Research Issues in Archival Provenance

Kenneth Thibodeau

Abstract This chapter reviews the conceptualization of archival provenance and the related concepts of archival fonds and original order and the ways these concepts have guided practice. Section 1 identifies problems entailed by the qualitative and imprecise conceptual approach traditionally applied to the issues involved in the management, use, and preservation of records, and suggests a multidisciplinary strategy to reformulate basic concepts. Section 2 describes problems arising from the traditional approach in greater detail. Section 3 indicates how a multidisciplinary strategy might be applied to both clarify theory and improve practice.

1 Introduction

The management of records and their preservation in accordance with the dictates of archival science and well-established practices has been guided by concepts articulated in a purely qualitative, largely philosophical, and often rhetorical mode. In a world where the information used in the conduct of affairs is increasingly encoded in digital bits and where digital information is growing exponentially [1], the conceptual foundations for managing records, both for current business and to enable exploitation of their long term value for diverse and often unforeseeable purposes, need to be reformulated with greater precision in representation, clarity in distinctions, and verifiability in implementation. This reformulation is needed to ensure that sound concepts and methods are implemented unambiguously and effectively, to increase the adaptability and efficacy of the governance of institutional information, and to respond to the challenges posed by the continuing stream of new forms of information and new ways of communicating and using information enabled by information and communication technologies (ICT).

K. Thibodeau (Retired) (✉)
National Archives and Records Administration, College Park, MD, USA
e-mail: KThibodeau@Fordham.edu

This chapter explores the potential for reformulation of the basic concepts that guide the management of records and archives through the application of systemic functional linguistics and mathematical graph theory.

2 A State of Confusion

Currently, the concepts that guide management, preservation and communication of records suffer from a combination of narrowness and vagueness. Perhaps surprisingly, the one follows from the other. Archival thinking and practice are constrained by the two overarching and complementary concepts of provenance and original order. While there are substantial disagreements about what the two concepts entail, there seems to be universal acceptance of their fundamental importance. Reverence towards the two concepts leads those who perceive issues with either their conceptualization or their implementation to focus on broadening the concepts, with the net result that they are overloaded to the point of becoming confusing and impractical.

Problems with the traditional conceptual approach to managing and preserving records begin at the most basic level. They include disagreements about both the intensive and extensive definition of 'record;' confusion between the quality of a record and its existence; failure to distinguish categorically or consistently the properties of the things that are designated as records from the properties of that which makes them records; overly zealous application of definition to justify the exclusion of documents from the application of records management controls that would serve organizational interests; and the mutation of successful techniques for addressing basic problems of managing records into abstract concepts that are proffered as essential requirements.

In principle, any type of persistent information object can be a record; moreover, an information object could be a record in one context and not in another. Similarly, the same object can be different records in different contexts. The criteria that determine whether an information object is a record are independent of the characteristics of the object itself. The properties of a document as such can make it a better or worse record, but they cannot make it a record in the first place. The key criteria that determine whether any given object is a record are (1) a record is a document that is used in an activity of the person or organization whose record it is and (2) it is kept, ideally under records management control. These criteria are often stipulated in laws, regulations, standards and policies related to records [2–4]. For example, the U.S. Federal Records Act defines records as "all recorded information, regardless of physical form or characteristics, made or received by a Federal agency under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation . . . ." [5].

Thus, there is no such thing as a record pure and simple. It must be a record of something: some act, activity, or state of affairs. What makes a document a record is the link between the document and the context in which it was used. This link may

not be evident in, or even reliably inferable, from the document itself. For example, one would infer from its content that a written contract for the acquisition of technical services by a government agency is a record of the agency's procurement activity; however, that is not necessarily the case. If the contract were found in the case files of a law enforcement agency, it would be a record of an investigation of some alleged illegal activity, such as bribery of a government official. The different contexts in which a document can be a record may even be independent of one another. Records created for weather forecasting can become important in criminal cases; for example, they could be used to impugn eye-witness testimony if they showed that, in spite of a full moon, there were dense clouds and therefore poor visibility the night a crime occurred.

