The Language of Archives: Essence and Identity

Richard Pearce-Moses
Presented at the NHPRC Archival Research Fellowship Symposium
Boston, Massachusetts
Friday, 13 June 2003

> After all, the cultivated person's first duty is to be always prepared to rewrite the encyclopedia.
>
> —Umberto Eco, *Serendipities*[1]

In his book on dictionaries, Sidney Landau noted, "Not the least value of lexicography is that one learns to be humble about one's own knowledge of the language."[2] I would extend his observation by noting that writing a professional glossary greatly adds to one's knowledge of the field.

Today, I will reflect on my experience of being immersed in the language of archives for the past eighteen months. I will start with a short discussion of the mechanics of writing the glossary, then offer some observations on archivists' language, and finally touch on some of the hard words I've wrestled with.

AN ODD HOBBY

As a bit of background, I thought I would comment on why I tackled a revision of the Bellardo's *Glossary*.[3] Some people like baseball; I don't. A good friend plays golf and fishes. I am so happy for him. I like words and historical lexicography. I also like typography. In other words, I'm a word wonk. There are worse addictions; crack cocaine or gambling, for instance.

I am fascinated by the fact that language – something so essential to us – is largely transparent. Words surround us like air, and we are largely oblivious to them. Only when there is some disturbance do we notice them. When I began this project I recalled a poem by Theodore Roethke that I think is appropriate.

> THE CHAIR
>
> A funny thing about a Chair:
> You hardly ever know it's *there*.
> To know a Chair is really it.
> You sometimes have to go and sit.[4]

---

1. p. 21.

2. *Dictionaries: The Art and Craft of Lexicography*, 2nd ed. (Cambridge: Cambridge University Press, 2001), p. 115.

3. Lewis J. Bellardo and Lynn Lady Bellardo, *A Glossary for Archivists, Records Managers, and Manuscript Curators* (Chicago: Society of American Archivists, 1992).

4. Originally published in *I Am! Says the Lamb* (1961). Included in *The collected poems of Theodore Roethke* (Garden City, N.Y: Anchor Press/Doubleday, c1975).

I had already begun developing a glossary of terms relating to electronic records to facilitate the Arizona State Library and Archives' efforts to build collaborative partnerships between the diverse community of information technologists, records managers, archivists, and others with responsibility for e-records. I also had a bit of experience in terminology, having worked on the scope notes for photographic terminology in the *Art and Architecture Thesaurus*. Tackling a revision of the SAA glossary seemed like a logical thing to do.

<div align="center">Ↄ Ↄ Ↄ</div>

MECHANICS

A COLLABORATIVE EFFORT

Before I go any further, I want to note that the glossary is a collaborative effort and acknowledge my colleagues. Mark Greene of the American History Center at the University of Wyoming, Diane Vogt-O'Connor of the National Archives and Records Administration, and Rob Spindler of Arizona State University serve as advisors. They help guide decisions about the work in general and review definitions. I'd also like to note that Diane Vogt-O'Connor is getting input from Lew Bellardo, who co-authored the second edition, and several others at the National Archives.

Conversations with the advisors were lively, interesting, and occasionally a little heated. I am not using those words to avoid 'contentious'. To the contrary, it was my honor to work with a group that was so collegial and harmonious, even when disagreeing. Over the years, and during this project, they have taught me as much about what it means to be a professional as anything else. I am humbled by their knowledge, generosity, support, and friendship.

One of the themes of this paper will be language and identity of the archival community. The glossary is also a work of autobiography. The selection of words and their definitions reflect my own career and interests. While I have sought to be as objective as possible, working with advisors, seeking others' opinions, and justifying definitions with citations from the literature, the work will ultimately be a reflection of me. Another person would have written another glossary. I say this not as self-aggrandizement, but to clearly paint the target for any criticisms on my own forehead.

AUDIENCE

The first two editions of the glossary were "for archivists, manuscript curators, and records managers." I am targeting a wider audience of anyone who needs to understand records because they work with them. In particular, I want to create a tool that can help interpret archival concepts to people working with records in information technology and business.

The glossary is based primarily on archival practice in the United States and Canada, in that order. In a few instances, I have included terms, definitions, and citations from other English-speaking communities when I think they are relevant to the primary audience.

The Bellardos' glossary contains about 800 defined terms plus references. The current edition has more than 2,000 defined terms plus some 400 headings that are lead-ins to other terms.

The glossary includes terms related to the profession that an archivist is likely to encounter when reading the literature or when working with a fairly typical collection of archival materials. I've made it a point to include terms necessary to understand electronic records. In addition to terms clearly relating to archives and records management, the glossary includes important terms from preservation, information technology, law, and micrographics. In particular, I have sought to include terms that shift meanings between different communities of use, such as record or archives. I have included some words that are no longer in common use, but will be useful when reading older literature; for example, Spindex.

In general, I have not included words that have no particular archival connotation. While aisles are common in archives, any good dictionary will provide an adequate definition of the term. I have included a few common words which function as guide terms; these terms help illustrate the relationships between other terms. For example, 'box' needs no definition, but is included as a bucket term to group cross-references to related terms, such as Hollinger box, Bankers Box, and phase box. Many words specific to some affiliated professions, such as rare books, printing, or bookbinding, were not included.

Some of the more animated discussions with my advisors have had to do with words that I find personally interesting, but may be of limited use to the audience. Sometimes I think a word is cool, like *eschatacol*, precisely because it is arcane. More often, I found that the word – and its underlying meaning – captures a piece of the profession's history and provides context for and insight into current practice. JoAnne Yates' *Control Through Communication* was a valuable source of terms from the nineteenth and early twentieth centuries. Those words, based in paper records, gave me significant insights into how people used paper records to solve problems in the past, and from that I gained a better understanding of parallel problems in the digital present. One reason I tend towards including those terms is that I see modern parallels. *Ars dictamen*, manuals of letter writing with established forms that could be used as boilerplate, persists today and some word processing programs come with canned text that are, in my mind, equivalent. Some terms represent what is now obsolete practice, but might be worth revisiting. Red-to-black headings point from a topical heading (say, Arizona – Governors) to more specific headings (Evan Meacham, Rose Mofford, and Fife Symington); for archival catalogs, red-to-black headings (regardless of color) might be a particularly useful tool to complement provenance-based access, leading patrons from general subjects to relevant collections.

DEFINITIONS, CORPUS LINGUISTICS, AND LITERARY WARRANT

Samuel Johnson believed that one of the lexicographer's jobs was "to correct or proscribe . . . improprieties and absurdities." However, he also stated that his intent was "to register the language . . . ; not [to] teach men how they should think, but relate how they have hitherto expressed their thoughts."[5]

---

5.   Cited in Landau, *Dictionaries*, p. 62.

The previous two editions of the glossary were prescriptive. The first edition sought to standardize the language, presenting preferred terms and meanings. The Bellardos continued that practice, noting that they "identified preferred terms and developed definitions reflecting the practices of leading archival institutions and professionals."[6]

The current edition of the glossary is descriptive. Rather than serving as an arbiter of correctness, the glossary documents the different ways a term is used within the profession; while some archivists carefully distinguish archives from personal papers, others do not. The glossary also notes when meanings differ between communities of interest; for example, many terms have a sense that is specific to cataloging, to information technology, or to law.

I think it is important to document these differences in meaning. The variations point to areas where there is no professional consensus. As a result, the glossary points out horizons of understanding where key concepts are being reconsidered or established. Further, I believe that including a single definition for a term could be confusing to a reader who finds in the glossary a definition that is very different from the meaning intended by an author.

Although designed to be descriptive, the glossary can be used prescriptively in some instances. I would like the glossary to serve many communities and purposes, and some communities need a prescriptive definition. *Describing Archives: A Content Standard*, being developed by CUSTARD, has a technical glossary of descriptive terms which will be included in the glossary. Those definitions will be flagged so that someone looking to the glossary for descriptive terms would be able to distinguish the definition preferred for that purpose.

Within limits, the definitions are written according to the lexicographical principle of substitutability. It should be possible to replace a word in a sentence with the definition and have the sentence scan correctly. To do that perfectly, it would be necessary to include a definition for each part of speech of a word; for example, rely (verb), reliable (adjective), and reliability (noun) would have separate definitions. The glossary typically includes a definition for only the noun form, although other parts of speech are included if there is a shift in sense.

To provide additional understanding, the definitions may be supplemented by notes. My model here is Fowler's *Modern English Usage*.[7] Since beginning this project, John Ciardi's *A Browser's Dictionary* has become another model. For both authors, clarity is the proper measure of correctness. I hope that like these two authors' works, the glossary will provide insight into the language to enable people to communicate clearly and effectively.

Many terms have citations from the literature illustrating use and to provide additional understanding of difficult concepts.

Finally, the glossary includes extensive cross-references that help illustrate the relationships between terms. Initially, I had intended to follow the ISO/ANSI standard for thesaurus construction. However, the cross-reference structure in that standard is fairly simple. It provides for hierarchical relationships

---

6.   p. v.

7.   *A Dictionary of Modern English Usage*, 2nd ed., revised and edited by Sir Ernest Gowers (Oxford: Oxford University Press, c1965).

(broader and narrower terms), equivalency relationships (use and used for), and a third catch-all relationship of related terms. I am revising the cross-reference structure using an expanded set of relationships.

NORMALIZATION

One of the temptations facing lexicographers is a desire to normalize the language, to make it more rational. In its most innocent form, it's a compulsion to tidy up those areas of ambiguity around the details of meaning, to correct the exception by making it fit the mold. A darker side of normalization is for the lexicographer to interject personal bias into the work by preferring one definition over another, often in deference to a particular point of view. One of the reasons I'm committed to a descriptive glossary based on literary warrant is that I do not want to be accused of preferring Bearman, Cox, Duranti, Hedstrom, Jenkinson, Norton, Schellenberg, or myself.

Jenkinson provides an excellent example of normalization when he synthesizes a definition of archives in his *Manual*.[8] Not satisfied with the definition in the *Oxford English Dictionary*, he constructs his own "by comparing in some well-known case documents which are obviously Archives with others which are obviously not."[9] Constructing a definition from examples is reasonable, and I have tried hard to think of the variety of things to which a term refers so that a definition is not too narrow. However, another person in another situation could just as legitimately select different examples and wind up with a very different definition. Having begun my career some sixty years after Jenkinson published his *Manual* and working with photographic archives and local history collections in the United States, Jenkinson's definition strikes me as archaic, alien, and artificial. The foundation of Jenkinson's definition does not fit my experience, so his logic falls apart.

The issue here is not right or wrong. Rather, the tension results from the fact that, given language's irregular nature, it is possible for different people to impose or discern different orders on language.

METHODOLOGY

The glossary is based on principles of corpus linguistics, which establishes definitions based on how a term is used.

I began compiling a list of headings by rekeying the Bellardos' glossary. I then spent about a year looking at other glossaries, such as the UNESCO glossary of A/V terms, the *Art and Architecture Thesaurus*, and the AIIM and ARMA glossaries. I looked for works of archival and related professions' literature that had glossaries, such as RAD and AACR2. I reviewing past issues of *American Archivist* and – to help capture Canadian terms and use – *Archivaria*. I read key monographs. I'm relying on *Black's Law Dictionary*, the *Corpus Juris Secundum*, and the second edition of *American Jurisprudence* for legal definitions. Currently, more than 300 sources are cited.

---

8.   Hilary Jenkinson, *A Manual of Archive Administration* (London: Percy Lund, Humphries, 1966), p. 2-15.

9.   p. 3.

As I reviewed the literature, I am building a corpus of quotations which provide literary warrant for the definitions. Some of those quotations will be included in the published version. Currently, the database contains more than 4,500 quotations.

Originally, I expected that I would have a solid draft completed in August. I based that projection on the assumption that the glossary would be only a little larger than the second edition. Now that it contains nearly three times the number of defined terms, I am shooting to finish the draft by the end of the year. That's an aggressive deadline, and some days I don't think I'm going to make it.

When the draft is completed, it will be submitted to the SAA Publications Board for review. I expect additional comments on the definitions from the reviewers, and I would hope that any additional work would be finished within six months. My hope is to have the final copy finished by the beginning of summer 2004.

The work will be published in hard copy and will be made freely available on the Web.

I already envision a fourth edition. Illustrations would be enormously valuable. I keep finding more sources to check for additional terms and meanings. Many subdisciplines have fairly technical definitions based on genres and formats that fall outside the more general scope of the current edition. For example, audiovisual archives have many obsolete formats, some obscure and others rather common.

<center>ಡಿ ಡಿ ಡಿ</center>

CAUGHT IN A WEB OF WORDS

When I told a colleague that I was interested in revising the SAA glossary, he counseled me, "No, you aren't. Too many battles to be fought, and you'll be caught in intellectual cross-fire. You can't win."

