Bearman, 1995, 382). It took the solutions that had been suggested to the problems of processing as its starting point, not to provide more solutions, but rather to consider

the framings and meanings of archival processing implied by those solutions.

In this way, consideration of David Bearman's suggestion that 'archivists will need to identify that information which can be obtained from outside, and import it into their systems automatically' led to a conception of archival processing in terms of the process of moving information across different systems (Bearman, 1989). Then again, consideration of the suggestion that 'archives should not describe records, but, rather, document records-creating activity' led to a conception of archival processing in terms of the process of keeping track of and recording information about a large number of different sorts of entities and the relationships between them (Bearman, 1995, 401). In both cases, another framing was highlighted, which distinguished between manual and automatic processing and between the need to communicate information in a way that made that information available for processing by both human and computer agents. This ultimately led to one more framing of archival processing, this time in terms of the process of information processing.

One of the purposes of this chapter was to highlight the way in which, by defining archival processing solutions in certain terms, we also frame both archival processing and the problems its faces. If we are to realise different futures for archival processing, we will need to be alert to such framings and be able to move beyond them. For example, being alert to the framing of moving information across different systems can lead us to question the boundaries we choose to place around those systems and whether we are taking a broader or a narrower view of them. Then again, a framing of keeping track of and recording information about many different sorts of entities and the relationships between them, can lead us to a renewed recognition that the magnitude of the task we set ourselves is indeed truly immense.

The final re-framing, in terms of information processing, perhaps allows for the possibility of a solution, this time in terms of an answer to the question: how can we combine the computer's information processing, admittedly faster and more accurate as it is over large quantities of digital material, with our own sense-making? David Bearman framed the processes 'of analysing the records themselves in order to find clues to context, and then entering this data into our own descriptive systems by keyboard' as 'costly and inefficient steps'

(Bearman, 1989). It may well be that, with tools like DROID, we are rapidly removing the need for the last of those steps. However, the first step is, to my mind at least, not one we should eradicate, but a skill we should celebrate and develop. Relearning this skill in respect of the new material we deal with, and the new tools and vastly increased processing power available to us, is the future of archival processing.

References

Bearman, D. (1989) *Archival Methods: Archives and Museum Informatics Technical Report #9*, Archives and Museums Informatics, www.archimuse.com/publishing/archival_methods.

Bearman, D. (1992) Documenting Documentation, *Archivaria*, **34**, 33–49.

Bearman, D. (1993) Record-Keeping Systems, *Archivaria*, **36**, 16–36.

Bearman, D. (1995) Archival Strategies, *American Archivist*, **58** (4), 380–413.

Chandler, K. (2016) *Exploring the Principle of Provenance with Social Network Analysis*, Masters of Archival Studies Thesis, University of British Columbia.

Chandler, K. (2017) Investigating original order with Cybernetics and Community Detection Algorithms, *Archival Science*, **17** (3), 267–83.

Duff, W. and Harris, V. (2002) Stories and Names: archival description as narrating records and constructing meanings, *Archival Science*, **2** (3), 263–85.

Evans, J. (2007) *Building Capacities for Sustainable Recordkeeping Metadata Interoperability*, PhD Thesis, Monash University.

Greene, M. and Meissner, D. (2005) More Product, Less Process: revamping traditional archival processing, *American Archivist*, **68** (2), 208–63.

International Council on Archives Experts Group on Archival Description (2016) *Records in Contexts: a conceptual model for archival description*, Consultation Draft v0.1, www.ica.org/sites/default/files/RiC-CM-0.1.pdf.

ISO 23081-2:2009 *Information and documentation – Records Management Processes – Metadata for Records – Part 2: Conceptual and implementation issues*, International Standards Organization.

Meehan, J. (2009) Making the Leap from Parts to Whole: evidence and inference in archival arrangement and description, *American Archivist*, **72** (1), 72–90.

Oxford English Dictionary (2018) Oxford University Press, http://www.oed.com/.

Society of American Archivists (n.d.) *A Glossary of Archival and Records Terminology*, www2.archivists.org/glossary.

Stanley, L. (2017) Archival Methodology Inside the Black Box: noise in the archive! In Moore, N., Salter, A., Stanley, L. and Tamboukou, M., *The Archive Project: archival research in the social sciences,* Routledge.

The National Archives (2016) *The Application of Technology-Assisted Review to Born-Digital Records Transfer, Inquiries and Beyond,* research report, www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf.