Bringing Things Together: Aggregate Records in a Digital Age

GEOFFREY YEO

RÉSUMÉ Les recoupements sont nombreux dans l'univers numérique, tout comme ils le sont dans l'analogique. Dans le monde des documents sur papier, les recoupements sont rassemblés et le classement est finalisé avant que l'usager ne soit impliqué, mais plusieurs critiques prétendent que la révolution numérique a renversé ces conventions; les usagers peuvent maintenant établir l'ordre de leur choix et construire leurs propres recoupements, à leur guise et lorsqu'ils le veulent. Pour les archivistes qui croient que des hiérarchies stables de dossiers et de séries sont essentielles, ces idées peuvent paraître troublantes. Cependant, plutôt que de se sentir menacés par la fluidité de l'ordre numérique, nous devrions nous servir de ces nouvelles capacités pour nous aider à dépasser certaines des limites des méthodes du monde du papier. Par exemple, si les cadres technologiques appropriés sont en place, nous serons capables de construire de multiples séries de documents en parallèle afin de rencontrer divers besoins ou réaliser différentes conceptualisations des séries. Nous n'avons pas à privilégier le rassemblement de documents pour des raisons de preuve plus que d'autres rassemblements qui favorisent différents cadres de référence, mais nous devons nous assurer que nous puissions rassembler des recoupements de documents et présenter des aperçus du contexte lorsque nécessaire. La mise en œuvre pratique sera possible seulement quand les outils appropriés seront disponibles. Nous aurons besoin de systèmes extensibles et conviviaux qui permettront la construction de recoupements de documents, ainsi que de recoupements de « non-documents » (« *'non-record' aggregations* »), et qui conserveront l'information au sujet du contexte logique et du classement physique imposé dans le passé. Si de tels systèmes peuvent être créés, il sera possible de présenter les ressources archivistiques de plusieurs autres façons, reflétant leurs divers ordres « originaux », les différentes interprétations de leur contexte, ainsi que les autres ordres que désireront les usagers au cours de leur recherche et de leur expérimentation.

ABSTRACT Aggregations abound in the digital realm, as they do in the analog. In the world of paper records, aggregations are brought together and arrangement is fixed before the user arrives on the scene, but many critics argue that the digital revolution overturns these conventions; users can now make orders of their own choosing and construct their own aggregations dynamically when they wish. To archivists who believe that fixed hierarchies of files and series are essential, these ideas may appear disquieting. However, far from feeling threatened by the fluidity of the digital order,

ARCHIVARIA 74 (Fall 2012): 43–91

we should be able to take advantage of the new capabilities to help overcome some of the limitations of paper-world methods. For example, if appropriate technological frameworks are in place, we will be able to build multiple overlapping record series to meet different needs or realize different conceptualizations of series boundaries. We need not privilege the assembly of records for evidential purposes over other groupings that accommodate different frames of reference, but we must be sure that we can bring together aggregate records and present views of context when they are required. Practical implementation will be possible only when appropriate tools are available, and we will need scalable and user-friendly systems that enable the construction of aggregate records as well as "non-record" aggregations, and also preserve information about logical contexts and about physical arrangements imposed in the past. If such systems can be developed, it will be possible to present archival resources in many different ways, reflecting their various "original" orders, different interpretations of context, and other orders newly desired by users in the course of research and experimentation.

Introduction

In his presidential address to the Society of American Archivists in 2008, Mark Greene spoke of archivists' belief "that aggregation is ... an essential reflection of the organic nature of recordkeeping." According to Greene, a "focus on the aggregate is part of what sets us apart from librarians and museum curators. ... Our ability to work in the aggregate is ... an important source of our power."[1] This paper sets out to explore some aspects of our "ability to work in the aggregate" in a digital age.[2] It reviews ideas about series and files as conceptual and physical units, considers how far it is appropriate for us to "focus on the aggregate" in the world of digital records, and examines suggestions that the digital revolution has changed the rules about fixity of aggregation. Many critics affirm that users can now make orders of their own choosing and construct their own aggregations when they wish. What impact might this have on the perceived role of the aggregate in recordkeeping? What might be the implications for our understandings of "original" order and contextualization? In attempting to tease out some answers to these questions, this paper builds on earlier work in which I developed a view of records as representations and discussed the intersections between conceptual fonds and physical collections. After investigating conventional definitions of the series and the file in professional literature, the paper indicates some new approaches to

1 These quotations are from the published version of Greene's address: Mark A. Greene, "The Power of Archives: Archivists' Values and Value in the Postmodern Age," *American Archivist* 72, no. 1 (2009): 23–24.

2 Parts of this paper were presented at the 36th Annual Conference of the Association of Canadian Archivists, Toronto, June 2011, at a meeting of the Norsk Arkivråd, Oslo, May 2012, and at the 17th Congresso Brasileiro de Arquivologia, Rio de Janeiro, June 2012.

archival practice in a world no longer constrained by traditional assumptions about stable aggregations and hierarchical systems.

A Representational View of Records

My starting point is a question much debated among archivists: what might we mean when we speak of a "record"? In papers published in 2007–8, I sought to characterize records as *representations of occurrents*. Occurrents, I suggested, are temporal phenomena such as activities, functions, processes, transactions, or events, and records serve to represent these in a persistent manner; the representation remains available after the conclusion of the occurrent that is represented. Records are typically created at the level of a single activity or a single step within an activity, and these "elementary" records can be aggregated to form records of functions, processes, or other occurrents at higher levels.[3]

Of course, it can be argued that characterizing records as persistent representations merely shifts the debate from one field of contestation to another. If we perceive a record as a representation of something else, we may still want to ask how far, or in what way, one thing can be said to represent another. Ideas about representation help us to make sense of many aspects of human behaviour and experience, but they remain controversial, and some critics have seen them as targets for censure or scorn.[4] Persistence, too, can be a problematic concept, not merely because traces or objects that persist over time may be open to continuous reinterpretation, but also because notions of persistence confront us with troublesome questions about what it might mean to claim that objects can remain identical from one moment to another. In practice, however, most of us accept that humans – and machines designed by humans – can construct linguistic, symbolic, or iconic representations, and that such representations can endure and be carried forward in time. Records, like other representations, may be interpreted in many ways, but their persistence influences and shapes the interpretations to which they are subject.

I also suggested that it can be helpful to see elementary records as representations *created by participants or observers or their authorized proxies*. I did not seek to impose a definition of who or what might qualify as a "participant" or an "observer" of an activity or event; a degree of uncertainty is inevitable, and there seems little merit in attempting to pin down these terms too

3 Geoffrey Yeo, "Concepts of Record (1): Evidence, Information, and Persistent Representations," *American Archivist* 70, no. 2 (2007): 315–43; "Concepts of Record (2): Prototypes and Boundary Objects," *American Archivist* 71, no. 1 (2008): 118–43.

4 See, for example, Claire Colebrook, *Ethics and Representation: From Kant to Post-structuralism* (Edinburgh, 1999); Richard Rorty, *Objectivity, Relativism and Truth* (Cambridge, 1991).

precisely.[5] Nevertheless, participation in, or observation of, an activity remains central to our notions of the creation of elementary records. Participants and observers may not be neutral or impartial witnesses to the activity, but they have knowledge unavailable to those who did not experience it.

Assembly of aggregate records, however, is not always the responsibility of those who participate in or observe the occurrents that the records represent. In an organizational context, such records characteristically come together as a result of decisions and actions taken by records management staff, computing specialists, administrators, or secretarial assistants, as well as creators of elementary records. All these people work for the organization and presumably act on its behalf, but many of them may have little immediate connection with the occurrents represented in the records they assemble. Aggregate records can also be put together by others at a greater distance. For example, when Professor Plum organizes a course of seminars and after each seminar sends notes of the proceedings to the local archives, it is the archivist who aggregates these to construct a record of the course as a whole; when John's friend intermittently sends him photographs of her children's activities, it is John who puts them together to form an aggregate record of what her children have done.[6] Although aggregations assembled by third parties have often been considered suspect in the world of archives, it seems possible to conclude that aggregate records might be brought together by anyone with the interest and the means to do so.

Physical and Conceptual Groupings

Questions about archival aggregations inevitably lead us to examine ideas about fonds and collections. In "The Conceptual Fonds and the Physical Collection" (*Archivaria* 73), I argued that a fonds is purely a conceptualization and that physical aggregations of records are collections.[7] Both the fonds and the collection have the character of bringing, or keeping, things together,

5 Yeo, "Concepts of Record (1)," 338. Further scope for ambiguity arises from the somewhat infelicitous phrase "authorized proxies," which I inserted because our society recognizes some representations (such as birth registers) as records even though their creators do not directly participate in or observe the events they represent. Who or what authorizes a "proxy"? How far can distinctions be drawn between representations created by "proxies" and those created by third parties? These questions admit no easy answers. There are many third-party representations (such as writings by historians long after an event) that we would not normally label as records of the events they describe, but ultimately the boundaries of the concept of "record" remain fuzzy.

6 Yeo, "Concepts of Record (2)," 134–38.

7 Geoffrey Yeo, "The Conceptual Fonds and the Physical Collection," *Archivaria* 73 (2012): 43–80. I argued that neither the fonds nor the collection "occupies the exact niche ... traditionally assigned to it" (p. 44).

but the fonds, as Terry Cook affirmed twenty years ago,[8] is an intellectual construct. Collections are what we manage, preserve, deliver, and examine in the "real" world. They are ubiquitous in digital as well as analog environments.

Unlike "fonds," the term "collection" is not restricted to our own field; writings about collections span many disciplines. Some cultural commentators have claimed that collections and their contents are divorced from everyday utilitarian roles and translated to a qualitatively different plane.[9] Such claims might seem to associate record collections with "historical" archives rather than organizational records management,[10] but I would suggest that rigid distinctions between utilitarian business roles and "non-utilitarian" cultural purposes are unsustainable and that any ends for which a collection is employed are ultimately utilitarian. Collections can be assembled by creators of elementary records and records managers, as well as by archivists and private collectors. I propose a wider application for the term "collection" than European and Canadian archivists normally allow.

Undoubtedly, archivists have usually wanted to set themselves apart from private collectors and to differentiate their holdings from so-called "artificial" collections. Because understandings of records and archives are closely associated with activities and other occurrents in organizational work or personal lives, archivists have traditionally attributed a special significance to aggregations of records brought together at or near the time of these occurrents. Many writers about archives belittle human participation in the assembling of record aggregations and claim that these aggregations emerge naturally or organically as life or work takes its course. Their position is reminiscent of conservative forms of self-organization theory in the social sciences, whose proponents assert that systems and hierarchies can be self-organizing, emphasize the supposed spontaneity of organizing processes, and deny a role for human agency in the emergence of order.[11]

8 Terry Cook, "The Concept of the Archival Fonds: Theory, Description, and Provenance in the Post-Custodial Era," in *The Archival Fonds: From Theory to Practice*, ed. T. Eastwood (Ottawa, 1992), 73.

9 Russell W. Belk, "Collectors and Collecting," *Advances in Consumer Research* 15 (1988): 548–53; Susan M. Pearce, *Interpreting Objects and Collections* (London, 1994), 2–3.

10 See Krzysztof Pomian, *Collectors and Curiosities* (Cambridge, 1990), 9. For Pomian, "the category of collection" includes "most libraries and archives," but not records that "remain part of ... economic circuits of activity."

11 I know of no evidence that conventional ideas about archives have been informed by self-organization theory, but the parallels are instructive. Self-organizing systems are said to be more than the sum of their parts (Christian Fuchs, *Internet and Society: Social Theory and the Information Age* (New York, 2008), 13, 34). As in other strands of systems thinking, the notion of "emergence" is pre-eminent, and hierarchies are said to be produced as new order or new "qualities" emerge on a higher level; these qualities in turn enable and constrain the operation of the system at lower levels (ibid., 11–14). Ideas about self-organization can be

However, I argued that – even if we accept that elementary records can sometimes be *created* more or less naturally in the course of business or daily life – distinctions between "organic" and "artificial" are hard to maintain when we examine the *aggregation* and *retention* of such records.[12] Despite claims by some commentators that low-cost digital storage will eliminate the need for selectivity and enable us to retain everything, we cannot (yet) live in an ideal world where no record is ever destroyed. Choices about which records to keep are "artificial" human decisions; the collections that result from them are also to some degree artificial. Even "original" aggregations assembled by elementary records creators are purposeful and thus bear a sense of artificiality as well as a close connection to the contexts of production. In an organizational environment, a collection of records would appear to be a necessary component of what David Bearman called a "recordkeeping system," although if such systems include equipment, people, policies, and procedures, as Bearman suggested,[13] it is clearly not their only component. We are also likely to encounter many collections, within and outside organizational workplaces, that do not form part of a recordkeeping system defined in this way.

A fonds, on the other hand, is not determined by the decisions or accidents that lead to the formation of physical collections. We can perceive it as the totality of interdependent elementary records produced and received in the course of the life of a person or a group of persons or the functions and activities of an organization. Conceptually, it may include records considered ephemeral as well as those formally designated for retention. Its membership can be described but need not be physically brought together. Nevertheless, while description of its membership may meet the needs of some who seek knowledge of a fonds, others are likely to insist that no fonds is fully available unless it can be realized as a collection, a physical aggregate record that we can access and use.

traced back to G.W.F. Hegel and Friedrich Engels in the nineteenth century, but are now commonly associated with the biologists Humberto Maturana and Francisco Varela and with social scientists such as Niklas Luhmann. The pertinence of their thinking to archival description has been explored by Jennifer J. Bunn, "Multiple Narratives, Multiple Views: Observing Archival Description" (PhD diss., University College London, 2011); see also Angelika Menne-Haritz, *Business Processes* (Dordrecht, 2004), 19–24. In social science, notions of emergence in self-organization theory do not necessarily exclude human participation, but Luhmann saw humans largely as observers rather than active participants, and the economist F.A. von Hayek argued that self-organizing systems are spontaneous and that human interference in their operation is harmful (Fuchs, 26–27, 40). Assertions about the spontaneous production of archives (e.g., in Elio Lodolini, *Archivistica: Principi e Problemi*, 2nd ed. (Milan, 1985), 14, 127) seem remarkably similar.

12 Yeo, "The Conceptual Fonds," 51–52, 58–59; "Concepts of Record (2)," 126–29, 134–35.

13 David Bearman, *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations* (Pittsburgh, 1994), 36.

Viewed in this way, fonds and collection are not unrelated, as traditional archival science has often supposed. In some sense, the fonds always bears the mark of a possible realization. In practice, however, fonds are rarely realized as co-located physical wholes; the collections we encounter seldom correspond exactly to conceptual fonds. Records that represent aspects of organizational work or individual life are often divided across multiple sites (a practice that seems likely to become even more frequent for digital records in an era of cloud computing). Over time, some elementary records may be inadvertently lost, others consciously eliminated in appraisal exercises. Collections often comprise the segments of a fonds that have been allocated to a particular repository or the residues that have survived the processes of selection or the accidents of neglect. Items may be added to or removed from a collection, and the components of collections may sometimes include items of differing provenance.

Like any conceptualization, a fonds can be difficult to circumscribe. Because membership of fonds may be non-exclusive, one fonds may not be wholly discrete from another and there is potential for a multiplicity of overlapping fonds. Recognition that fonds are conceptual also opens the possibility of varying interpretation of their boundaries. Do Professor Plum's publications fall within the perimeter of his fonds? Or the souvenir objects he acquired during his trip to India, or his press cuttings of the Taj Mahal? Borderline cases such as these will almost always be disputed.[14] Identification of collections, however, is less subjective. As material entities, they may change over time, but at any given moment their boundaries are usually easy to recognize. Determination of the extent and shape of collections is the prerogative of the collector, although it may be subject to constraints within which the collector operates.

The circumstances of the digital age both enable and require us to protect the provenance and contexts of archives by means that go beyond mere attempts at preserving or documenting the collections assembled by records creators and custodians. Nevertheless, unlike Cook and Bearman, I do not believe that archivists should "ignore the physical and concentrate on the conceptual."[15] The formation of physical collections is one of our responses

14 The "prototype" effects that can be observed in connection with perceived boundaries of the concept of record (Yeo, "Concepts of Record (2)," 121) are also observable here. Where fonds membership is concerned, the status of Plum's correspondence file is less likely to be disputed than the status of his box of press cuttings, and the press cuttings in turn may be less disputed than his souvenir woodcarving. Prototype theory (Eleanor Rosch and Barbara B. Lloyd, eds., *Cognition and Categorization* (Hillsdale, NJ, 1978); D.A. Cruse, "Prototype Theory and Lexical Semantics," in *Meanings and Prototypes: Studies in Linguistic Categorization*, ed. S.L. Tsohatzidis (London, 1990), 382–402) suggests that the core of a conceptual category is usually uncontested, while the inclusion of items further from our mental prototype may be hotly debated.

15 Terry Cook, "The Impact of David Bearman on Modern Archival Thinking," *Archives and*

to the impossibility of comprehending everything and is a necessary part of human experience. Meanings can be attached to collections (including those we cannot hope to label as aggregate records) as well as to conceptual fonds, and documentation of past and present collections is no less important than documentation of fonds.

The File and the Series

Collections, observably, have part-whole structures; they are assemblages of items, and each item in the collection is acknowledged as a member of the whole. A collection may transcend its parts, but the relationship of the brick to the wall, or the soldier to the army, is always that of a part to a whole. The structural aspect of collections also makes it possible to identify sub-collections. An army, it must be admitted, is not usually perceived as a collection, but it provides a useful exemplar. It can be divided into battalions and regiments; soldiers form part of a battalion, which in turn forms part of a regiment. Battalions, regiments, and other "sub-collections" of an army can be reorganized over time, but at any one moment each soldier belongs to only one line of the hierarchy. A further characteristic of sub-collections is that they can be seen both as divisions of the larger whole and as collections in their own right. Within the army, the regiment has its own identity; returning the discussion to a place where the word "collection" may seem more at home, we can likewise perceive the "ABC Museum Watercolours Collection" both as a collection in itself and as a division of the larger collection held by the museum. Groupings that are sub-collections in one context are collections in another.

In the pre-digital world, archival aggregations appear to work in just this way. Archival "items" are commonly believed to have membership of aggregations at several higher levels, including levels known as "files" and "series"; at any one time, it seems that an item can be housed in only one file and that a file is part of only one series. Files have their own identities but are also seen as parts of a larger whole. Some archivists have sought to confine the term "record" to the level of the "item" or the "archival document,"[16] but in our book *Managing Records*, published in 2003, Elizabeth Shepherd and I suggested that records can be identified with aggregations at numerous different levels.[17]

Shepherd and I took an approach broadly derived from systems theory and used the notions of "activity" and "process" to distinguish records at various

Museum Informatics 11, no. 1 (1997): 20.

16 See, for example, Luciana Duranti and Randy Preston, eds., *International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2): Experimental, Interactive and Dynamic Records* (Padua, 2008), 832.

17 Elizabeth Shepherd and Geoffrey Yeo, *Managing Records: A Handbook of Principles and Practice* (London, 2003), 65.

levels. We gave a simple example of claims handling in an insurance company. The activity of handling a claim from a customer comprises a number of steps, and each step generates a record. The records of the various steps can be aggregated to form a record representing the activity in total. Thus, we argued, records exist at two levels: the elementary level of the individual steps and the higher level of the activity. Moreover, the company handles many claims over time; the handling of each claim is an instance of a claims-handling process. The records of all the claims can be aggregated in their turn to form a record at a third level, representing the process as a whole.[18]

When records are managed in paper form, attempts can be made to manifest these aggregations physically as paper files or as the contents of cabinets or other storage units. With computer systems, underlying storage arrangements are usually driven by technical economy, but related item-level records may appear to be brought together in electronic folders in the user interface. At upper levels, files or folders may be brought together in a series.

In *Managing Records*, Shepherd and I were cautious about identifying the records of particular activities or transactions as "files," but we offered a definition of a "series" as "the records of all the activities that are instances of a single process." We suggested that series are not defined by location or size but by their relations to particular processes.[19] With this move, we sought to identify the series in conceptual terms. We perceived it as an aggregate record that has the potential to be manifested as a physical unity, even if it is not always manifested thus in practice. Terry Eastwood and Chris Hurley have advocated broadly similar views.[20] From this perspective, even if no attempt has been made to assemble a record of a process physically and maintain it within a single paper or digital system, its components nonetheless form a single series. Conceptually, the series grows while the process continues; its expansion ceases when the process is terminated.

Most definitions of series are less precise than the definition offered in *Managing Records* and assume that the common characteristic of a series need not be limited to functional relationships. In this view, some series are composed of elementary records that result from a single process or activity, but others comprise records related in some other way. The definition

18 Ibid., 52–55. We attempted to supply more precise definitions of "activity" and "process" than records management literature usually provides. We also argued (p. 65) that, at the highest levels, records of processes can be aggregated to form records of business functions and of the work of the entire organization.

19 Ibid., 65, 86–88.

20 Terry Eastwood, ed., *The Archival Fonds: From Theory to Practice* (Ottawa, 1992), 11; Chris Hurley, "Relationships in Records" (c. 2004), http://www.infotech.monash.edu.au/research/groups/rcrg/publications/relationships-in-records-rev-3b.rtf, sec.R1.11 (accessed 1 December 2011). See also Dan Zelenyj, "Linchpin Imperilled: The Functional Interpretation of Series and the Principle of Respect des Fonds," *Archivaria* 42 (1996): 126–36.

offered by Kathleen Roe suggests that the components of a series may "result from the same function or activity" or may "have a particular form." Laura Millar affirms that they "relate to the same processes" or have "a common form, purpose or use."[21] North American descriptive standards (*RAD*, *APPM*, and *DACS*) allow series to be related by subject matter as a further alternative to function and form. Many of these definitions conclude with a phrase permitting other shared characteristics; *DACS*, for example, suggests that the contents of a series might share "some other relationship arising out of their creation, receipt, or use."[22]

As Dan Zelenyj remarked in 1996, definitions such as these are "intentionally broad in order to provide flexibility in interpretation." Zelenyj contended that ascribing series "on the basis of form or subject is to ... flout some of the most elemental archival concepts."[23] This contention may not win universal acceptance, but it is clear that definitions of the kind that Zelenyj attacked do not support an understanding of the series as a representational record. A series as Shepherd and I characterized it does not merely "relate to" but *represents* a process; however, no series can be a representation of its purpose or use.

In Millar's definition, the series appears to be conceptual; nothing is said about physical arrangement. However, many definitions introduce a further requirement that the components of a series must be brought together or organized in a certain way. *APPM* defines a series as "file units or documents *arranged in accordance with a filing system or maintained as a unit because* they relate to a particular subject or function, result from the same activity, have a particular form, or because of some other relationship ..."[24]; many other North American definitions of series similarly insist on connections between filing arrangements and a list of possible common characteristics.

All these definitions are presumably based on observation of groupings of records held in repositories and the ways in which these groupings had been shaped by their creators. Even if functional series are often felt to be desirable, real-world groupings are not always functionally based. Regardless of whether the definitions make specific mention of filing, their wording is intended to acknowledge that creators often arrange records on the basis of form, subject,

21 Kathleen D. Roe, *Arranging and Describing Archives and Manuscripts* (Chicago, 2005), 61; Laura A. Millar, *Archives: Principles and Practices* (London, 2010), 147.

22 Bureau of Canadian Archivists, *Rules for Archival Description*, rev. version (Ottawa, 2008), http://www.cdncouncilarchives.ca/RAD/RADComplete_July2008.pdf, Appendix D (accessed 1 December 2011); Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, 2nd ed. (Chicago, 1989), 8; Society of American Archivists, *Describing Archives: A Content Standard* (Chicago, 2004), 206–7.

23 Zelenyj, "Linchpin Imperilled," 127–28.

24 Hensen, *Archives, Personal Papers, and Manuscripts*, 8 (my italics).

or other criteria.[25] The underlying assumption is of a world in which records are controlled by means of physical ordering. In this world, definitions that confine the series to functionally based groupings are predicated on a capacity to maintain records in strict accordance with ideas on functional classification.

All the definitions so far discussed imply that physical series have some level of intellectual coherence, but we also encounter other definitions that lack explicit conceptual underpinning. For example, most definitions that circulate in the records management community view series largely as physical units convenient for retention scheduling. In a recent textbook, a "records series" is described merely as "a group of related records filed and used together as a unit and evaluated as a unit for retention purposes."[26] When digital curators characterize a series as any assemblage of digital material "that a user chooses to collect in one bucket," conceptual understandings are absent and the series is perceived only in physical terms.[27]

These issues are not unique to series; similar uncertainties arise at other levels of aggregation. For example, the recordkeeping metadata standard ISO23081 defines a file as "a sequence of items, physically or virtually linked, which evidences an organizational/business activity" and asserts that "individual items on the file have relationships ... which are preserved by being kept on file in the right order."[28] In this standard, the file is defined conceptually in terms of its evidential connection to activity, but its internal relationships are seen as both conceptual and actual. When ISO23081 affirms that files can be either analog or digital, the emphasis clearly shifts to the file as an actual object or set of objects. The reference to keeping items "in the right order" makes sense only if the file is perceived as a physical container. The question of what constitutes "the right order" is unanswered (and perhaps unanswerable), but the definition implies that intellectual qualities are embedded in physical arrangements and sequences and that conceptual understandings of the file coincide with real-world aggregations.

25 For an examination of the range of criteria that may be used, see Deborah K. Barreau, "Context as a Factor in Personal Information Management Systems," *Journal of the American Society for Information Science* 46, no. 5 (1995): 327–39; Deborah Barreau, "The Persistence of Behavior and Form in the Organization of Personal Information," *Journal of the American Society for Information Science and Technology* 59, no. 2 (2008): 307–17.

26 Judith Read and Mary Lea Ginn, *Records Management*, 9th ed. (Mason, OH, 2011), 419. Similar definitions can be found in Mary F. Robek, Gerald F. Brown, and David O. Stephens, *Information and Records Management*, 4th ed. (New York, 1995), 585, in ARMA International's *Glossary of Records and Information Management Terms*, 3rd ed. (Lenexa, KS, 2007), 21, and elsewhere.

27 Patricia Hswe et al., "The Web Archives Workbench (WAW) Tool Suite: Taking an Archival Approach to the Preservation of Web Content," *Library Trends* 57, no. 3 (2009): 450.

28 ISO23081-2:2009 *Metadata for Records: Part II: Conceptual and Implementation Issues*, 11.

In practice, a physical file in the paper world may contain records of an activity, records of two or more activities, or a part of the records of a single activity; alternatively, it may contain records related by form, subject, date, or some other criterion. Depending on circumstances, records "which evidence an organizational/business activity" may be housed in one file or distributed across many files.[29] Sometimes – for example, if they are of unusual size or shape – they may not be housed in files at all. In hybrid environments, they may also be divided between paper and digital storage systems.

In everyday English, the term "file" usually connotes a particular type of physical container. Attempts to use it to refer to a *conceptual* level of aggregation often give rise to ambiguity. We might do better to follow the French tradition of using the term "dossier" to describe conceptual units that may or may not have a one-to-one correspondence with physical storage containers. In paper systems, we could then unambiguously use "file" to refer to physical devices that facilitate the housing, retrieval, movement, and inspection of documents.[30]

Unfortunately, no such alternative readily presents itself in the case of the word "series."[31] "Series" has to do duty in both a conceptual and a physical sense; the lack of separate terminology doubtless reinforces a tendency to confuse the two. Since definitions of series run a gamut from the wholly conceptual to the wholly physical, it is often uncertain whether archivists who speak of a series are alluding to a "conceptual record of a process," a "physical aggregation that realizes a conceptual record of a process," a

29 Shepherd and Yeo, *Managing Records*, 81–88. The issue extends beyond the necessity to divide "fat" paper files into file parts or volumes; in a "day file" system, for example, the records of an activity extending over many days will normally be distributed across many files, even (or especially) if each file is "thin." Moreover, where recordkeeping systems are poorly designed, paper files (like physical collections at any level of aggregation) may contain assortments of apparently unrelated records, perhaps intermixed with unused forms and envelopes, advertising circulars, magazines, or other materials.

30 Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, MD, 1998), 120–21, citing Gérard et Christiane Naud, "L'Analyse des archives administratives contemporaines," *Gazette des Archives* 115 (1981). This solution is perhaps not wholly unambiguous: in English, there is still the problem of the different meaning of "file" in information technology. In paper environments, however, the term "file" remains redolent of the stationery cabinet; archivists' use of it as a generic label for a level of aggregation seems to reflect their experience of twentieth-century methods of bureaucratic recordkeeping. In practice, item-level paper records can be stored not merely in files but in bundles, boxes, and a variety of other containers, which do not always fit neatly into archivists' standard hierarchy of files and series.