The fact that a record is defined not by what it is intrinsically, but by its relationship to activity creates an anomalous situation with respect to the provenance of a record because traditionally archival provenance has been limited in scope to the records of a single records creator or a succession of records creators [6, 37]. Even when a broader scope is adopted, the provenance of records does not encompass actual creation of documents that are received from outside sources, such as incoming correspondence and reports.

In archival science and practice, provenance has both internal and external dimensions. Externally, provenance is delimited by the archival fonds, conceived as the totality of records created by a single records creator. Internally, provenance reflects the relationships between records and the activities in which they were instrumental. This reflection is seen as embodied in the way records were organized by their creators. Both dimensions are the cornerstones of archival theory and practice epitomized in the principles of *respect des fonds* and respect for original order [1]. Provenance encompasses the relationships between records and the entities that create, keep or use them [7]. Original order is "The organization and sequence of records established by the creator of the records" [8]. There are several problems with the ways both provenance and original order have been conceptualized and used.

To explain, archival provenance is determined when a document is captured or set aside as a record. This appears reasonable in the context of traditional hard-copy records. When records are inscribed on physical media in a hard and fast manner and placed in a physical folder, there is little likelihood that they would be changed, either in themselves or by relocation to a different folder. In this environment, information about who made the document a record in what activity could be a sufficient description of its provenance. However, even for hard copy records, the concept is overly restrictive. As the Australian concept of records series emphasizes, records can be used in the same activity by more than one actor. Governments, for example, carry out specified functions for as long as the laws that authorize or require them are in effect. But during that time government entities are often reorganized, resulting in different organizations carrying out the same function successively, often using the same records, or the same types of records, arranged in the same way [9–12]. The definition of provenance as a univocal and unalterable attribute of a set of records needs to be modified to reflect these and similar situations.

A second aspect of the arbitrary narrowness of the concept of record provenance is that it is bounded by the recordkeeping of an actor or succession of actors. In most cases the records of any given creator include a large proportion of documents that were not produced by the records creator, but acquired from other sources in a variety of ways. Records received from external parties, such as incoming correspondence, are obviously shaped by their authors. Records created in response to received records are also influenced by the authors of the incoming documents. The extent and limitations of acquired records, whether received from outside parties acting independently or purposely acquired at the initiative of the records creator, can also influence the records creator's decisions on whether it needs to acquire or produce other records to carry out an activity and, if so, what should be the form and content of the additional records.

Even when a records creator has extensive control over the submission of documents, their authors still have a decisive role. For example, the U.S. National Institutes of Health (NIH)[1] dictates the form of applications for research grants, publishes formal guidance on applying for grants, provides an online system for development, submission and tracking of applications, and publishes notices identifying specific areas where program managers believe that there is both a need for new knowledge and a significant probability of successful research [13–15]. Nonetheless, the specific subjects, methods, scope, and other details of the research projects proposed in grant applications are determined by the competence and creativity of the applicants, and shaped by factors such as their understanding of the state of science, their estimates of likely competition and alternatively of possible collaborations, the research resources available to them, the influence of their mentors and peers, the availability of other sources of funding, and even their knowledge of the government and external experts likely to be involved in reviewing their applications. The importance of researcher initiative and insight is reflected in the longstanding belief held by NIH managers that the best research results from ideas formulated by the researchers themselves, rather than by government officials [16]. Thus the substance of the research conducted with NIH grant funding is substantially determined by the applicants and their success or failure influences the subsequent direction of the research funding. Thus, limiting provenance to the records of the NIH arbitrarily truncates any effort to understand how and why the records are what they are. This observation can be extended to most contexts in which records are created, except in the uncommon and mostly uninteresting case of activities carried out entirely by one party.