My friend had a good point. Many people are passionate about words, and monkeying about with something so emotionally loaded was a bit risky. Especially risky in that passion, rather than rational intellect, tends to drive arguments about use. The linguist Stephen Pinker observed that

> Whenever pedants correct, ordinary speakers hypercorrect, so the attempt to foist "proper" Greek and Latin plurals has bred pseudo-erudite horrors such as *axia* (more than one *axiom*), *peni*, *rhinoceri*, and [*octopi*].[10]

Nevertheless, I ignored my friend's advice and dove in. Some eighteen months later, I am surfacing to offer some observations. I'd like to stress that there will always be exceptions to these generalizations. Also, I intend none as criticism. The more I think about the problems of language, the more I am amazed that we can communicate at all. Language is hard work. Effective use of the language is even harder.

---

10. *Words and Rules: The Ingredients of Language* (New York: Basic Books, c. 1999), p. 55. He continues, "It should be "Fellow octopuses." The *-us* in *octopus* is not the Latin noun ending that switches to -i in the plural, but the Greek *pous* (foot). The etymologically defensible *octopodes* is not an improvement."

LANGUAGE AND IDENTITY

No doubt, people are passionate about language because it defines the boundaries of our identity. Archivist – the word most clearly associated with our identity – may be the most hotly contested word.

Who can call themselves an archivist? Must an individual be certified? Must an individual have a graduate degree? Does that degree have to be in archival science? What amount of experience, if any, is a substitute for those measures? Is there a difference between a manuscript curator and an archivist? Is an archivist also a records manager? Subscribers to Archives and Archivists debated, discussed, and flamed the issue in the 1990s.

The flip side of inclusion is exclusion. Some define archivists by pointing those outside the professional boundary, such as clerks or records custodians.

An underlying issue here is how we define ourselves. We can base our understanding of 'archivist' on a number of criteria. We can define an archivist based on certain characteristics; for instance, education. We can define an archivist based on activities, such as preservation, or on the nature of those activities, organizing materials with respect to provenance and original order. We can define an archivist on the basis of materials; for example, one who works with records.

Complicating the issue, archivists do not have the exclusive right to define themselves. Others recognize us as a group, and they apply their own criteria. More often than not, those outside the profession define us by function: the people who care for old records. They may care very little about many of the distinctions we make.

LANGUAGE IS AD HOC

The challenge of defining 'archivist' is typical. A word's meaning is derived from its use. Answering Juliet's question, "What's in a name?", linguist Stephen Pinker observes

> What's in a name is that everyone in a language community tacitly agrees to use a particular sound to convey a particular idea. Although the word rose does not smell sweet or have thorns, we can use it to convey the idea of a rose because all of us have learned, at our mother's knee or in the playground, the same link between a noise and a thought.[11]

Unfortunately, there is seldom exact agreement on meaning. People have a consensual understanding of the general concept, but disagree on the details. Thus, most everyone would agree that archivists work with records, but whether they must have a certain level of professional knowledge or follow a certain methodology is less certain.

This fuzziness surrounding meaning gives words their easy-to-miss nuance. NPR's Morning Edition interviewed Daniel Jurafsky, a winner of a MacArthur Fellowship and a computational linguist. He noted that there's a difference between uh-huh and yeah. Uh-huh means, 'Go on, I'm listening.' Yeah

---

11.   *Words and Rules*, p. 2.

means, 'I want to say something.'[12]   Unless one pays very close attention to words, the subtlety escapes conscious awareness.

Most of us don't learn language by reading dictionaries and grammars. We learn language the way most of us learn to drive. We observe people around us. In Pinker's words, "at our mother's knee." Before I got behind the wheel the first time, I could name the steering wheel, the pedals, and the stick shift, and I knew what they did. I had to transform the observational knowledge that I had acquired from watching other people drive me around to the practical knowledge of shifting gears and balancing my attention between many different things. We learn language by hearing how other people use the language. We try out new words; if we use those words wrong, we get feedback in some manner that corrects our behavior.

Some of us go to driving school for formal education, and I assume we all went to school and learned something about language. Even in school, though, a great deal of our education is more literature than language. Only those who study a new language are likely to be instructed through formal grammars and vocabulary lists. When I walked into my first course in Russian in college, I had never heard anyone speak more than a да or нет in a movie; I had no one to observe, no opportunity to learn by osmosis.

As archivists, few of us have formally studied the language. We have studied the concepts of archival science, and we have learned to associate those concepts with the words. Few of use are instructed in methods of historical linguistics or lexicography. The observation is not a criticism, and it is not unique to archivists.

Last year, I was reading a draft of a paper when one particular sentence threw me into high dudgeon.

> With more and more information now created in electronic, rather than physical form,
> records may need to be re-defined for an electronic medium . . . .