31 In the UK, archivists have sometimes used "class" as a synonym or near-synonym for "series" (see Michael Cook, *The Management of Information from Archives*, 2nd ed. (Aldershot, England, 1999), 111), but this practice conflicts with the growing use of "class" to refer to *entity classes* in relational and object-oriented modelling and in standards such as ISO23081.

"physical aggregation that displays one or more of several possible common characteristics," or simply a physical aggregation irrespective of its character. In archival discourse in English, unless the use of "dossier" is adopted, understandings of the file also fluctuate between the conceptual and the physical.

Of course, records managers and archivists can design systems that attempt to ensure a coincidence between conceptual and physical entities. In its *Principles and Functional Requirements for Records in Electronic Office Environments*, the International Council on Archives (ICA) provided records managers with a "basic model" that equated *series* with records of a "business function," *files* or *containers* with "activity," and *items* with "transaction" (though it also recognized that "aggregation to more than three levels may be necessary ... for … complex topics").[32] In *Managing Records*, Shepherd and I recommended a similar approach to intellectual and physical arrangement at series level, but we proposed greater flexibility at levels below series to accommodate user requirements for other groupings.[33] Such methods need not be limited to systems for organizing records at the time of creation or initial capture. At later moments, archivists faced with apparently chaotic accumulations of papers may decide to arrange them in physical series or files that purport to represent functions, processes, or activities. Others reject this approach and insist that, following the principle of original order, establishment of files and series is a prerogative of the records creator and the archivist's task is not to impose order retrospectively but to maintain the structure (or lack of structure) that already exists. Conflicts can emerge between traditional emphases on archival non-intervention and the newer world in which recordkeeping professionals attempt to ensure the arrangement of records on a functional basis.

Since conceptual and physical files or series do not always coincide, many of the issues that I discussed in "The Conceptual Fonds and the Physical Collection" recur at file and series level. Just as investigations of the fonds are sometimes impeded by uncertainty as to whether it is a physical or a conceptual grouping, there are tensions between physical and conceptual understandings of the series. If we seek to resolve these tensions, we must acknowledge that, like conceptual fonds, conceptual series are not always realized physically. Like the components of a fonds, the components of a conceptual series need not be co-located. They may be distributed between onsite and offsite storage, different divisions of an organization, or paper and

32 International Council on Archives, *Principles and Functional Requirements for Records in Electronic Office Environments, Module 2: Guidelines and Functional Requirements for Electronic Records Management Systems* (2008), http://www.adri.gov.au/products/ICA-M2-ERMS.pdf, 20 (accessed 1 December 2011).

33 Shepherd and Yeo, *Managing Records*, 81–85.

digital media. They may be in disorder, scattered randomly across numerous locations. Just as the aggregations that archivists describe as "fonds" are often parts of larger conceptual fonds that have been divided or diminished, the archivist's "series" may be part of a larger conceptual series that also embraces more recent files or items held in a current records system. Many aggregations described as "series" are residues of larger series that have been reduced by acts of selection.

In 1994, Sue McKemmish affirmed that "the record is only partly manifest in what is ... on the repository shelves" and that archivists "may need to liberate themselves from the notion that the series is a physical grouping of records."[34] From this perspective, the series is a conceptualization that transcends most physical groupings. Nevertheless, it would be unwise to ignore the physical aggregations that are often labelled as "series" and "files." In the paper world, these are often necessary for pragmatic reasons; large undifferentiated assemblages are almost always unmanageable and have to be broken down into smaller physical units for administrative control and retrieval. Hierarchies are formed, with each series comprising one or more files and each file containing several documents. I know of no definitions of "series" or "file" that employ the word "collection," but I would argue that physical series and files can be identified as collections (or sub-collections) and that their attributes and roles are similar to those of the collections discussed in my earlier article. They are intentionally brought together and their boundaries are determined by collecting decisions. Their contents indicate the combinations that their collectors thought might constitute a suitable aggregation for some practical purpose. Physical series and files are also the configurations that users experience. Conceptual series or dossiers are not immediately available to the user, except insofar as they have been realized physically. In a prototypical retrieval system, users encounter physical series as the frameworks within which access is granted and physical files as the tangible units delivered for inspection. In the world of paper archives, collections are brought together, hierarchies are established, and arrangement is fixed before the user arrives on the scene.

Digital Dynamics

In recent years, many critics have argued that the digital revolution overturns established conventions and offers new modes of cultural understanding and opportunities to see the world in different ways. In 1995, the textual scholar

34 Sue McKemmish, "Are Records Ever Actual?" in *The Records Continuum: Ian Maclean and Australian Archives First Fifty Years*, ed. S. McKemmish and M. Piggott (Clayton, Victoria, Australia, 1994), 200–1.

Jerome McGann drew on ideas developed in the 1960s and 1970s by computer visionary Ted Nelson to argue that, in a world of hypertext, "historical orderings" of documentary materials are replaced by systems in which each document can "be connected to every other document ... in any way one chooses." "Centralized or hierarchical structures" are arbitrary, McGann asserted, and users should be "encouraged not so much to find as to make order."[35] The visual culture theorist Lev Manovich pursued similar themes in 2001, when he wrote that hierarchical file systems assume "that the world can be reduced to a logical ... order, where every object has a distinct and well-defined place." Manovich advocated a new model in which all objects are deemed equal and no single way of structuring experience has any special status.[36] Elena Esposito, in 2002, wrote of dynamic digital models where considerations of "meaning" play no part in organizing materials until users ask to manipulate them. For Clay Shirky, arguing in 2005 that "we don't need ... hierarchy," Google's search engine provided an archetypal example: "Google can decide what goes with what after hearing from the user, rather than trying to predict in advance what it is you need to know."[37]

Probably the best known of writings in this vein is David Weinberger's *Everything is Miscellaneous*, published in 2007. According to Weinberger, paper environments limit us to a single ordering. In Weinberger's terminology, this is the "first order of order," in which we recognize the impossibility of arranging physical objects in more than one sequence simultaneously. In the "second order of order," which is also characteristic of paper environments, we can have a small number of alternative sequences, but only by using card indexes or other laborious representational surrogates. In the digital world, however, multiple needs can easily be achieved without duplication of effort, and we are no longer constrained by necessities to predefine the range of possible orderings or give priority to those orderings that seemed important in the past. Weinberger quoted the president of the Encyclopaedia Britannica Publishing Group, who announced that predetermined chronological or alphabetical "ways of ordering events and ideas no longer seem so incontrovertible, so natural." Instead, his company aimed to produce an electronic encyclopedia "composed of small units of information" that can be organized in what-

35 Jerome McGann, "The Rationale of Hypertext" (1995), http://www2.iath.virginia.edu/public/jjm2f/rationale.html (accessed 1 December 2011). Nelson's ideas on hypertext owed much to a famous paper by Vannevar Bush, published at the end of the Second World War (Bush, "As We May Think," *Atlantic Monthly* 176 (1945): 101–8).

36 Lev Manovich, *The Language of New Media* (Cambridge, MA, 2001), 16, 219–20.

37 Elena Esposito, *Soziales Vergessen* (Frankfurt, 2002), cited by Rudi Laermans and Pascal Gielen, "The Archive of the Digital An-archive," *Image [&] Narrative* 17 (2007), http://www.imageandnarrative.be/inarchive/digital_archive/laermans_gielen.htm (accessed 1 December 2011); Clay Shirky, "Ontology is Overrated: Categories, Links, and Tags" (2005), http://www.shirky.com/writings/ontology_overrated.html (accessed 1 December 2011).

ever way the user wants. This is Weinberger's "third order of order," in which resources can be arranged into as many sequences as may be desired and users can organize their work independently of the limitations imposed by analog systems.[38]

The digital world also allows collections to be built dynamically as and when they are needed. Instead of preordaining the groupings our users will encounter, we can employ technologies that make it easy to organize materials into multiple collections that reflect users' individual interests. Innumerable online resources now work in this way, encouraging visitors to use metadata or thumbnail images to build their own collections as they want them. To take just one of many possible examples, the website of the University of Kent, England, displays images of British cartoons;[39] users can assemble collections of cartoons by particular artists, from particular newspapers, depicting particular people or subjects, or published at particular dates; they can also select their favourites to build a collection based purely on personal choice. In digital spaces of this kind, many of them inspired by larger-scale ventures such as the photo-sharing site Flickr, no judgment is made that any one arrangement is better or more "correct" than another. Users can group and regroup, order and reorder, at will. In Shirky's language, there are no "officially approved choices."[40]

This, we are told, is the world of Web 2.0, of Gilles Deleuze and Félix Guattari's rhizomes, of Vilém Flusser's co-operative network dialogues:[41] a world in which democratic production and participatory experience replace older paradigms of centralized control. The rules that seemed to determine the shaping and stabilizing of collections in earlier "orders" are absent or diminished. Hierarchies are said to be broken down, linear connections lose their monopoly, and binary choices disappear. Users make their own orderings, select their own collections, construct their own narratives in whatever way they wish. "Content" can be repurposed or reused as desired. Freed from traditional constraints, forms of culture previously controlled by an elite are popularized and seemingly made available to everyone.

Archival Practice and the "Third Order"

To archivists, these ideas may appear disquieting. Questions inevitably arise about their implications for archival practice. Are traditional fixed arrangements becoming obsolete? Should we now encourage users of archives

38 David Weinberger, *Everything is Miscellaneous: The Power of the New Digital Disorder* (New York, 2007); the quotations are from p. 31.

39 See http://www.cartoons.ac.uk (accessed 1 December 2011).

40 Shirky, "Ontology is Overrated," unpaginated.

41 Gilles Deleuze et Félix Guattari, *Mille Plateaux* (Paris, 1980); Vilém Flusser, *Kommunikologie* (Frankfurt, 1996), cited by Fuchs, *Internet and Society*, 240–42.

to group and regroup "atomic" objects at will and build their own collections on demand, to make their own order rather than discover orders that already exist? If "online distribution of culture challenges traditional 'off-line' formats," as Manovich suggested,[42] can our familiar hierarchies of predefined collections and sub-collections be seen as outmoded in a digital era?

In the paper world, we had little choice but to impose single orderings on our holdings. At best, we were restricted to Weinberger's "second order of order." We knew, for example, that archival holdings need not be arranged on shelves in the sequence used to describe them in a finding aid, but we also knew that there were practical limits to the number of alternative orderings we could achieve. Many experts affirmed that there was a single right order – a "correct" way to organize records – and that the archivist's role is to identify this single correct arrangement and present it in the finding aid. There also seemed to be little room to dispute what the basis of this arrangement should be. The archivist's responsibility, we were informed, was to preserve (or, if necessary, reconstitute) and stabilize the particular configuration that reflected the work of the records creator.

Following these precepts, archivists privileged particular groupings of records (those whose shared characteristic is creatorship or provenance); within such groupings they sought to protect the internal arrangement imposed on the records by those who created them. In Italian archival theory, these ideas are associated with the notion of an "archival bond," a *vincolo originario* between documents that allows and demands a particular ordering that is (supposedly) objective, inevitable, and non-arbitrary.[43] Ordering that reflects this bond might be imposed through physical storage or by using a classification scheme in which each item logically occupies a single place; the necessity of preserving it and securing its fixity gives us the principle of original order, which is said to be the only valid basis for the arrangement of archives.[44]

In traditional practice, this ordering usually has two aspects: identification of the files or other containers into which records are to be grouped and identification of the sequence in which the contents of each container are juxtaposed and presented to users. Containers keep related items together, while presentation sequences are believed to preserve evidence of specific linkages between one item and the next. In the digital world, physical arrangements within storage media are arbitrary, but recordkeeping systems may introduce virtual containers or presentation sequences at the level of the user interface so that items appear to be organized in a way that reflects a perception of the bonds among them. Outside the recordkeeping domain, designers of interfaces

42 Manovich, *The Language of New Media*, 333.
43 Lodolini, *Archivistica*, 127–28.
44 Ibid., 127.

to "content management systems" are less likely to impose static presentation sequences, but often provide hierarchical folder structures that allow "content" to be grouped in ways that broadly correspond to archival practices. Other orderings, it is alleged, do not objectify the implicit bonds between records that arise from their intimate connection to organizational business or personal life. Elementary records sometimes lack meaning when viewed in isolation, and their aggregation into immutable higher-level units determined by their creators is often claimed to support their usability, their contextualization, and perhaps also their authentication.

Several recent commentators have used arguments like these to dismiss dynamic or "ordering on demand" approaches as irrelevant to evidential recordkeeping. Fiorella Foscarini, for example, has asserted that, from a records management viewpoint, we need an "organizing principle which determines, once and for ever, how records accumulate" and that only a hierarchical classification scheme can serve this purpose. From this perspective, "ordering on demand" endangers evidentiality because it destroys the fixed single arrangement that supposedly reflects a natural ordering of records and provides essential information about their context.[45]

Unsurprisingly, archivists whose frame of reference is informed by postmodernism take a different view. For writers such as Brien Brothman, organizing records in a particular way is merely an attempt to tell one particular story out of many possible stories. According to Brothman, arrangement by provenance is "not coincident with any natural informational order, because there is nothing 'natural' about classification systems or file order. Information ordering is social, not natural."[46] Where "original" order is concerned, archivists in this school of thought sometimes suggest that there may be many orders, all perhaps equally subjective, and that it makes no sense to claim that a single order, even one imposed with the aim of objectifying contexts of creation, should invariably be privileged over others. In Lara Moore's words, "any classification scheme inevitably creates its own stasis, preventing alternate conceptions of the same objects and ideas." Such views resonate with Weinberger's assertion that "in the third order of order, knowledge doesn't have *a* shape. There are just too many ... ways to make sense of our world."[47]

45 Fiorella Foscarini, "Function-based Records Classification Systems: An Exploratory Study of Records Management Practices in Central Banks" (PhD diss., University of British Columbia, 2009), 50, 58, 288. Cf. Luciana Duranti, "More than Information, Other than Knowledge: The Nature of Archives in the Digital Era," *Cadernos BAD* 2 (2003): 12–13; Ruth Frendo, "Disembodied Information: Metadata, File Plans, and the Intellectual Organisation of Records," *Records Management Journal* 17, no. 3 (2007): 157–68.