An arbitrary limitation in the application of provenance in archival practice is that archival descriptions of provenance tend to be parsimonious [17]. There are, of course, some extensive descriptions of the provenance of records [18]; however, even in such cases the description of provenance tends to be limited to unembellished basic factual information that provides a picture of the overall context in which records were created and kept, and it is often articulated at the

---

[1]The author was the NIH Records Management Officer from 1978 through 1988.

level of the person or organization that created the records, rather than the specific processes in which they were created. This level of description of provenance rarely offers any in-depth insight into the genesis of any individual record or aggregate of records below the level of the archival fonds.

A third aspect of the arbitrary narrowness of the concept of record provenance is that, in the digital realm it cannot be assumed that records are not changed once they are filed. Record provenance could logically be extended from origin to the entire life cycle of records to address changes that occur after creation in a comprehensive and coherent framework.

The concept of provenance also suffers from overextension. Professional literature over the last several decades has often argued for broadening the scope of both the concept and what is included in it. However, these arguments render the concept so broad and vague as to be extremely difficult to apply practically [19].

Like provenance, the concept of original order also has problems at both the conceptual and practical levels. Original order is fundamentally important because the way an actor organizes the information used in carrying out its activity supplements the information contained in the records. First of all, very few actions are carried out using a single document. Thus, the set of records kept in an activity indicates the range and variety of information the actor considered relevant and valuable in that activity. Second, the act of keeping records indicates at least an implicit judgment by the actor that the value of the records extended beyond their initial use. Third, the organization of the records reflects how they fit into the activities in which they were used or expected to be used.

Original order is closely linked to the existence of records because, obviously, to be a record a document must be kept. In essence the organization of records is a solution to a requirement; namely, readily accessing all of the information that is most likely to be relevant to a current action and only that information. In many cases, that requirement is best satisfied by grouping together records that were used in a prior action. This solution is effectively regarded as a requirement in records management standards [2, 20–23, 36], in government regulations [3], in policies of private sector and non-governmental organizations [24–26], and in archival theory [27]. The organization of records in filing systems is conceptually sound and demonstrably beneficial; nevertheless, it is an artifact of the technology of hard copy records and is limited by the fact that, with that technology the optimal, if not the only, effective and efficient way to arrange records is by physical proximity. The success of filing systems in practice does not justify an assertion that they are either the only or the best way to manage records. ICT opens new possibilities that might prove superior; for example, managing records on the basis of where and how often they are communicated.

The organization of records materializes relationships among records and these relationships can reflect how they were used by the recordkeeper, but this is not necessarily the case. One could assume that records organized in case files assemble the most significant records used in the activity that defines the case and arrays them in the sequence in which they were created, but this assumption is not valid for other ways of filing records, such as in subject files or correspondence files.

A subject does not necessarily correlate to an activity or set of related activities, and a given correspondent may communicate about a range of subjects or activities. Moreover, even when files are organized to correlate with activities, it is unlikely, except in the simplest cases, that all of the records that had a significant impact on a particular case are assembled in the case file [28]. Overemphasis on original order can lead to the assumption that the filing of records equates with the totality of relationships among those used in a given activity or to a mistaken assumption that it represents a complete and unbiased expression of the relationships between records and activities [29]. As Georgio Cencetti stressed in articulating the concept of the archival bond, the relationships among records that result from their use in the same activities are inherent in the records [30]. They may be reflected in the way records are aggregated, but they are prior to and independent of the organization of records. Even when it does reflect the archival bond, treating the aggregation of records in files as the equivalent or even the preferred expression of the bond between records and activities leaves us ignorant of other evidence of this bond.

Like the concept of provenance, original order has been the subject of critical scrutiny in recent decades. Criticisms have included that the conceptualization and implementation of original order has failed to distinguish physical collections from logical arrangements [31]; that the assumption that there is a single, static set of relationships among records flies in the face of their recontextualization over time, notably in the processes of archival management [32]; that records creators may not explicitly organize records; and that the way the concept has been articulated is ill suited to the records of individuals [33].