With some incredulity I pondered, *Who* was going to re-define record? As my thesis advisor Robert Crunden pointed out numerous times, the passive voice is a thing to be avoided because it masks the agent of change. If the authors rewrote the sentence in the active voice, Who would they expect to provide the new definition? The passive voice masks a presumption that the active voice would have underscored; there is no authority that establishes definitions for English words.

Technical standards are one of the few places we have precise definitions. We know what A4 means because an ISO standard tells us. However, this technical language is exceptional because it is created artificially. Ordinary language is not created systematically or rationally; it is constantly changing through an ad hoc process. It is a living thing.

In ordinary language, words evolve and their meanings shift.
›   Photographs have traditionally been distinguished from prints; the former made by photochemical reactions, the latter by mechanical processes. However, people refer to images made using digital cameras and printed on their computers as digital photographs, even though they would technically be classified as prints.

---

12.   http://www.npr.org/ramfiles/me/20020925.me.17.ram.

› Hacker was originally a term of praise, referring to an individual who was an expert programmer. Today, it's come to mean an individual who breaks into a system without authorization.

› Coming from the Southwest, I can tell you that everyone knows what cacti are. However, the word is a questionable Latinate backformation. Ceasar's Romans knew no variety of cactuses.

Arguing about which meaning is correct verges on pedantry. However, It *is* possible to determine which meanings are prevalent and any that are idiosyncratic or obsolete. I quit using the term hacker in its original sense; I didn't like giving up a word that had been part of my own identity, but I didn't have much choice if I wanted to be understood.

New words regularly become part of the language. Technology has changed significantly since the Bellardos' edition was published in 1992. Big iron still reigned. Intel had recently introduced the 486. Gopher technology was the hot new thing, and few people had heard of the http and html standards that would lead to the Web. Technology has been a significant source of new terms that archivists need to know.

Although it's a trivial observation that language is messy, I'll repeat it to emphasize that we don't do a particularly good job communicating because we don't pay enough attention to words. We tend to load a lot of meaning onto common words. We tend to forget that the rich context surrounding a word differs from individual to individual. When I compiled a union guide, I asked repositories for information about their collections of records. Two repositories responded that they didn't have collections, although it turns out they had large quantities of relevant materials. I was using collections to mean any example of records, individually or collectively. They understood collection more narrowly to mean purposefully created bodies of records.

JARGON, BUZZWORDS, AND MAGIC WORDS

Professions use some words as shorthand. Jargon used within a professional context where it is understood is a handy shortcut. When using the language outside that community, it is often confusing. The more archivists work with other communities, the more we must take the time – and words – to explain ourselves to those communities. It slows us down, but it's essential for effective communication.

Unfortunately, people who are concerned with intellectual trendiness often make a mess of the language. They use hot new jargon – buzzwords – without quite knowing quite what those words mean. After looking at the many often conflicting ways post-modernism was used in the literature, Bruce Handy asked

> "What do these people mean? We aren't sure. More to the point, they aren't sure. And just to make sure, we've tried to track down some of the writers . . . in order to discuss the postmodern question.
>
> ◦ Elle editors were unwilling to explain why ski parkas were postmodern. "We're busy," complained a spokeswoman.
>
> ◦ "Mr. Safire hasn't really addressed himself to the question," said Safire's *New York Times* assistant in response to our request for an interview regarding Paul Simon's postmodernism. "You can read about it in his column if it becomes timely."

◦ "It has become kind of vague and catchall, hasn't it," replied Ben Bantley when asked about his reference to the postmodern age. "The quote [from Schnabel] got distorted at the end – it's a reference to a Bardot sort of thing."

Handy concluded that postmodern had become "culturespeak, short for Stuff That's Cool in 1988. It's the current version of groovy – except that using it makes you sound smart."[13]

Fowler defined cant as "insincere or parrotlike appeal to principles, religious, political, or scientific, that the speaker does not believe in or act upon, or does not understand."[14] Staff at the National Endowment for the Humanities calls cant 'magic words', which often appear in grant applications. Applicants use words they think NEH wants to hear, but their use of the words made it clear that they don't understand them. For example, they may promise that finding aids will be in MARC format because they know that MARC has something to do with description and is a standard NEH looks for. Unfortunately, the sample finding aid clearly had nothing to do with MARC.

In the archival community, hot new buzzwords and magic words include evidence, digital object, resource, metadata, trustworthy, and the phrase 'authentic and reliable'. I can't tell you how many times I've read that electronic records must be authentic and reliable without any indication of what that means.