46 Brien Brothman, "Orders of Value: Probing the Theoretical Terms of Archival Practice," *Archivaria* 32 (1991): 84.

47 Lara J. Moore, *Restoring Order: The Ecole des Chartes and the Organization of Archives*

It is evident that we now face some critical theoretical and practical questions. Must we refuse to extend the ideas and practices of the "third order" to the archives and records management disciplines? Do the folder model and the classification scheme still deserve the paramountcy they have long enjoyed in records management? Can archivists still cling to the beliefs that fixed hierarchies of files and series are essential and that "the core values of the profession ... prescribe a hierarchical approach"?[48] Must levels of aggregation always be defined in advance of episodes of use, if archival principles are to be maintained? I wish to argue that the answer to all these questions is no, and that – as the Australian "series system" has already shown us[49] – we can respect provenance in ways that do not depend on the stability of particular orderings of the world. Far from feeling threatened by the fluidity of the "third order," archivists and records managers should be able to take advantage of the new capabilities to help overcome some of the contextual limitations of hierarchical classification schemes and paper-world methods.

Process Boundaries and Conceptualizations of Series

To take these arguments forward, it may be helpful to review some of the constraints that traditional practices impose. Over the past thirty years, several researchers outside the archives and records domain have suggested that people who are not classification specialists often find classification systems cognitively challenging and difficult to apply.[50] In records management, it has long been known that many users dislike *functional* classification schemes or find them unhelpful.[51] In an Australian study, users reported that a functional scheme employed in government agencies was cumbersome, unintuitive, and hard to comprehend.[52] In digital environments, however, appropriate use of

and Libraries in France, 1820–1870 (Duluth, MN, 2008), 129; Weinberger, *Everything is Miscellaneous*, 83.

48 Jenn Riley and Kelcy Shepherd, "A Brave New World: Archivists and Shareable Descriptive Metadata," *American Archivist* 72, no. 1 (2009): 110.

49 The "series system" allows a series to be linked to as many different record-creating entities as context documentation requires; see Adrian Cunningham, ed., *The Arrangement and Description of Archives amid Administrative and Technological Change* (Brisbane, 2010). I discussed the series system in Yeo, "The Conceptual Fonds," 64–65.

50 Thomas W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems* 1, no. 1 (1983): 99–112; M. Lansdale, "The Psychology of Personal Information Management," *Applied Ergonomics* 19, no. 1 (1988): 55–66; Olha Bondarenko and Ruud Janssen, "Documents at Hand: Learning from Paper to Improve Digital Technologies," in *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, 2005), 121–30.

51 Stuart Orr, "Functions-based Classification of Records: Is it Functional?" *Archives and Manuscripts* 34, no. 1 (2006): 61–63, 80.

52 Pauline Singh, Jane E. Klobas, and Karen Anderson, "Information Seeking Behaviour of

third-order methods can be expected to offer much higher levels of responsiveness to users' needs.

Consider some of the dilemmas records managers face when they attempt to design and implement a classification scheme in the paper world. Should the scheme be based entirely on functions, processes, and transactions? Or should it acknowledge users' possible preferences for other arrangements and group certain records by date, client, or creating department? Current orthodoxy in records management suggests that structural arrangements should not be used; even when responsibility for a function or process is perceived to extend across several departments of an organization, records managers usually recommend functional classification on the grounds that structural change is frequent and functional groupings are considered more stable. Foscarini's research, however, revealed substantial evidence of user preferences for structural groupings of records.[53] Internal accountability considerations or local business needs often require aggregate records of all the work done by a division, department, workgroup, or individual employee, even though such groupings may exclude records of functionally related work performed elsewhere in the organization. An organization may also have legitimate requirements for aggregate records of all the business conducted at particular locations or with particular clients or suppliers, or of all the work done in specific time periods. All of these criteria offer possible bases for assembling a record and each of these groupings carries aspects of context,[54] but traditional classification schemes oblige records managers to select only one of them.

Let us suppose that I operate a small business and decide on a functional scheme for the business's records. Inescapably, I find that further decisions must be made. Where do I think the boundaries of each function or process lie? When I repaint my rusty delivery truck with my new corporate colour scheme and logo, am I contributing to a vehicle maintenance function or a marketing function? Arguably, the elementary record of the repainting activity – the photograph of myself with paintbrush in hand, or the sign-off sheet completed by the painter I employed – forms part of a high-level record of both these functions. Faced with a classification system that requires me to file it in a single location, I am likely to be uncertain which file or folder to choose.

Electronic Records Management Systems (ERMS) Users," *Human IT* 9, no. 1 (2007): 166.

53 Foscarini, "Function-based Records Classification Systems," 263–73.

54 Although non-functional groupings are unfashionable among records managers, aggregate records do not have to be based on a purely functional arrangement. The work done by a particular team (or at a particular place, or on a particular date) can be identified as an occurrent, and a group of items representing it would seem to constitute a record just as much as a group of items representing a particular process.

Records management literature often presents functional or process analysis as an exact science, but it has become increasingly clear that this view is unsustainable. In the words of business consultant Fred Nickols, "analyses of processes are ... adrift in a sea of undefined terms, unclear boundaries, different perceptions and experiences."[55] Several studies have demonstrated that, when processes are under investigation, disputes about borders are normal. Jan Fülscher and Stephen Powell, for example, reported that, when a Swiss team began to analyze an insurance sales process, "almost immediately a controversy arose over the appropriate boundaries for the process"; participants could not agree whether it included product development or marketing activities as well as interaction with customers.[56] Similarly, in the world of health care, analysts may be uncertain whether the diagnosis and treatment of patients are one process or two. If patients or their insurance companies are billed, is the billing part of the same process or a separate financial process? Some observers may assume that processes are confined within single departments of an organization, but others may seek to identify broader processes that transcend organizational structures.[57] In many organizations, procedure manuals, workflow charts, or software configurations imply that processes are clearly bounded, but such precision derives only from the work of the developers of the manual, the chart, or the software. Even legal or regulatory definitions of processes, which seek to introduce firm boundaries where they would otherwise be absent, can be subject to revision if regulators change the boundaries

55 Fred Nickols, "The Difficult Process of Identifying Processes," *Knowledge and Process Management* 5, no. 1 (1998): 17.

56 Jan Fülscher and Stephen G. Powell, "Anatomy of a Process Mapping Workshop," *Business Process Management Journal* 5, no. 3 (1999): 214–15. Cf. Nickols, 17; Andy Crabtree, Mark Rouncefield, and Peter Tolmie, "There's Something Else Missing Here: BPR and the Requirements Process," *Knowledge and Process Management* 8, no. 3 (2001): 166.

57 Manuel Laguna and Johan Marklund, *Business Process Modeling, Simulation and Design* (Upper Saddle River, NJ, 2005), 114–16. It is instructive to compare attempts to impose precise definitions of process boundaries with the long-running debate about "maximalist" and "minimalist" notions of the fonds (Michel Duchein, "Theoretical Principles and Practical Problems of Respect des Fonds in Archival Science," *Archivaria* 16 (1983): 64–82; Terry Eastwood, "Putting the Parts of the Whole Together: Systematic Arrangement of Archives," *Archivaria* 50 (2000): 93–116), which is also driven by a perceived need to make authoritative decisions about boundaries. The paradigm for the maximalist position is perhaps the practice in the former Soviet Union of identifying the records of the entire Soviet state as a single fonds; the minimalist position might attribute fonds status to the records of each minor subordinate agency. Both positions assume a need for a single "officially approved choice." However, as noted by Chris Hurley, "Documenting Archives and Other Records" (2008), http://www.infotech.monash.edu.au/research/groups/rcrg/publications/ch-documenting-archives.pdf, 4 (accessed 1 December 2011), what is perceived as a fonds from one viewpoint can be a sub-fonds from another. The participants in Fülscher and Powell's process mapping study likewise discovered that a process can be a sub-process from another point of view. The choice of labelling is arbitrary except insofar as it is driven by politics or custom.

they previously imposed. Processes and functions are not objective and mutually exclusive realities, but constructions whose precise borders exist only in the minds of bureaucrats, analysts, or other observers.

If the boundaries of a patient care process or a financial process are a matter of interpretation, then the limits of the record series that represent them must also be in the eye of the beholder. Instead of a single series representing (for example) the care process, it is possible to conceptualize a multiplicity of series, each corresponding to a different understanding of the process boundary. It should also be possible to recognize that series membership can overlap: that a patient billing record can be perceived *both* as a component of one or more conceptual series representing a patient care process *and* as a component of one or more conceptual series representing a financial process. Classification schemes that attempt to identify mutually exclusive series by precise determination of process boundaries are necessarily arbitrary; they give an illusion of definiteness where none exists.[58] If records managers seek to evade these difficulties by producing classification schemes that lack precision, they merely defer the burden of arbitrary decision-making to the moment when records are filed.

Records managers (and archivists) may also find that a process can be represented by many overlapping series even when they, or others to whose authority they defer, have tried to assign it a single boundary. A financial process, for example – even one that has been closely defined – may be associated with a series of records designated for formal retention, a series that also includes ephemeral records scheduled for routine destruction, or a series that encompasses instructions for carrying out the process as well as records of its execution. All of these are possible conceptual series whose common characteristic is their connection to the process in question.

If we accept that multiple representations are the norm, any lingering notions of the series as a single definitive record aggregation must be abandoned.[59] Even in the paper world, elementary records are not always physically grouped into series, but we can often perceive conceptual series even when physical groupings are absent or disturbed. In digital environments, predefined physical series may be largely redundant, but conceptual series can still be meaningful. Moreover, just as my earlier paper suggested that over-

58 I have borrowed the phrase "illusion of definiteness" from Louis Bucciarelli's study of project charts, cited by Kjeld Schmidt, "Of Maps and Scripts" (1997), http://www.itu.dk/~schmidt/papers/maps_and_scripts.pdf, 140 (accessed 1 December 2011).

59 Cf. Jim Suderman, "Defining Electronic Series: A Study," *Archivaria* 53 (2002): 31–46. In 2008, I wrote that "records of individual steps can be aggregated to form the record of the activity" and that records of activities can be aggregated "to form the record of the process as a whole" (Yeo, "Concepts of Record (2)," 133), but my choice of language was ill advised; in many cases, such acts of aggregation form just one of a number of possible records of possible understandings of an activity or process.

lapping conceptual fonds can be realized if we have the ability to construct aggregations on demand,[60] overlapping conceptual series can also be realized if we are able to use the capabilities of the third order. When we are no longer obliged to identify a single "best" ordering, deterministic notions of series become obsolete; instead, we may perceive them as overlapping conceptualizations that we can seek to realize as the need arises.

From an archival viewpoint, these conceptualizations all reflect different ways in which we might understand or manifest bonds or relationships among records. Their variety also indicates the range of different arrangements and physical orders that creators or records managers could have chosen in an analog environment where selection of a single "original" order was required. It turns out that there are not single "inevitable" groupings that emerge naturally at the point of creation but many possible groupings. Each could be seen as valid, but no single arrangement captures all the interconnections that might be of interest to us.

We might even imagine conceptual series extending beyond the borders of a single organization to embrace records kept by other participants in a process. To take an extreme example: it has been reported that obtaining permission to build houses on state-owned land in Peru requires 207 "administrative steps" in 52 separate government offices.[61] We can surmise that each office might require records only of the steps in which it is directly concerned, but from the point of view of the applicants – and of the researcher investigating bureaucratic obstacles to house building – all 207 steps may constitute a single process. Archivists and records managers have traditionally assumed that series do not transgress organizational boundaries; however, if processes can reach across a number of organizations or a wider community, we may wish to perceive conceptual series that are similarly extensive. Activities or transactions can also cross organizational borders, and commentators who seek to label records of activities as "dossiers" might also acknowledge that the components of a conceptual dossier need not be restricted to elementary records maintained by a single organization. Series and dossiers that correspond to such wider views of processes and activities may not be easy to realize, since they breach the limits of the physical collections to which we have been accustomed in a world where one organization is discrete from another. Nevertheless, we should accept that – unlike physical groupings – conceptual series and dossiers need not reside within a conventional part-whole structure.

Paper systems and hierarchical folder models struggle to cope with elementary records that can be imagined as components of more than one higher-level record. Since a paper item can only be in one place at any one time, conflicting

60 Yeo, "The Conceptual Fonds," 74.

61 Hernando de Soto, *The Mystery of Capital* (New York, 2000), 19–20.

demands cannot easily be satisfied at a physical level; arbitrary choices have to be made when items are brought together to form aggregate records. By way of contrast, in the digital environment we no longer have to "make binary decisions about where things go."[62] Like the boundaries of conceptual fonds, the boundaries of conceptual series are open to varying interpretation. When third-order systems are in place, I should be able to constitute a series with whatever boundaries I consider appropriate; if you take a different view, you should be able to form a series that matches your own perception. Elementary records should be linkable in ways that folder-based approaches cannot easily replicate and aggregate records realizable by constructing collections on demand.

Earlier in this paper I suggested that, in principle, an aggregate record might be brought together by anyone with the interest and the means to do so. Traditionally, however, archivists and users have expected to find aggregate records ready assembled and hierarchical levels defined in advance of episodes of use. Third-order methods now offer the potential for radical change. There need be no predetermined structure, if tools are available for users to build an aggregate record when it is required.

Alternative Groupings

According to the ICA's *Principles and Functional Requirements*, "an aggregation of ... records may collectively constitute a narrative of events."[63] However, as postmodernist writers often remind us, any narrative is only one of many possible narratives. If an elementary record can be a component of many different aggregations, we need not confine it to those aggregations whose narratives are of events. Besides using it in assembling aggregate records representing events, processes, or other occurrents, we can also use it in constructing groupings that serve other purposes. An aggregation that represents an occurrent may be assumed to support needs for evidential record-keeping; an alternative grouping may be more helpful when other needs are to be met.

Several studies have shown that individuals engaged in multi-tasking frequently want to group items together in ways that assist task prioritization. Research by Barbara Kwasnik has demonstrated that, in office environments, people often find it helpful to employ "intended use" as an initial basis for arrangement. According to Olha Bondarenko and Ruud Janssen,

62 Weinberger, *Everything is Miscellaneous*, 83. Digital objects cannot occupy multiple locations simultaneously, but the abilities of computers to support pointer systems, and to generate seemingly indistinguishable copies of digital objects, largely serve to eliminate the kind of binary decision-making that was often unavoidable in the paper world.

63 International Council on Archives, *Principles and Functional Requirements, Module 2*, 18.

"by (re)grouping their documents people ... support ... task management" and identification of the urgency of tasks to be performed.[64] In many working situations, there may also be needs for groupings of items awaiting replies or other actions such as checking, scanning, indexing, or filing. Besides records creators, later users may also want to introduce more or less temporary physical groupings of this kind.