The provenance and original order of records are closely related. Provenance describes the origination of records and original order their organization. In practice, however, use of the two concepts is essentially independent. Descriptions of the provenance of records focus on their creators, rather than the records themselves; moreover, they are commonly articulated at the level of the entire archival fonds. Application of the principle of provenance to actual collections of records is guided by and expressed in the "principle of provenance," which dictates that fonds be kept separate, but this principle fails to impose or even indicate any practice that would elucidate the provenance of individual records or record aggregates below the level of fonds. Respect for the original order of records impacts the management of records preserved in archives and is a cornerstone for the development of finding aids for those records, but both archival functions are accomplished in the main with little more than a bow in the direction of provenance.

3   Out of the Morass

Any path out of the morass described in the previous section needs to incorporate valuable observations that have surfaced in criticisms of the traditional approaches to provenance and original order while concentrating on articulating concepts in a manner that can be readily translated into practice. An obvious way to do this

would be to disentangle the multifaceted and not necessarily compatible insights that have been advanced in recent criticisms, reformulating the conceptual foundations of records management and archives to achieve clearer differentiation and easier implementation. The challenges posed by explosive growth and increasing diversification of digitally encoded records make it highly desirable that reformulation express the requirements for managing records in a manner that enables automated implementation, verification, and measurement.

The open-ended growth in both quantity and variety of born-digital records presents both substantial challenges and unprecedented opportunities for the management and preservation of records. Besides the quantitative challenge, novel intrinsic properties of digital records, including genres that do not and even cannot exist outside of the digital realm, and different ways of expressing and preserving their relationships compel rethinking traditional concepts and re-examining established methods. Digital records offer an important opportunity to enrich the concept and expand the use of provenance below the level of the archival fonds. While this would be impractical with hard copy records, provenance of individual and aggregate digital records can be captured from metadata generated automatically when digital records are created, revised and used; for example, using transmission data for email and audit trails on system use. Furthermore, ICT provides tools that could be used to capture automatically additional data related to provenance.

This reformulation in practice could be complemented at the theoretical level by adopting and adapting concepts, methods and tools from other disciplines. Propitious opportunities for enriching the theoretical constructs applied in managing records come from systemic functional linguistics, which focuses on language "that is doing some job in some context" [34]. This emphasis on the function of language and the concomitant recognition that function cannot be understood apart from context parallels the view of records as instruments and by-products of the conduct of affairs. More specifically, systemic functional linguistics offers a systematic approach to capturing and organizing different aspects of the provenance of records through the adoption of its differentiation of context into field, tenor and mode of discourse. Field of discourse refers to the action or interaction in which language is employed. Tenor of discourse refers to the parties who participate in the activity, their roles, relationships and relative status. Mode of discourse refers to the role both spoken and written language plays in each context, and addresses how it is expressed, how it is organized, and what it achieves. Activities, participants, and the modes and functions of expression are all crucial in understanding records. Adopting the specific constructs of field, tenor and mode of discourse as they have been articulated in systemic functional linguistics offers the opportunity for describing facets of the provenance of records in a clear and empirically verifiable manner.

Another discipline that can enhance the treatment of archival provenance is the branch of mathematics called graph theory. Graph theory offers suitable, quantitative methods of analysis and opens possibilities for the use of automated analytical and visualization techniques that are well suited to the objectives of records management. Graph theory can be applied to capture, but still distinguish, different

aspects of provenance. In graph theory, a graph consists of a set of nodes connected by arcs, where the nodes represent things and the arcs relationships between nodes. Graphs are differentiated according to the types of things represented as nodes and the types of relationships between the nodes. The different aspects of context distinguished in systemic functional linguistics can be used to define corresponding archival graphs. In a graph of the field of discourse, the nodes could be either different activities or the various steps or stages of a single process and the arcs the transitions from one activity or step to another. The tenor of discourse could be graphed by identifying each person or organization involved in an activity as a node. Different graphs would result from selecting different types of relationships between parties; for example, one graph might depict interactions between parties as arcs, while another might indicate the relative status of individuals within an organization. Graphs of the mode of discourse would encompass records as nodes. The organization of records in a record-keeping system would depict the classification and placement of records aggregations as arcs. Another graph of mode could display derivation relationships among records revealing connections that would not surface in a graph or the arrangement of records. For example, many records are articulated in accordance with directives, but regulations, policy statements and other directives are usually not filed with the records representing the instances in which such directives and other policy documents are applied. Graph theory opens possibilities for extensive and precise description of relationships among activities, parties, and records through the superimposition of graphs of each of them [35]. System and human generated data could also be used to construct heterogeneous graphs showing the involvement of persons and organizations in activities and their participation in the generation and use of records.