Buzz words are likely to have important meanings, but many people use them as filler words that could mean a dozen different things. The issue is not purity of language. My concern is that sloppy language reflects sloppy thinking.

<div align="center">ও ও ও</div>

HARD WORDS

I've found that defining some words is more difficult than others. Some of the more challenging words that I've struggled with include

> accession, acquisition, appraisal, archives, authenticity, digital object, document, evidence, genuine, permanent, record, reliability, and trustworthy.

Many of these definitions were difficult because the terms were interrelated. For example, I finally realized that accession and acquisition were synonyms when nouns, but not when they were verbs.

Many of the words are charged by considerable discussion in the professional literature. For example, should archives be defined as records of permanent value or enduring value? Trying to tease meaning from polemics can be a bit like walking on a mine field, and I often hear the voices of different pendants in my head telling me I'm wrong. It would be a lot easier to ignore those voices if I didn't have a great deal of respect for them.

---

13. "The Rise and Fall of a Great American Buzzword."

14. Fowler, p. 315.

The rise of electronic records has made it particularly difficult to define some words because archival terminology has been appropriated by information technologists. In other cases, different people use a word with different meaning, reflecting an evolution of how the word is understood.

I'd like to spend a little time talking about a few more problematic words.

ARCHIVIST

> n. ~ An individual responsible for appraising, acquiring, arranging and describing, preserving, and providing access to records of enduring value, and who relies on the principles of provenance, original order, and collective control to protect the materials' authenticity and context.

> Notes: In the United States, archivists are typically associated with collections of inactive records. However, the European tradition includes management of active records as well, which in the United States is often the responsibility of a separate records manager. In the United States, archivists may be called 'manuscript curators', especially if they are responsible collecting and administering collections of historical documents acquired from individuals or other organizations.
>   In some organizations, an archivist may be responsible for management of active, inactive, and archival records. In other organizations, an archivist may be responsible only for those records transferred to the archives. In a large repository, a practicing archivist may specialize in only one or a few archival functions noted above. A teaching archivist may not be currently responsible for collections, but is familiar with the theory and practice of archival functions.

I've already touched on the first hard word, archivist, so I won't belabor it much further. The literature touches on a number important issues relating to the definition of archivist. Two stand out for me.

› To what extent should archivists work with current records? Luciana Duranti believes that current records remind archivists why people create records, and that historical records connect archivists to the their role as guardian of societal memory.[15]

› To a list of fairly widely accepted archival function, William Maher adds "authenticating the validity of the evidence held."[16] Kaplan, on the other hand, suggests that archivists' efforts to preserve authentic history is far from the objective truth many expect. "As archivists, we do not seem to recognize that ours is a subjective endeavor, and we rarely present it as such. Yet authentic voices

---

15. Duranti, Luciana. "Meeting the Challenge of Contemporary Records: Does It Require a Role Change for the Archivist?," *American Archivist* 63:1 (Spring/Summer 2000), p. 7-14. "If the archivist, focusing on the needs of the researchers, becomes detached from current archives, he divorces real life, renounces his responsibility as guardian of people's rights, loses contact with the experience of change, and gets lost in the multiple expressions of archival research [citing Oddo Bucci.] If the archivist, focusing on the needs of the creator, becomes detached from the historical records, he divorces the life of the spirit, renounces his responsibility as guardian of society's memory and culture, loses contact with future generations, and gets lost in a myriad of administrative tasks."

16. William J. Maher, "Archives, Archivists, and Society," *American Archivist* 61:2 (Fall 1998), p. 252-265.

are authentic only because they declare themselves to be so, or because they reflect an authenticity that we have projected onto them."[17]

If we use the creation of the National Archives (1934) as a benchmark to date the beginning of the profession in the United States, we are fairly young. Before we could achieve a clear sense of professional identity, external factors are forcing us to re-examine who we are. Electronic records have profoundly changed the nature of our work, and information technologists have staked a claim on what we thought to be our turf. If we are to thrive in this new environment, we must have a clear understanding of what we uniquely bring to the table. It is that expertise and those values that defines our profession.

ARCHIVES

n., –also archive ~ 1. Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value of the information they contain or because of their enduring value as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control. – 2. The division within an organization responsible for maintaining the organization's records of enduring value. – 3. An organization which collects the archives of individuals, families, or other organizations; a collecting repository. – 4. The building (or portion thereof) housing archival collections. – 5. A published collection of scholarly papers.