There is also evidence that, in the workplace, people often want to assemble items on the basis of their form or subject matter. Studies by Sarah Henderson and by Christopher Khoo et al. found that, in academic and office settings where folder systems were employed, groupings by "genre" or "document type" (both more or less equivalent to "form") were used more widely than any other mode of arrangement.[65] Possible needs for grouping by subject matter have been the focus of a long-running debate in the field of medical records: should the contents of a patient's file be grouped to match the hospital's structures or work processes or should they be oriented to the patient's medical problems? The former is often easier to implement and useful for short-term administration and costing, but the latter is thought to provide a firmer basis for long-term patient care.[66] In other contexts, too, users often favour subject arrangement for groupings that are intended as "permanent"; 43 percent of the responses in a study undertaken in an Australian city council indicated a preference for subject classification.[67] As noted earlier, the frequency with which groupings by form or subject are imposed in paper-world systems has influenced many archival definitions of "series"; groupings of this kind often constitute "original" orders, if an original order is an ordering made when elementary records are classified at or near the time of their creation. It is observable, however, that the studies cited above do not present a unanimous preference for a single type of ordering that could be universally supported when hierarchical classification schemes are employed. Within the

64 Barbara H. Kwasnik, "How a Personal Document's Intended Use or Purpose Affects its Classification in an Office," in *Proceedings of the ACM-SIGIR 12th Annual International Conference on Research and Development in Information Retrieval* (New York, 1989), 207–10; Bondarenko and Janssen, "Documents at Hand," 125. See also Barreau, "Context as a Factor."

65 Sarah Henderson, "Genre, Task, Topic and Time: Facets of Personal Digital Document Management," in *CHINZ '05: Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer Human Interaction* (New York, 2005), 77–78; Christopher S.G. Khoo et al., "How Users Organize Electronic Files on their Workstations in the Office Environment," *Information Research* 12, no. 2 (2007), http://informationr.net/ir/12-2/paper293.html (accessed 1 December 2011).

66 James F. Fries, "Alternatives in Medical Record Formats," *Medical Care* 12, no. 10 (1974): 871–81; Huibert Tange, "How to Approach the Structuring of the Medical Record?" *International Journal of Bio-medical Computing* 42 (1996): 27–34.

67 Tina Calabria, "Evaluating Caloundra City Council's EDMS Classification" (2004), http://www.steptwo.com.au/files/kmc_caloundracouncil.pdf, 6 (accessed 1 December 2011).

workplace, different groupings (and systems that support them) will bring benefits to different users.

Later users, more distant in time or space from the moment of records' creation, also have diverse wants and needs. Not all users seek evidence of the occurrents that records represent, or look for groupings of records based on contextual provenance. Because archivists know this, they sometimes provide subject indexes or thematic guides to supplement conventional finding aids. But these, too, are paper-world solutions with limited capacities. Using the more powerful tools available in the digital world, enthusiasts for coins, clocks, or cucumbers can put together collections of elementary records that mention their favoured topic, regardless of the records' provenance, as and when they wish. Assembling a collection of interest to a particular community group may encourage dialogues about community identity or improve relations with the community concerned. A collection based on documentary form may enhance knowledge of diplomatics. A recent paper by Sue Breakell discussed the formation of groupings based on colour or visual content; a paper by Tarez Samra Graban explored the possibility of forming assemblages of records on the basis of their rhetorical or discursive style.[68] Collections of items of a certain size or shape; presented in Helvetica font; discovered in the basement of Building X; rescued from a fire; employing a special photographic technique; written by a named author; or consulted by a particular official: all may be of value to one user community or another. Any of them might be formed by subdividing a larger collection whose common characteristic is provenance or by assembling discrete items across the boundaries of conceptual fonds or existing physical collections. Potentially, they might all overlap, but in the third order this no longer inhibits our ability to assemble them. When collections are dynamic, some may appear to be subsets of others, but subset hierarchies are no longer a constraint imposed by the system.

In fact, some aspects of the third order may not be as revolutionary as they seem. The sixteenth-century astronomer Tycho Brahe is said to have acquired new knowledge of the universe by combining and recombining data derived from the separate records and writings of the astronomers who preceded him. Two centuries later, Thomas Jefferson devised a rotating reading stand so that he could bring together varying selections of books from his library and examine them simultaneously.[69] Later still, the social reformers Sidney

68 Sue Breakell, "For One and All: Participation and Exchange in the Archive," in *Revisualizing Visual Culture*, ed. C. Bailey and H. Gardiner (Farnham, England, 2010), 106; Tarez Samra Graban, "Emergent Taxonomies: Using 'Tension' and 'Forum' to Organize Primary Texts," in *Working in the Archives*, ed. A.E. Ramsey et al. (Carbondale, IL, 2010), 206–19.

69 Bruno Latour, *Science in Action* (Cambridge, MA, 1987), 226–27; Matthew G. Kirschenbaum, "The Remaking of Reading: Data Mining and the Digital Humanities" (2007), http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf,

and Beatrice Webb and the socialist historian Raphael Samuel chose to use loose sheets of paper for their note-taking, so that they could shuffle and reshuffle the sheets as often as they wished and rearrange them in whatever way their work suggested or required.[70] At a more mundane level, systems for cross-referencing and for making duplicates that can be filed or presented in different sequences became routinely available in the twentieth century as means of providing alternative access routes to paper-based resources. It is not only computer technology that allows us to manipulate and combine discrete objects to support new learning.

However, in the pre-digital world, alternative orderings can only be achieved at a cost; the available solutions are invariably time-consuming and their scope is limited. Some of them remain laborious when translated to digital environments; in particular, digital cross-referencing can be provided by "shortcut" or "alias" features that allow links to a single digital object to be inserted in different folders in a directory system, but these features are rarely used in everyday computing because they are difficult to administer and maintain. But digital technology also offers less cumbersome approaches; by forsaking digital folders and replacing them with third-order methods, we can easily assign objects to as many collections as we wish. If we have an appropriate technological framework, little time or labour will be required to group or regroup resources to fit the interests of the user or the needs of the occasion.

Of course, a grouping of objects based on subject content, colour, size, or rhetorical style will rarely constitute a representation of an occurrent. Although such a grouping may be composed of elementary records – in Eastwood's phrase, it may be "made up of documents once individually produced in the course of activity"[71] – we are unlikely to be able to perceive the grouping as an aggregate record in itself. Provenancial interrelationships between its components will almost certainly be absent or confused. Nevertheless, like groupings based on connections to process or activity, a grouping of this kind is a conceptualization that can be realized using third-order methods. When we use these methods to assemble a collection of items that mention our favourite subject or items that are big and green, we are realizing one kind of conceptual

unpaginated (accessed 1 December 2011).

70 Beatrice Webb, *My Apprenticeship* (London, 1926), 426–34; Alison Light, "A Biographical Note on the Text," in Raphael Samuel, *Island Stories* (London, 1998), xix. Samuel's papers may cause us to question archival practices that inhibit further reordering of records. Since Samuel did not accept that work could ever be "finished" (Light, xxi), and since any order in which we find his papers is merely one of myriad possible orders that he may have employed during his life, can archivists justifiably insist on presenting them to users in a single fixed order? I am grateful to Wendy Russell for drawing these references to my attention.

71 Terry Eastwood, "A Contested Realm: The Nature of Archives and the Orientation of Archival Science," in *Currents of Archival Thinking*, ed. T. Eastwood and H. MacNeil (Santa Barbara, CA, 2010), 7.

aggregation, just as physically assembling a collection of items that represent a process realizes another. Moreover, just as processes and activities usually have fuzzy boundaries, notions of (for example) bigness and greenness are also open to a multiplicity of interpretation. Users are likely to have different views on the precise location of borderlines between small and big, or between blue and green, and consequently different opinions on the extent of a conceptual grouping of big green items. We might not choose to label such conceptual groupings or their physical realizations as "series" or as "dossiers" or "files" – once we move away from process-based groupings it is far from clear that these labels remain relevant in the third order – but, like the conceptual series discussed earlier, all these conceptualizations could be realized with a variety of boundaries. Possibly some of them will never be realized.

If, as I have argued elsewhere, many elementary records are boundary objects – entities shared by multiple communities but comprehended or used in varying ways[72] – it is unsurprising that different users favour different groupings and that a single physical hierarchy does not accommodate their diverse requirements. As well as digital systems that support multiple orderings of elementary records, we can also envisage systems that bring records together with other materials. A user might seek to assemble, for example, a selection of records of a manufacturing company alongside images of the company's factories, workers, or products, or of the localities where the company operated, and perhaps might also want to interlink these with published maps and plans or with oral history interviews. It is becoming increasingly apparent that the ability to juxtapose one item with others previously unconnected to it, to form and re-form temporary collections that may sometimes cross the boundaries of provenance, can enhance user experiences and provide scope for innovative modes of research and intellectual discovery. Users who are accustomed to such capabilities in other domains will expect no less from archivists.

Records and the "Database Form"

Much of the literature about the third order focuses on its ability to support information retrieval; writers with a computer science background often assume that this is the main or only issue to be addressed. To take just one example, a paper by Frank McKenna, published in 2009, affirmed that "with computers ... we actually don't need a classification scheme ... we just need a way to index and then search for information."[73] In writings of this kind, concerns about contextualization and evidentiality are conspicuously absent.

72 Yeo, "Concepts of Record (2)," 131–32.

73 Frank McKenna, "Do You Really Need a Taxonomy/Classification Scheme with a Records Management System?" *Records Management Bulletin* 152 (2009): 13.

But this absence does not mean that third-order methods are irrelevant to evidential understandings of recordkeeping. If appropriate frameworks are in place, such methods can enable us to assemble a multiplicity of aggregate records, as well as a wide variety of other aggregations.

The assembly of aggregate records can rarely be left to chance. It normally depends on the availability of formal mechanisms for collating elementary records that are created on separate occasions but perceived as logically related. Without such mechanisms, in the paper world, there will be no files or cabinets whose contents realize a conceptual record of a process, and little possibility that such an aggregate record could ever be brought together. Paper records management needs effective systems for assembling elementary records into aggregations, but it is limited to systems that predefine the shape of the aggregations that can be assembled. In digital environments, systems also need to be robust but need not be restricted to single fixed aggregations; we should be able to assemble different aggregate records as they are called for. Records management has traditionally been concerned with controlling aggregate records that have a stable physical form, but we may now reconceptualize it as equally concerned with ensuring that aggregate records can be constructed when we require them.

For Manovich, the paradigm of the digital realm is the database, which appears as a "new symbolic form" that supplements or replaces previous linear approaches. New media that follow this form "do not have any development ... that would organize their elements into a sequence." A database may present the world as a list of items, but "it refuses to order this list."[74] Manovich's arguments might suggest disparities between databases and the world of documents. The internal structure of individual documents is defined by their creators, and even electronic documents allow only limited scope for variation at the point of retrieval or output, but databases are commonly designed so that presentation of their components can be determined at the time of use rather than the time of creation. Users have traditionally been expected to traverse a document in a preordained sequence, but in the database, as perhaps also in the hypermedia networks of the Internet, creator-defined narratives are replaced by a multiplicity of presentation options. In Manovich's words, the database form becomes "a support for individual users' trajectories." Database users may choose to present items in a way that supplies a particular narrative, but such a narrative has no privileged status; it is "just one method of accessing data among many."[75]

74 Manovich, *The Language of New Media*, 218–19, 225.

75 Ibid., 220, 284. Geoffrey C. Bowker, *Memory Practices in the Sciences* (Cambridge, MA, 2005) also argued that "databases ... do not impose a hegemonic solution" (p. 190), although Bowker recognized that the database is not wholly benign; it is one of the tools that can be used by the state to enumerate and control its citizens (p. 30).

Archivists and records managers now recognize that data and database technology are increasingly used to record transactions and activities.[76] In a database, with appropriate permissions, data can be reused, rearranged, and regrouped largely as we wish. Some possible groupings of data do not appear to serve recordkeeping purposes, but others can constitute different aggregate records. In the words of the ICA's *Principles and Functional Requirements*, "it is ... possible for a single data element to form part of more than one record."[77] Data-centric records may be aggregated in different ways and identified at different levels of aggregation. Moreover, despite the apparent gulf between data-centric and document-centric approaches, in the third order these practices are also applicable to records that take the form of documents; the database model can be applied, not perhaps to the internal structure of an individual document, but to collections of documents that can be assembled in multiple ways to form a variety of aggregate records.

Nevertheless, assembly will be possible only when systems are available to facilitate it. In digital environments, all records (or at least all human-readable records) depend on the processing capabilities needed to assemble their components. In the language of the InterPARES project, "stored records" or "fixed content data" must be processed to output "manifested records" that can be read and interpreted.[78] Since storage is largely random, even a record in the form of a single digital object requires technological tools that can locate its components, assemble them appropriately, and present the result to a user, though hardware and software can be expected to perform these actions seamlessly and failures are likely to be rare until technological obsolescence intervenes. In the case of compound digital objects or records composed of multiple objects, higher levels of processing are needed and the risk of failure may be greater. Records in databases, especially those distributed across several tables in a complex relational database, may need special provision. Data-centric transactional business systems are usually configured to assemble records for display or output, but these may be limited to certain types of records pre-

76 Philip C. Bantin, *Understanding Data and Information Systems for Recordkeeping* (New York, 2008), 103–10; Geoffrey Yeo, "Rising to the Level of a Record? Some Thoughts on Records and Documents," *Records Management Journal* 21, no. 1 (2011): 11–14.

77 International Council on Archives, *Principles and Functional Requirements for Records in Electronic Office Environments, Module 3: Guidelines and Functional Requirements for Records in Business Systems* (2008), http://www.adri.gov.au/products/ICA-M3-BS.pdf, 14 (accessed 1 December 2011). The word "record" is used here in the sense familiar to archivists; readers of *Archivaria* will be aware that database designers often use it with a different meaning.

78 Yvette Hackett, "Methods of Appraisal and Preservation: Domain 3 Task Force Report," in *International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2): Experimental, Interactive and Dynamic Records*, ed. L. Duranti and R. Preston (Padua, 2008), 217; Luciana Duranti, "From Digital Diplomatics to Digital Records Forensics," *Archivaria* 68 (2009): 46.

selected by the system supplier, and the robustness of the configuration over time remains a critical issue.

Ensuring that a range of aggregate records can be brought together on demand from components that are maintained as discrete objects will require innovative approaches and more advanced systems. It will also be important to secure the ability to reconstitute aggregate records that have been assembled or viewed on previous occasions and to ensure that their particular components can be reassembled when necessary. Two decades ago, McKemmish famously asked whether records are "ever actual."[79] My answer would be: records do not have to be *constantly* actual, but they must have enduring potential for actuality. In digital space, the persistence of aggregate records lies in their continuing capacity for realization, which in turn depends on the continuance of structures that allow appropriate sets of elementary records to be brought together and presented to users. We need not privilege the assembly of records for evidential purposes over other groupings that accommodate different frames of reference, but we must be sure that requirements for records can be met. The ability to realize aggregate records must be a crucial part of our third-order systems.