Graphs of the different facets of the context of records creation, keeping and use would also benefit researchers who use records. Contextual graphs could help them to discover related records. Moreover, they could expand or extend such graphs by adding nodes and/or arcs that are of particular interest in their research, even extending the graphs to include outside parties that interact with the persons or organizations that create or keep records, such as correspondents, customers, and authorities.

To conclude, this chapter has reviewed the conceptualization of archival provenance and the related concepts of archival fonds and original order and the ways these concepts have guided practice. In doing so, it has identified problems entailed by the qualitative and imprecise conceptual approach traditionally applied to the issues involved in the management, use, and preservation of records, and suggested a multidisciplinary strategy to reformulate basic concepts, positing the possibility of applying systemic functional linguistics and graph theory as possibilities for finding a way out of the existing conceptual morass.

References

1. Horsman, P.: The last dance of the phoenix, or the de-discovery of the archival fonds. Archivaria **54**, 1–23 (2002)
2. International Organization for Standardization: ISO 23081-1 Information and Documentation – Records Management Processes – Metadata for Records: Principles. ISO, Geneva (2006)
3. National Archives and Records Service of South Africa: Rec. Manag. Policy Man. **7**, 15–20 (2007)
4. United Nations Secretariat: Secretary-General's Bulletin: Record-keeping and the Management of United Nations Archives. United Nations. ST/SGB/2007/5 (February 17, 2005), https://archives.un.org/sites/archives.un.org/files/ST_SGB_2007_5_eng.pdf
5. 44 United States Code § 3301(a)(1)
6. International Council on Archives – Committee on Descriptive Standards, "Glossary of Terms Associated with the General Rules," General International Standard Archival Description ISAD(G), 2nd edn. International Council on Archives, Ottawa (1999), http://www.icacds.org.uk/eng/ISAD(G).pdf
7. InterPARES 2: Dictionary, http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=6354166&CFTOKEN=11363511
8. Pearce-Moses, R.: Glossary of Archival and Records Terminology. Society of American Archivists, Chicago (2005). http://www2.archivists.org/glossary/terms/o/original-order
9. Scott, P.J., Finlay, G.: Archives and administrative change: some methods and approaches (Part I). Arch. Manuscr. **7**(3), 115–127 (1978)
10. Scott, P.J., Finlay, G.: Archives and administrative change: some methods and approaches (Part 2). Arch. Manuscr. **7**(4), 151–165 (1979)
11. Scott, P.J., Finlay, G., Smith, C.: Archives and administrative change: some methods and approaches (Part 4). Arch. Manuscr. **8**(2), 51–69 (1980)
12. Scott, P.J.: Archives and administrative change: some methods and approaches (part 5). Arch. Manuscr. **9**(1), 3–18 (1981)
13. National Institutes of Health: Application Submission System & Interface for Submission Tracking (ASSIST) User Guide. System Version 2.15. Document Version 6.3.0 (October 16, 2015), https://era.nih.gov/files/ASSIST_user_guide.pdf
14. National Institutes of Health: NIH Grant Application Submission and Review. Useful Web Links, http://public.csr.nih.gov/aboutcsr/NewsAndPublications/Publications/Documents/ReivewWebLinks.pdf
15. National Institutes of Health: NIH Guide for Grants and Contracts, https://grants.nih.gov/grants/guide/index.html
16. Collins, F.: Opportunities for research and NIH. Science **327**, 36–37 (2010). https://www.genome.gov/Pages/Newsroom/Webcasts/2010ScienceReportersWorkshop/FCIntro_Opportunities-for-Research-and-NIH.pdf
17. National Archives and Records Administration (U.S.): Online Public Access, http://www.archives.gov/research/search/
18. Provincial Archives of Alberta, Canada: An Administrative History of the Government of Alberta 1905–2005 (2006), http://www.culture.alberta.ca/paa/archives/research/documents/contents.pdf
19. Douglas, J.: Origins: evolving ideas about the principle of provenance. In: Eastwood, T., MacNeil, H. (eds.) Currents of Archival Thinking, pp. 23–43. CA. Libraries Unlimited, Santa Barbara (2010)
20. International Organization for Standardization: ISO 15489: Information and Documentation - Records Management. ISO, Geneva (2001)
21. International Organization for Standardization: ISO 15801 Trustworthiness and Reliability of Records Stored Electronically. ISO, Geneva (2009)
22. Object Management Group: Records Management Services (RMS) Version 1.0 (2011), http://www.omg.org/spec/RMS/1.0/PDF/