Notes: In the vernacular, archives is often used to refer to any collection of documents that are old or of historical interest, regardless of how they are organized. Within the professional literature, archives are characterized by an organic nature, growing out of the process of creating and receiving records in the course of the routine activities of the creator (its provenance). In this sense, archives are differentiated from artificial collections.

Many archivists, especially those in the United States who are influenced by the thinking of Theodore Schellenberg, follow an inclusive definition of archives, which encompassed a wide variety of documents and records. Schellenberg also distinguished between the primary and secondary value of the materials; only materials with secondary value, value beyond their original purpose, could be considered archival. For Schellenberg, archivists appraise records for transfer to the archives on the basis of their secondary, research value.

Other archivists follow the writing of Hilary Jenkinson, who argues that archives are 'documents which formed part of an official transaction and were preserved for official reference.' For Jenkinson, the records' creator is responsible for determining which records should be transferred to the archives for preservation. Because Jenkinson emphasized that records are evidence of transactions, he did not recognize any collections of historical documents as archives, although he noted that collections of personal papers were of value to historians because they complemented archives.

'Manuscript repository' is sometimes used to distinguish an organization that collects archives of other organizations from the division of an agency which collects the records of its parent organization.

---

17. Elisabeth Kaplan, "We Are What We Collect, We Collect What We Are: Archives and the Construction of Identity," *American Archivist* 63:1 (Spring/Summer 2000), p. 125-151.

North American archivists sometimes deprecate the use of the form 'archive' as a noun, but that form is common in the United Kingdom. However, the noun 'archive' is commonly used to describe collections of backup data in information technology literature.

One of the questions I heard most frequently is whether archives ends in an 's'. Every standard English dictionary I've checked enters the word under the s-less form, but then notes 'usually archives'. Within the archival community, 'archives' is more common in the United States. Other English-speaking countries tend to use 'archive'; Jenkinson's manual uses this form in its title. One dictionary indicates that the s-less form was common in the United States before the 20th century.[18] Outside the archival community and especially in information technology, 'archive' is more common. This shift is due, no doubt, to the use of the word to mean data that is stored offline or a backup.

Archive is a backformation made by people who, following the convention that plural nouns end in 's', used the form when speaking of one, rather than several, archives. The convention is not absolute; a collection may have several series, but one alone is not described as a serie.

Professionals within the community tend to use 'archives' with the 's' as a mark of identity. Those who use the s-less form are *auslander*. This use of language to establish professional boundaries has a parallel among scientists, who continue to use data with a plural verb, even though in the vernacular it now commonly takes a singular. The 's-less' form is fairly common in recent archival literature, and I suspect it will be the prevalent form within a couple of decades.

To the extent we understand each other, the form is irrelevant. A colleague tells a story of archivists and technologists having trouble understanding each other. Finally, it came out that the technologists were using the word to mean those records that could be deleted from the system, the antithesis of the archival sense of records that must be preserved. The absence or presence of an 's' couldn't have fixed this failure of communication.

More problematic than orthography is the stuff of archives. If we define archives by the character of their contents, we run into differences of tradition. Archivists who follow Jenkinson view their collections as documents which were "drawn up or used in the course of an administrative or executive transaction (whether public or private) of which itself formed a part; and subsequently preserved in their own custody for their own information by the person or persons responsible for that transaction and their legitimate successors."[19] For those who follow Schellenberg, archives were made up of records acquired for their secondary value and may include many documents that fall outside Jenkinson's definition and which are acquired for purposes beyond those Jenkinson envisioned.

The language of electronic records seems to follow the Jenkisonian tradition because it focuses on records that are the result of transactions. Based on personal conversations with several archivists who specialize in electronic records, this emphasis is less likely a philosophical affinity for Jenkinson. Rather, it is because managing transactional records is much easier than other types of records which are not the product of routine transactions.[20] My sense is that a practical, strategic decision to focus on transactional

---

18. *World Book Encyclopedia* (Chicago: A Thorndike-Barnhardt Dictionary published exclusively for Field Enterprises Educational Corp., 1966).

19. Jenkinson, p. 11.

20. The discussion begs the question, What is a transaction?, which further complicates the question.

records has had a side effect of limiting what some archivists and information technologists now consider to be a record. Here we may want to challenge this tendency – as I did with the call to redefine record – to prevent the language, and hence the profession, from evolving away from a body of materials that it has traditionally worked with.