Matters of Context

Discussion of users' requirements for records must also address contextualization. Michael Roper has recounted how, at the UK Public Record Office (PRO) in the early 1900s, the principle that records should be presented in accordance with their provenance or context overcame the previously prevailing view that only their information content was important. In the years immediately before the First World War, members of the PRO's staff, perhaps influenced by the Dutch *Manual* published in 1898, began to argue that researchers would "wish to know ... under what conditions [government records] were written, received, and preserved in official custody" and to study them "in their proper setting as products of an administrative machine, and not merely as evidence of isolated facts."[80] A century later, some commentators have begun to express anxiety that the flexibility and fluidity of the third order may presage a return to a world of isolated facts used out of context. Francis Blouin and William

79 McKemmish, "Are Records Ever Actual?" 187.

80 Michael Roper, "The Development of the Principles of Provenance and Respect for Original Order in the Public Record Office," in *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, ed. B.L. Craig (Ottawa, 1992), 134–53. The quotations are from Hubert Hall, *Studies in English Official Historical Documents* (Cambridge, 1908), 83; *Oxford Dictionary of National Biography*, s.v. "Crump, Charles George," cited by Margaret Procter, "Life Before Jenkinson: The Development of British Archival Theory and Thought at the Turn of the Twentieth Century," *Archives* 119 (2008): 151.

Rosenberg, for example, have argued that "when content is ... stored in digital form," materials may be "structured apart from the contexts of their creation" and thus deprived "of important kinds of historical meaning." Robert Cole and Chris Hackett have also apprehended that the ability to "reorder ... information in a manner more amenable to ... specific research interests" may be "at odds with ... the need to examine sources in their totality and to understand them in the broader context of their creation."[81]

At the opposite pole, some twenty-first-century archivists have asserted that contextual wrapping is often redundant. Bill Stockting has noted that many users show little concern for contextual detail; Bradley Westbrook claimed that knowledge of wider contexts is unimportant for users who "want to mine ... only the pieces of information that satisfy their purposes."[82] Others, more radically, might want to argue that notions of definable historical context are becoming outmoded in a postmodern culture and that users should be encouraged to conduct their own explorations on their own terms. Just as, for example, YouTube presents clips from old television shows without demonstrating much concern for the dates or other contexts of their original production, the Web 2.0 world in which we live is said to be a world in which uses of the past are increasingly linked to the fleeting interests of the present.[83] In these modes of thinking, memory and identity (as perceived by those alive today) supersede history (of the world as it was, or as it might have been); language or text is proclaimed to be self-sufficient, and the contexts that matter are those of the reader, not the originator.

Not every user of archives shares these postmodernist concerns. Even researchers wishing to read archives "against the grain" may feel the need for an understanding of the "grain" that they are trying to counteract, in terms of the contextual milieux in which documents were created and initially captured. Decontextualized access seems acceptable in some circumstances, but not in all. To many users, entries in a register of baptisms may appear self-explanatory, but some users – such as those unfamiliar with Christian baptism practices – may need to seek elucidation. Many other records are likely to be far less transparent. Ronald's scribbled note "OK, let's do it" (or a digital

81 Francis X. Blouin and William G. Rosenberg, *Processing the Past: Contesting Authority in History and the Archives* (New York, 2011), 203; Robert Cole and Chris Hackett, "Full-text Repositories, Granularity, and the Concept of 'Source' in the Digital Environment," in *Better Off Forgetting? Essays on Archives, Public Policy, and Collective Memory*, ed. C. Avery and M. Holmlund (Toronto, 2010), 113, 115.

82 Bill Stockting, "Time to Settle Down? EAD Encoding Principles in the Access to Archives Programme and the Research Libraries Group's Best Practice Guidelines," *Journal of Archival Organization* 2, no. 3 (2004): 16; Bradley D. Westbrook, "Prospecting Virtual Collections," *Journal of Archival Organization* 1, no. 1 (2002): 79.

83 Lynn Spigel, "Housing Television: Architectures of the Archive," in *Media History and the Archive*, ed. C. Robertson (Abingdon, England, 2011), 68–69.

image of the note) may suffice without contextualization for a user seeking a sample of Ronald's handwriting, but most other users may legitimately want to know whether the proposal to which he consented was for ordering lunch or invading the Evil Empire. Arguably, the handwriting student might also seek to identify a context for the note, as reassurance that the writing she is examining is indeed Ronald's.

Similar variations can be expected if users are empowered to construct their own collections in a third-order environment. When they set out to build a collection, some users may feel they need little or no knowledge of previous contexts, but others are likely to seek a degree of contextual understanding both in identifying items for inclusion in the collection and in subsequent analysis. Moreover, since users have different aims and objectives, the contexts of interest can be expected to vary from one user, or one user community, to another. Many users – not least those wanting to realize a conceptual record series or other aggregate record for evidential purposes – will attach particular significance to the social or administrative contexts in which each elementary record originated and the activity that it represents; some, however, might focus their concern on its technological context – the interactions of hardware and software that underwrote its production – or on the contexts of its subsequent career of storage, movement, retrieval, and exploitation.

In a world of reusable data and third-order systems, the risk of loss of context should not be underestimated. If we are to allow or encourage users to create their own collections and construct their own hierarchies, we also need to find ways of presenting larger or previous contexts and of enabling users to contextualize each item in their collections. Equally, we should recognize that traditional finding aids have sometimes been perceived as providing users with more contextual detail than they want;[84] new approaches should avoid imposing excessive contextual information on users who do not require it.

"Original" Orderings

Archivists have often believed that physical and conceptual orderings have a natural correspondence and that documentary relationships are reproducible in linear sequences and classification schemes. These beliefs underlie many of the formal definitions that assert or assume coincidence of conceptual and physical aggregations. They also underlie the importance often attributed to preserving pre-existing orders and physical arrangements, as a means of protecting and presenting logical contexts and associations. Blouin

84 Anne J. Gilliland-Swetland, "Popularizing the Finding Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups," in *Encoded Archival Description on the Internet*, ed. D.V. Pitti and W.M. Duff (New York, 2001), 207–8; Westbrook, "Prospecting Virtual Collections," 79.

and Rosenberg, for example, contended that "files and fonds need to be kept intact" if we are to avoid the dangers of decontextualization.[85] I wish to argue that preserving information about previously imposed orders and physical arrangements can be important even when the methods of the digital world are available to us, but that logical contexts can be better protected in other ways.

The principle of original order rests on an assumption that "each group of records is ... an articulate whole, the meaning of which is lost if its natural arrangement is disturbed."[86] Increasingly, however, archivists are questioning allusions to "natural arrangement" and notions that groups of records might have a single definable "meaning." If we see meanings as multiple and arrangements as artificially imposed, we are unlikely to be content with approaches that seek to capture contextual meaning by identifying and preserving immutable physical series and files. Contextual relationships are too complex to be encompassed within the hierarchical order of a classification scheme. In their study of architectural practice, Kjeld Schmidt and Ina Wagner demonstrated some of the intricate and diverse relationships between records, actors, and activities. Schmidt and Wagner argued that architects, their clients, and other parties interact and coordinate their activities through collections of "digital documents as well as paper and other tangible artifacts," but their behaviour involves a "mixture of concurrent, sequential and reciprocal action" more elaborate than any single ordering of records could manifest.[87] Many physical arrangements (whether "original" or not) undoubtedly indicate how creators, records managers, or other custodians felt they could best try to represent documentary relationships within the constraints of a hierarchical filing system, but such relationships are not always linear. Any attempt to use physical containers to construct representations of complex temporal phenomena is almost certain to be compromised.

Where does this leave archivists' concerns for *original* order? If individuals (like Sidney and Beatrice Webb) or organizations rely on "ordering on demand" for their recordkeeping, it is clearly appropriate to carry over this original ordering method into archival settings. However, most organizations and individuals still employ hierarchical approaches to organize their records; the widespread use of files in hard-copy storage is closely paralleled by the use of folders in directory systems supplied by computer vendors. Evidence that

85 Blouin and Rosenberg, *Processing the Past*, 205. Cf. Alistair Tough and Michael Moss, "Metadata, Controlled Vocabulary and Directories: Electronic Document Management and Standards for Records Management," *Records Management Journal* 13, no. 1 (2003): 24–31; Tough and Moss argued for the continued use of classification schemes ("file plans") alongside other approaches that treat digital documents as discrete units.

86 Charles Johnson, *The Care of Documents and Management of Archives* (London, 1919), 11.

87 Kjeld Schmidt and Ina Wagner, "Ordering Systems: Coordinative Practices and Artifacts in Architectural Design and Planning," *Computer Supported Cooperative Work* 13 (2004): 349–408.

functional classification schemes imposed by records managers are unpopular is counterbalanced by a considerable body of evidence that computer users like folders when they are free to construct and arrange them in their own way.[88] Arguably, their arrangement choices may not always match archival "best practice" or the needs of their employing organizations. Nevertheless, from a longer-term perspective, preserving information about "original" order – or other orders imposed subsequently – can show how people organized their archives at a certain point in time and thus tell us something about the priorities and perceptions of the people concerned. Moreover, although we may doubt whether the ordering of even the most meticulously organized collection is sufficient to capture the intricate richness of context, it may still provide contextual clues that are unavailable elsewhere and thus enable us to make inferences about circumstances of creation or use that are implicit but not formally documented. Even when a past order reflects little beyond happenstance or the whim of a filing clerk, it can supply evidence of the shape in which the archives appeared to those who viewed them at some moment in the past; it may also allow us some understanding of how they could have been used and interpreted at that time.

For all these reasons, it is important to retain knowledge of how records have been organized in the past. At a basic level, it is not difficult to combine linear descriptions reflecting a past or present physical arrangement with interactive features that offer the openness and flexibility of the third order. For example, the website of the Minnesota Historical Society provides two different approaches to the papers of Jerome Hill; metadata about these papers can be organized "on demand" on the basis of theme or time period, but users can also view a conventional finding aid that depicts a single "fixed" arrangement apparently imposed on the papers by their previous custodians.[89] In the UK, the innovative "Archive as Event" website sets out to present the papers of the artist John Latham (or, more precisely, digital images of them) in accordance with Latham's cosmological theories of time and event structures. When fully developed, the site will offer users a variety of third-order approaches, including a random slide show, a time base, and search options to form groupings

88 Ofer Bergman et al., "Improved Search Engines and Navigation Preference in Personal Information Management," *ACM Transactions on Information Systems* 26, no. 4 (2008): 20; Philip Jones, "The Role of Virtual Folders in Developing an Electronic Document and Records Management System: Meeting User and Records Management Needs," *Records Management Journal* 18, no. 1 (2008): 55; William Jones et al., "Don't Take My Folders Away! Organizing Personal Information to Get Things Done," in *CHI '05: Extended Abstracts on Human Factors in Computing Systems* (New York, 2005), 1505–8.

89 See http://collections.mnhs.org/jeromehill; http://www.mnhs.org/library/findaids/00565.xml (accessed 1 December 2011).

"on demand," but it also offers a more stable view of the papers that lists them in the order in which the archivists found them after Latham's death.[90]

A list of this kind need do no more than represent the physical ordering of a collection; it need make no attempt at explicating logical relationships, which can be better handled in other ways. However, even as a means of depicting physical arrangements, the approach used with the Hill and Latham archives has its limitations. Each of the websites presents one supposedly "original" order, but a collection frequently has multiple and varying orders imposed at different times in its life. We have seen that initial orderings to support task prioritization often precede attempts at imposing a long-term arrangement. Even when putative long-term arrangements are in place, records creators sometimes change the ordering in an attempt to introduce a better system; later custodians may do the same. It often seems futile to try to distinguish orders that are "original" from those that are not. No ordering is likely to be perfect, but all are potentially significant. Ideally, as Peter Horsman has suggested, we should uncover each consecutive original order, not just the first (or the last).[91] Hypothetically, we could do this by providing separate listings of every order that has been applied to a collection at different times, but this would incur considerable redundancy of descriptive information. We need to develop a more efficient approach that makes appropriate use of the tools available in the digital age.

An Approach to Third-Order Contextualization

As Terry Cook foresaw twenty years ago, assumptions that the arrangement of archives reveals contextual provenance have proved increasingly inadequate.[92] Although understanding of context can often be crucial, information about past physical orderings alone will rarely provide more than a glimpse of it. Since logical contexts transcend physical arrangements, we must distinguish information about such arrangements from knowledge of the wider contexts in which archival resources have been created and used.

90 See http://www.ligatus.org.uk/jla; http://www.ligatus.org.uk/aae (accessed 1 December 2011). Further information about this site was presented by Athanasios Velios at the Archives and Society seminar, University of London, 1 February 2011. At the time of writing, the site was still under development.

91 Peter Horsman, "Taming the Elephant: An Orthodox Approach to the Principle of Provenance," in *The Principle of Provenance: Report from the First Stockholm Conference on Archival Theory and the Principle of Provenance,* ed. K. Abukhanfusa and J. Sydbeck (Stockholm, 1994), 58.

92 Terry Cook, "The Concept of the Archival Fonds in the Post-Custodial Era," *Archivaria* 35 (1993): 26.

In the world of the third order, where "multiview is the way people work,"[93] we need to be able to construct different aggregate records as well as a variety of "non-record" aggregations; we also need the abilities to preserve information about past orderings and to provide richer evidential contexts wherever they are wanted. These are demanding requirements, and robust, scalable, and user-friendly systems will be required if we are to meet them all. A recipe for success will require a number of ingredients; third-order methods will need to be combined with other emerging technologies.

Granularity

I suggest that the first ingredient is an emphasis on granularity. In a seminal paper published in 1996, Bearman argued for the importance of what he called "item level control." He affirmed that archivists manage paper records collectively only because "it is too expensive to manage them individually" and that in digital environments it is more efficient "to control and describe records at the item level from the moment of their creation."[94] Bearman said little about the wider discourses that may have shaped his thinking, but it seems likely that he was influenced, directly or indirectly, by visionaries such as Nelson who were advocating a shift away from "top-down" or hierarchical models in the theory and practice of computer science.