23. U.S. Department of Defense: Electronic Records Management Software Applications Design Criteria Standard (2007), http://www.dtic.mil/whs/directives/corres/pdf/501502std.pdf
24. Bilotto, A., Guercio, M.: The management of corporate records in Italy: traditional practice and methods and digital environment. Rec. Manag. J. **13**(3), 136–146 (2003)
25. British Broadcasting Corporation: Records Management Policy, Version 1.4 (2010), http://www.bbc.co.uk/guidelines/dq/pdf/media/records_management_policy_v1.4.pdf
26. Chachage, B., Ngulube, P., Stilwell, C.: Developing a model corporate records management system for sustainability reporting: a case of the Iringa region in Tanzania. S. Afr. J. Inf. Manag. **8**(1), 1–17 (2006). http://www.sajim.co.za/index.php/SAJIM/article/viewFile/217/213
27. Boles, F.: Disrespecting original order. Am. Arch. **41**, 26–32 (1982)
28. Thibodeau, K.: Out of Bounds: Breaking the Chains of Original Order to Exploit the Potential of the Archival Bond. Association of Canadian Archivists, Annual Conference 2014. Victoria, British Columbia (2014), https://www.academia.edu/18470021/Out_of_Bounds_Breaking_the_Chains_of_Original_Order_to_Exploit_the_Potential_of_the_Archival_Bond
29. Duranti, L., Guercio, M.: Research issues in archival bond. In: Bearman, D., Barata, K., Trant, J. (eds.) Electronic Records Research 1997: Resource Materials. Archives & Museum Informatics, Pittsburgh (1998). http://www.archimuse.com/erecs97/s1-ld-mg.HTM#_jmp0
30. Cencetti, G.: Sull'archivio come "Universitas Rerum". Archivi **4**, 7–13 (1937)
31. Yeo, G.: The conceptual fonds and the physical collection. Archivaria **73**, 43–80 (2012). http://www.members-archivists.ca:8080/ojs2/index.php/archivaria/article/view/313/413
32. Macneil, H.: Archivalterity: rethinking original order. Archivaria **66**, 1–24 (2008)
33. Meehan, J.: Rethinking original order and personal records. Archivaria **70**, 27–44 (2010)
34. Halliday, M.A.K., Hasan, R.: Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective. Oxford University Press, Oxford (1989)
35. Lemieux, V.L.: Toward a "Third Order" archival interface: research notes on some theoretical and practical implications of visual explorations in the Canadian context of financial electronic records. Archivaria 78 (2014). http://journals.sfu.ca/archivar/index.php/archivaria/article/view/13493
36. Hilbert, M.: How much information is there in the "Information Society"? Significance **9**(4), 8–12 (2012). http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2012.00584.x/pdf
37. National Archives of Scotland: Records Management (July 8, 2013), http://www.nas.gov.uk/recordKeeping/recordsManagement.asp