In the archival domain, the first steps in this direction had already been taken in the 1960s in Australia, where the "series system" had displaced the record group as the principal level of control. In its stead, Peter Scott had identified the series as the lynchpin of description and record organization.[95] To Scott, and others who followed him, the series seemed unproblematic as a granular control unit. Series were thought to have a "fundamental organic nature";[96] they might perhaps be split between current and archival storage, but in other respects physical and conceptual series were assumed to coincide. In Scott's system, the series did not form part of a hierarchical record group, but items (and files and folders) within a series were assumed to have stable contexts and part-whole relationships that would allow them to be managed

93 Andries Van Dam, "Hypertext '87 Keynote Address," *Communications of the ACM* 31, no. 7 (1988): 895.

94 David Bearman, "Item Level Control and Electronic Recordkeeping," *Archives and Museum Informatics* 10, no. 3 (1996): 195–245; the quotations are from p. 201.

95 Peter J. Scott, "The Record Group Concept: A Case for Abandonment," *American Archivist* 29, no. 4 (1966): 493–504.

96 Peter J. Scott, "Introduction," in *The Arrangement and Description of Archives amid Administrative and Technological Change*, ed. A. Cunningham (Brisbane, 2010), 28. Cf. Max J. Evans, "Authority Control: An Alternative to the Record Group Concept," *American Archivist* 49, no. 3 (1986): 252; Cook, "The Concept of the Archival Fonds: Theory, Description, and Provenance," 70.

collectively. In practice, however, components of physical series can be fluid; creators or custodians may rearrange items or folders within a physical series, remove them from the series, or relocate them from one series to another. Conceptual series, as we have seen, are open to shifting interpretation; an elementary record may be perceived to have membership of a variety of conceptual series, depending on the viewpoint of the observer. The series is a much less stable entity than Scott assumed, and Australian work in the new millennium no longer gives it a privileged status.[97]

In 2010, I intimated that levels of "control" might be contingent on context and circumstances.[98] New approaches require a finer level of granularity than Scott envisaged, but even the notion of control at item level may not always be unambiguous; for example, there can sometimes be scope for debate about what constitutes an "item" in a complex business system or web resource. Nevertheless, as Bearman's 1996 paper affirmed, the item is the paradigmatic unit of control in the third order. Each item can be assembled with others in different ways; each has its own provenance, its own complex network of relationships with actors, actions and events, and potentially its own history of aggregation and use. In the nineteenth century, item-level control was the norm in many registry systems until the growing volume of paperwork rendered it impracticable. In the digital era, it seems both possible and necessary to return to some form of item-level management if we wish to move away from the constraints imposed by rigid hierarchical schemas.[99] As yet, archivists have generally been reluctant to embrace such moves, but for more than a decade shifts toward item-level control have underpinned research projects in computer science, such as the Placeless Documents project at the Xerox Palo Alto Research Center in 1999/2000, which aimed to develop a working system to support multiple categorization in document management, and the Fenfire project in Finland between 2003 and 2008, which explicitly built on Nelson's work with the aim of developing an "item-centred computing environment."[100]

97 Sue McKemmish, Barbara Reed, and Michael Piggott, "The Archives," in *Archives: Recordkeeping in Society*, ed. S. McKemmish et al. (Wagga Wagga, NSW, Australia, 2005), 170–71; Hurley, "Relationships in Records," sections 5.04–5.05.

98 Geoffrey Yeo, "Debates about Description," in *Currents of Archival Thinking*, ed. T. Eastwood and H. MacNeil (Santa Barbara, CA, 2010), 106.

99 For a further call for records managers and archivists to adopt item-level management, see Greg Bak, "Continuous Classification: Capturing Dynamic Relationships among Information Resources," *Archival Science*, forthcoming.

100 Paul Dourish et al., "Presto: An Experimental Architecture for Fluid Interactive Document Spaces," *ACM Transactions on Human–Computer Interaction* 6, no. 2 (1999): 133–61; Paul Dourish et al., "Extending Document Management Systems with User-Specific Active Properties," *ACM Transactions on Information Systems* 18, no. 2 (2000): 140–70; Benja Fallenstein and Tuomas J. Lukka, "Hyperstructure: Computers Built Around Things That You Care About" (2004), http://fenfire.org/manuscripts/2004/hyperstructure/ (accessed 1 December 2011).

Practical implementations of item-level control in information technology are now becoming widespread. Over the past five years or so, granular systems that do not impose single modes of ordering have steadily gained popularity. These developments extend far beyond the social spaces of Flickr and YouTube. Alongside a growing number of research papers proposing the abandonment of hierarchical file systems,[101] non-hierarchical models and multiple categorization approaches have been extensively implemented in, for example, Google's Gmail and Microsoft's SharePoint application. Desktop and enterprise search systems based on the same paradigm have been widely promoted. Mainstream computing can increasingly be expected to supply the technical infrastructures needed to support management at granular levels.

Relational Modelling

The second suggested ingredient is the use of relational models and systems. In 2003, Shepherd and I proposed that records managers might use authority files as a means of providing appropriate context without predefined hierarchies; more recently, Sabine Mas and her colleagues have suggested using faceted classification methods developed in library science, to provide a range of alternative views of electronic records.[102] But approaches derived from relational and object-oriented modelling seem preferable to either of these solutions; they are more adaptable, more responsive to diachronic change, and better able to encode complexity and reduce the risk of decontextualization in the third order.

As I noted in "The Conceptual Fonds and the Physical Collection," and as recent Australian work has amply demonstrated, relational models offer powerful means of documenting our perceptions of entities and their contextual relationships.[103] Perceived logical relationships between one record and another, and relationships of records to activities and to agents such as creators, collec-

101 See, for example, Margo Seltzer and Nicholas Murphy, "Hierarchical File Systems Are Dead" (2009), http://www.eecs.harvard.edu/~margo/papers/hotos09/paper.pdf (accessed 1 December 2011).

102 Shepherd and Yeo, *Managing Records*, 95–98; Sabine Mas, Dominique Maurel, and Inge Alberts, "Applying Faceted Classification to the Personal Organization of Electronic Records: Insights into the User Experience," *Archivaria* 72 (2011): 29–59.

103 Yeo, "The Conceptual Fonds," 71–78; Hurley, "Documenting Archives." Among existing models, probably the best known in the archival field is the Australian *SPIRT* model (Sue McKemmish et al., "Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema," *Archivaria* 48 (1999): 3–43), which models agents, records, business, and mandates, and seeks to depict relationships among them. The faceted classification approach advocated by Mas et al. also sets out to "link ... information objects to ... business processes" (Mas, Maurel, and Alberts, 38), but in practice its descriptions of "activities" and "projects" merely appear to form part of the metadata describing the record (ibid., 41–46). Relational approaches such as *SPIRT* seek a clearer separation of record and context.

tors, custodians, and users can all be modelled, as can relationships with other entities such as places and dates; viewing these through the lens of a relational model may offer levels of understanding that we cannot attain if objects are viewed in isolation. In these respects, relational models fulfill the context-documenting role traditionally claimed for classification schemes, but they fulfill it more effectively because they are unconstrained by fixed hierarchies and can support documentation of contexts that evolve over time. Physical as well as logical relationships can be modelled, as can diachronic changes to physical relationships. If our model is sufficiently detailed, we could use a relational system to generate descriptions of collection membership, details of an object's juxtapositions to other objects in a collection or a succession of collections, or sequential lists of container contents and orders at different times. Even the endlessly shuffled and never finalized papers of Raphael Samuel and the Webbs have histories of physical ordering as well as contexts of origin and use; all could potentially be documented using relational systems, which offer more expressive power than any other approaches currently available to us.

Most existing item-centred implementations are lightweight approaches that rely on tagging or search tools; few are likely to support the depth of contextualization that archivists may require. However, much recent computer science research, especially in the field of Semantic Web development, has focused on the possibility of combining item-level organization with relational methods. Both in Europe and in the US, researchers have proposed replacing hierarchical computer file systems with "semantic" file systems based on "context awareness" or "provenance relationships."[104] Another European research project, NEPOMUK, has sought to use Semantic Web technologies to relate digital resources to agents, tasks, and other contextual entities in order to develop a "social semantic desktop."[105] Computer scientists have also begun to seek ways of using technology to address the resource challenges that are implicit in the creation of complex relational models; for example, an Austrian project has investigated the modelling of computer users' task contexts in combination with automated task detection techniques.[106] As might be

104 Alexandros Karypidis and Spyros Lalis, "Automated Context Aggregation and File Annotation for PAN-based Computing," *Personal and Ubiquitous Computing* 11, no. 1 (2007): 33–44; Andrew W. Leung et al., "Copernicus: A Scalable, High-Performance Semantic File System" (2009), http://ssrc.cse.ucsc.edu/Papers/ssrctr-09-06.pdf (accessed 1 December 2011).

105 Gunnar A. Grimnes et al., "The Personal Knowledge Workbench of the NEPOMUK Social Semantic Desktop," *Lecture Notes in Computer Science* 5554 (2009): 836–40; Uwe V. Riss et al., "Knowledge Work Support by Semantic Task Management," *Computers in Industry* 61 (2010): 798–805. For an earlier project along similar lines, see Thomas Moran et al., "Unified Activity Management: Supporting People in E-business," *Communications of the ACM* 48, no. 12 (2005): 67–70.

106 Andreas S. Rath et al., "UICO: An Ontology-based User Interaction Context Model for Automatic Task Detection on the Computer Desktop," in *Proceedings of the 1st Workshop*

expected, researchers in all these projects have viewed context and provenance relationships from a computing rather than an archival science perspective, typically emphasizing their role as data management or access mechanisms. Nevertheless, it seems reasonable to foresee a time when research of this kind will provide archivists with valuable new tools to support documentation and understanding of context.

Some caution, however, is required. Especially where logical relationships are concerned, we would do well to question how far the construction of a relational model is an act of imagination or interpretation (a way of making sense of a domain on the basis of particular assumptions about the way the world works) and how far it can be expected to reflect realities that exist or once existed in the world. We must accept that both logical and physical relationships – particularly past relationships that have left only nebulous traces for us to examine – may be matters for inference or conjecture rather than secure knowledge that can be modelled with confidence. Ultimately, the complexities of human experience lie beyond the capacities of any model we can hope to design. To some commentators, the Semantic Web and the models that underpin it are "open, free, ... visionary, idealistic, experimental, ... part of Web 2.0, 3.0 or even some futuristic variant," but others have claimed that such models are authoritarian because they privilege particular world views or seek to impose unwarranted levels of certainty.[107] To employ relational models successfully, we will need to address the tensions between their enabling functionality and the constraints they impose.

Relationships among items, and between items and their contexts, lie partly in the history of the items themselves and partly in the perceptions of those who observe them (perceptions that may be shared by the wider communities to which the observers belong). Excessive rigidity could perhaps be avoided by using a combination of automated and participatory techniques to build relational systems, giving opportunities for records creators and users to contribute to their development and to suggest different perspectives on entities and relationships. If, as a recent commentary has advocated, we "engage communities not just to author content ... but also [to] serve as ... information

on Context, Information and Ontologies (New York, 2009), unpaginated. For further references to tools designed to automate the detection and capture of contexts, see Yeo, "The Conceptual Fonds," n90. See also Carlos Jensen et al., "The Life and Times of Files and Information: A Study of Desktop Provenance," in *CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems* (New York, 2010), 767–76.

107 Liam Magee, "Contemporary Dilemmas: Tables versus Webs," in *Towards a Semantic Web*, ed. B. Cope, M. Kalantzis, and L. Magee (Oxford, 2011), 226; Clay Shirky, "The Semantic Web, Syllogism, and Worldview" (2003), http://www.shirky.com/writings/semantic_syllogism.html (accessed 1 December 2011).

architects," we may be able to develop flexible structures whose shape is emergent, negotiable, and responsive to change.[108]

If such systems are to be effective, each entity will need rich metadata, to support its identification and retrieval and to supply further lines of defence against loss of context. It may be helpful to distinguish between metadata that are usually relatively uncontroversial, such as geographical location coordinates and dates of access, and metadata that are more discursive and open to contestation, such as content descriptions and descriptions of agents and activities. We can expect that the former will increasingly be captured automatically by software. Provision of the latter is more problematic. The labour-intensiveness of manual capture of item-level metadata has long been known; indeed, promoters of the "More Product, Less Process" approach have asserted that archivists need to "avoid ... descriptive work on an item level" to achieve increased productivity.[109] If creators and users can contribute metadata, the burden of capture is shared and – equally importantly – representation of diverse viewpoints may be enhanced, but attempts to persuade users of electronic records management systems to supply contextual metadata at the point of creation have often met considerable resistance. Even free-form tagging, which normally requires less effort than more formal methods of metadata capture, may not be adopted by a sufficient number of users. To achieve the volumes of metadata that may be needed, manual capture will almost certainly have to be supplemented by artificial intelligence tools that automatically analyze the form, content, and contexts of objects and populate descriptive systems accordingly. We cannot expect the capabilities of these tools to match human understanding, but in the future we can surely expect them to expedite work that would otherwise be impossibly time-consuming.

In "The Conceptual Fonds and the Physical Collection" I argued that, because fonds are conceptual groupings, they should not be represented as fixed entities in relational systems but should be seen as potential system outputs. This argument can now be extended to suggest that other aggregations whose boundaries are a matter of interpretation, such as processes and conceptual series, should likewise be considered and constructed as system outputs. Elementary activities, individual and organizational agents, item-level records, physical containers, and collections are less subjective entities whose inclusion in a formal model presents less difficulty. Relational systems also

108 Ramesh Srinivasan, Alberto Pepe, and Marko A. Rodriguez, "A Clustering-based Semi-automated Technique to Build Cultural Ontologies," *Journal of the American Society for Information Science and Technology* 60, no. 3 (2009): 611. See also Ramesh Srinivasan and Jeffrey Huang, "Fluid Ontologies for Digital Museums," *International Journal on Digital Libraries* 5, no. 3 (2005): 193–204.

109 Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (2005): 253.

support documentation of known logical and physical associations between these entities (for example: Letter 4 was written by Person P and received by Person R; Letter 4 is (or was) in Box A; Box A was part of Physical Series B). Information about such relationships can be expected to provide some immediate contexts in third-order environments and can often be captured with a degree of confidence, although relational systems must also be able to take account of past states or events that were once knowable but are now uncertain.[110] But factual certainty is impossible and irrelevant where membership of conceptual aggregations is concerned; attempts to specify their membership will always be arbitrary and contested. Instead, if entities and relationships are documented at a granular level, such aggregations can be assembled dynamically. Parameters can be set and components identified to match differing perceptions; conceptualizations can be realized as users wish.

System Interfaces and Functionalities

Thirdly, we will need appropriate interfaces and, behind them, the computing power to support the inputs and outputs of the system. Unconstrained by paper paradigms, systems and interfaces should enable archival resources to be presented in many different ways, reflecting their various "original" orders, different interpretations of context, and other orders newly desired by users in the course of research and experimentation. In addition, as Randall Jimerson has observed, "in the 2.0 world ... institutions [and individuals] will continue to need ... records for legal, evidential, accountability, administrative, and documentary purposes."[111] Technical ingredients must include the ability to realize a range of aggregate records and present them in ways that can meet institutional and individual needs. Australian experience with the relatively limited complexities of the series system suggests that the available interfaces have not been found user-friendly; considerable further research will be required to develop and test effective interfaces if users are not to be overwhelmed by the more elaborate structures of relational systems operated at item level.[112]

110 For example: Letter 4 was probably written by Person P; it may have been in Box A between Date 1 and Date 2, but we can no longer be sure.

111 Randall C. Jimerson, "Archives 101 in a 2.0 World: The Continuing Need for Parallel Systems," in *A Different Kind of Web*, ed. K. Theimer (Chicago, 2011), 325.

112 Andrea Rosenbusch, "Are Our Users Being Served? A Report on Online Archival Databases," *Archives and Manuscripts* 29, no. 1 (2001): 50. Relevant investigations have already been undertaken by computer scientists aware that user acceptance of Semantic Web technologies will require appropriate interfaces to complex interlinked representations. For the development of non-hierarchical interfaces see, for example, Dennis Quan et al., "User Interfaces for Supporting Multiple Categorization," in *Human–Computer Interaction: INTERACT '03 International Conference on Human–Computer Interaction*, ed. M. Rauterberg et al. (Amsterdam, 2003), 228–35, and (in a cloud-computing context) Lucia Terrenghi et al., "CloudRoom: A Conceptual Model for Managing Data in Space

Some cultural critics have envisaged a kind of techno-utopia where archival materials can be retrieved and examined without verbal labels or semantic indexing.[113] Their dream is of the computer as a transparent medium, providing unmediated access to "texts" unsullied by curatorial intervention or other forces of hegemony. In practice, however, such aspirations cannot be fulfilled. The size and shape of screens, the look and feel of web browsers, and the functionality of interfaces are all likely to be determined, if not by archivists, then by software designers and the computer industry. Despite the apparent breadth of democratic choice offered by, for example, the Hill and Latham websites, limits are in fact imposed; the materials that can be retrieved, the range of available access routes, and the presentation of the outputs selected by the user are all defined by the system.[114] Ideally, users as well as archivists and technologists might contribute to decisions about modes of access and presentation, but it seems inevitable that some degree of structure must be applied; even in the third order, we cannot achieve unlimited fluidity of design.

Despite their deficiencies, traditional linear finding aids offer what Ruth Frendo has called "transparency of organisation."[115] They support browsing, allow users to obtain an overview of available resources, and usually supply at least a modicum of contextual information. In third-order environments, where linear access does not operate, users still need the ability to assess the scope and possible significance of archival materials; they also need orientational tools to assist in planning research and assembling appropriate collec-

and Time," in *CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems* (New York, 2010), 3277–82. David R. Karger, "Haystack: Per-User Information Environments Based on Semistructured Data," in *Beyond the Desktop Metaphor*, ed. V. Kaptelinin and M. Czerwinski (Cambridge, MA, 2007), describes a system (intended for personal computers) that aims to offer users the richness of a relational model but with a flexible interface that does not exhibit the technical aspects of the underlying database.

113 Doireann Wallace, "Words as Keys to the Image Bank," in *Revisualizing Visual Culture*, ed. C. Bailey and H. Gardiner (Farnham, England, 2010), 83, citing Wolfgang Ernst, "Dis/continuities: Does the Archive Become Metaphorical in Multi-media Space?" in *New Media, Old Media*, ed. W.H.K. Chun and T. Keenan (New York, 2006).

114 For accounts of decisions about the form of the Latham website, see Athanasios Velios and Simon Gould, "Applying Artists' Methodologies to Archiving: A Case Study of John Latham's Archive," in *Archiving 2008: Final Program and Proceedings* (Springfield, VA, 2008), 125–29; Athanasios Velios, "Creative Archiving: A Case Study from the John Latham Archive," *Journal of the Society of Archivists* 32, no. 2 (2011): 255–71. Some writers have suggested parallels between the participatory empowerment offered by information technology and supposed trends toward democratization, decentralization, and flattening of hierarchies in organizations and the wider society, but others, following Manuel Castells, *The Rise of the Network Society*, 2nd ed. (Oxford, 2000), have argued that "network societies" can have their own power relations and are not necessarily progressive. See, for example, Darin Barney, *The Network Society* (Cambridge, 2004), especially 27–32; Geert Lovink, *The Principle of Notworking* (Amsterdam, 2005), 17–21.

115 Frendo, "Disembodied Information," 163.

tions to support it. Search engines alone will not provide these affordances. Visualization techniques, recently explored by Mitchell Whitelaw and others, offer intriguing possibilities;[116] when large quantities of complex information are to be assimilated, visual approaches often allow easier comprehension than textual modes of presentation, and can also provide effective ways of displaying and navigating multi-dimensional relationships that cannot easily be represented in linear text. Not every user will always want to examine every aspect of a relational model, but system interfaces must be able to present such models, or relevant parts of them, to those who need them. Users should be able to decide how much of the model and how much associated metadata they want to see, both when building a collection dynamically and when seeking additional contextual information about items in a collection they have already built. When search engines are employed, access to such models and their metadata can be expected to enhance understanding of search results that might otherwise be decontextualized. Interfaces should be capable of offering different views of a model, perhaps including views created interactively by users themselves.

In considering the systems we might want, this paper can do no more than set out some preliminary ideas for future development. Practical implementation will call for advanced technology that is scalable but not overly deterministic. Further work on relational modelling can be undertaken now, but its benefits will be achieved only when appropriate technical tools become widely available. Many of these tools are still in their infancy at present and will almost certainly develop in ways we cannot now foresee. For both archivists and users, a shift away from hierarchical modes of access will require significant cultural changes, whose implications have not yet been fully explored. A long journey lies ahead.

116 Mitchell Whitelaw, "Visualising Archival Collections: The Visible Archive Project," *Archives and Manuscripts* 37, no. 2 (2009): 22–40. See also Robert B. Allen, "Using Information Visualization to Support Access to Archival Records," *Journal of Archival Organization* 3, no. 1 (2005): 37–49; Jeanne Kramer-Smith, Morimichi Nishigaki, and Tim Anglade, "ArchivesZ: Visualizing Archival Collections" (c. 2008), http://archivesz.com/ArchivesZ.pdf (accessed 1 December 2011); Weijia Xu, Maria Esteva, and Suyog Dott Jain, "Visualizing Personal Digital Collections," in *JCDL '10: Proceedings of the 10th Annual Joint Conference on Digital Libraries* (New York, 2010), 169–72. At the time of writing, visualization tools for relationship mapping were being developed by, among others, the Crowded Page project, http://www.crowdedpage.org, the RoSE project, http://rose.english.ucsb.edu, and the Yaddo Circles project, http://vimeo.com/36929545 (all accessed 18 May 2012).

Conclusion

Archivists have often wanted to believe that logical relationships and physical orderings of archives naturally coincide. This paper and its predecessor in *Archivaria* 73 have set out to demonstrate that this cannot be so. Logical relationships are likely to be more complex than any single physical aggregation could possibly imitate. Within such an aggregation, neither physical juxtaposition nor traditional classification schemes that mimic its constraints can hope to replicate logical associations, which in any event are likely to extend beyond the borders of the aggregation. We must also recognize that the nature and extent of logical relationships can be open to dispute; a relationship between two activities (and hence between the records that represent those activities) may be apparent to one observer but not to another. Relational models that allow for uncertainty and diversity of interpretation will come closer to meeting twenty-first-century needs than reliance on the sequential structure of physical aggregations or hierarchical classification schemes.

However, while relational models can provide us with more powerful ways of documenting our understandings of context, this does not mean that information about hierarchical or sequential orderings imposed in the past is redundant. Such information can have its own significance, but we need no longer struggle to preserve past orders by means of shelving plans or linear description; relational approaches can also provide an effective basis for documenting previous arrangements and their evolution over time. Hitherto, archivists who acknowledged tensions between conceptual and physical groupings have often faced a difficult choice between maintaining representations of business activity and protecting existing or former physical arrangements. Now we can begin to design integrated environments that support both these objectives and also facilitate new readings and creative analysis of archival materials.

In archival literature, relational approaches have often been associated with notions of a postcustodial world in which physical collections are seen as outmoded, irrelevant, or non-existent. In the 1990s, Australian archivists Greg O'Shea and David Roberts characterized postcustodialism as an approach that "transcends a narrow ... collecting focus,"[117] and Australian writings still often present the work of Scott and his followers as a "virtual" solution that eliminates or supersedes physicality.[118] These ideas resonate with many writings in the 1990s by cultural critics who asserted or assumed that digital objects or

117 Greg O'Shea and David Roberts, "Living in a Digital World: Recognising the Electronic and Post-custodial Realities," *Archives and Manuscripts* 24, no. 2 (1996): 293.

118 See, for example, Barbara Reed, "The Australian Context Relationship (CRS or Series) System: An Appreciation," in *The Arrangement and Description of Archives amid Administrative and Technological Change*, ed. A. Cunningham (Brisbane, 2010), 352–54.

digital texts are immaterial.[119] The suggestion that digital objects have no physical existence is what textual scholar Matthew Kirschenbaum has called "the tactile fallacy": a presumption that such objects must lack materiality because "we cannot reach out and touch them." Although digital inscriptions are not permanently affixed to their storage media, at any one instant, as David Levy has observed, "the bits for a particular document are somewhere real and physical."[120] The same can be said of collections of digital objects: transient though such collections may sometimes be, at a given moment their components must have a material existence.

The collection is far from obsolete in the new world of dynamic ordering. For Manovich, as we have seen, the database exemplifies the symbolic forms of the digital age, yet Manovich has emphasized that databases "are collections of individual items."[121] On the Web, even an experimental site such as "Archive as Event" provides access to a defined collection. Flickr is a collection on a much larger scale, from which users can select collections of their own. Facebook, too, has been defined as a collection, albeit one that can be endlessly manipulated; according to Joanne Garde-Hansen, "each user's page is a database of their life, making this social network site a collection of collections."[122] The corpus of tweets maintained on Twitter's servers is a collection, as is the subset of these that was donated to the Library of Congress in the US.[123] The Internet itself can be perceived as a collection, perhaps the biggest collection of all.[124]

Every time I use a search engine on the Internet, I assemble a momentary collection of information about a small subset of that huge collection. When I browse through images on a website or database and place a selection of them into a basket labelled "My Favourites," I bring together a collection. Similarly, if I use third-order methods to construct an aggregate record, the resultant assembly is a collection of elementary records. Archival tradition presents record aggregation as an organic process, but collecting decisions seem inevitable whenever an aggregate record is realized. My ability to assemble aggre-

119 See, for example, Mark Poster, *The Mode of Information* (Chicago, 1990), 111–15.

120 Matthew G. Kirschenbaum, "Editing the Interface: Textual Studies and First Generation Electronic Objects," *Text: An Interdisciplinary Annual of Textual Studies* 14 (2002): 43; David M. Levy, *Scrolling Forward: Making Sense of Documents in the Digital Age* (New York, 2001), 156.

121 Manovich, *The Language of New Media*, 218.

122 Joanne Garde-Hansen, "My Memories? Personal Digital Archive Fever and Facebook," in *Save As ... Digital Memories*, ed. J. Garde-Hansen, A. Hoskins, and A. Reading (Basingstoke, England, 2009), 141.

123 Cf. Laura E. Campbell and Beth Dulabahn, "Digital Preservation: The Twitter Archives and NDIIPP" (2010), http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/campbell-27.pdf (accessed 1 December 2011).

124 Ernst, "Dis/continuities," 119.

gate records dynamically implies that there are both collections from which I select and collections that I make.

If my selections are to be made in a coherent manner, I will need adequate documentation of the collections from which I select and the perceived relationships of their components. What of the collections that I make? Do they also need to be documented? When a collection has been assembled for long-term use, it seems appropriate to declare it as an entity in a relational system and capture information both about the collection itself (when it was formed, who formed it, what adventures it has undergone) and about its contents. But third-order collections can be *very* temporary, put together to meet the needs of the moment and then seemingly cast into oblivion as the user moves on or the computer shuts down. Nevertheless, there may be a case for retaining information about them, if only to show which items were used and for how long. Moreover, third-order collections need not always be highly dynamic. When the *MoReq2010* standard for organizational records systems speaks of allowing item-level records "to appear in more than one aggregation," it seems to envisage that these aggregations will be maintained for an extended period.[125] Similarly, the overlapping collections assembled by visitors to a Web 2.0 site in the public arena may remain accessible to future visitors to the site. Regardless of whether a collection of item-level records is assembled by their creator or a subsequent user, information about the collection will offer insight into the conceptualizations and value judgments that underlay its formation. Privacy considerations may sometimes preclude documentation of collections assembled by individuals purely for their own temporary use, but even ephemeral collections can contribute to what Ramesh Srinivasan and Jeffrey Huang have called the "genealogy of ideas."[126] Any act of grouping item-level records forms part of their stories, and we may wish to consider documenting it in some way. Where accountability is critical, it may be essential to use technological tools to maintain audit trails of users and the collections they assembled or examined.

Collections abound in the digital realm, as they do in the analog. What changes is that fixed sub-collections lose their primacy. Broadly speaking, the overarching collection and the item survive, but the intervening layers (those we traditionally identified as physical series and files) become destabilized in digital space. Within a collection, visitors interact with elementary units at item level rather than with structures into which items have already been organized. Because sets of items can be brought together when we want them, structures emerge from interaction and no single structure need be canon-

125 DLM Forum Foundation, *MoReq2010 Modular Requirements for Records Systems, Volume 1: Core Services & Plug-in Modules* ([Luxembourg], 2011), 82.

126 Srinivasan and Huang, "Fluid Ontologies," 200.

ized. New collections can be constructed within or across the borders of those that existed previously; these new collections also need not have rigid internal structures, and their contents can likewise be mined and manipulated on demand. Some of these collections can be deemed to represent occurrents in the wider world and thus fulfill individual or organizational needs for records. In developing third-order systems that span the needs of different users, we can provide for the realization of aggregate records, but we can no longer adhere to customary assumptions about fixed boundaries and linear ordering.