RECORD AND DOCUMENT

§   1. Data or information that has been fixed on some medium; that has content, context, and structure; and that is used as an extension of human memory or accountability. – 2. The information created or received in the course of individual or institutional activity and preserved as evidence of that activity for future reference. – 3. A phonograph record (see sound recording). – 4. (Computing) A row in a database table.

Notes: Records function as an extension of memory. The term carries the connotations of being able to repeat, recite, or recall something. Records may be created specifically to preserve information over time or to prevent future misinterpretation of that information. Records are often stored in a public place to provide constructive notice of the information the contain. However, any document – no matter how ephemeral it was intended to be – may serve as a record if it is used as evidence of the thing to which it refers.

    Records may be in any format, including text, images, or sound. However, a record is ultimately independent of the medium. Paper records may be microfilmed, and electronic records may be transferred from memory to disk to paper.

    Records are often unique, as distinct from publications. However, a record may be a single copy of many that has been selected for preservation or special treatment (a record copy).

§   n. ~ 1. A unit of data or information, especially written or textual, that has been compiled and formatted for a specific purpose that includes content and structure, and may include context. – 2. Information that has been fixed in written form.

Notes: Historically considered to mean unpublished texts on paper. Modern use includes all media and formats, especially those that are (or are perceived to be) two-dimensional. Photographs, drawings, sound recordings, videos, and electronic texts may all be considered documents.

    Often used synonymously with record, although record connotes formal documents most commonly created by an organization or government. However, documents are generally not considered records unless they are created with the specific purpose of preserving information or until they are later used as evidence of the thing to which it refers. Document does not carry records' connotation of being fixed; web pages whose content may change with each viewing have been described as dynamic documents.

    Often used interchangeably with publication, although this use has the sense that there are many copies in distribution. This use is particularly common in federal depository libraries.

    A document's content will, to some degree, reflect formula and convention in its structure, including formal rules of representation, literary style, and specialized language that reflect the author's political, professional, or social cultures. A document's physical characteristics may also follow conventions relating to the medium, organization of internal elements, and presentation of the information.

If archivist is the most contentious word, I suspect that record is the most abused and confused. For the moment, the revised definition remains in flux. I find that as I define words I've used in the definition of record – to think about those words more carefully, I'm forced to revise the definition of record.

The statutory definition of a public record is extraordinarily broad.[21] I would argue that the definition, as written, describes virtually all information held by the government. As a result, those of us who work with governmental information no longer have a word for what we commonly call a record, something that serves as a memorial, gives public notice, or documents some information we wish to remember.

On the flip side, record is often used to mean something more than evidence of a transaction. Part of the problem here is intent. Is a record a document created with the intention of preserving memory? Or, are random documents that have survived by happenstance that we, as archivists, choose to preserve for their historical value at a later date also records? Many people would not consider an album of personal snapshots from the 1920s to be records; they certainly aren't official documents of a transaction, nor do they serve to give public notice. However, it would not be surprising to hear someone refer to those photographs as a 'record of the past'.

Record has a number of senses that complicate the definition. The term serves as shorthand for phonograph albums. Something 'on the record' is something appropriate for attribution.

What I find most fascinating is that a word that is so central to our profession is, in my mind, so poorly understood. I asked a group of archivists to tell me what made a record a record. Two in the group were Fellows of SAA. They stumbled, some of which was a function of their being caught off guard. The exercise came from my own realization that before I began this project I could not clearly and succinctly describe the characteristics of a record, that I could not have answered the question well.

For me, the most distinguishing characteristics of a record are its fixity, its physicality (boundaries and internal, material evidence, including a virtual materiality), and that it bears information that is used as evidence of something, whether intended as such, or not. This list is not complete; I did not mention context, which for some people would be a key aspect of a record. Different people will mention different characteristics; some will overlap, but the lists will – almost certainly – not correlate perfectly.

Document and record are closely related. They share many of the same problems, in large part because so many people use the words synonymously. An advisor pointed out that in much of the archival literature the words are used interchangeably, in part to add some variety rather than using record over and over.

Even though many frequently use the terms as equivalents, many others use the terms in a hierarchical relationship. All records are documents, but not all documents are records. Contrawise, in some uses it's, All documents are records, but not all records are documents. In the first case, a document is not a record unless it has some official capacity as evidence. In the second case, a three dimensional artifact is kept as the record of some activity, but fails the commonsense understanding of a document as information on paper.

---

21. The federal statute serves as the model for many states' legal definition.

I have a hunch that the confusion of document and record results from how we think about the two. To the extent we think about them in terms of their characteristics, they're largely synonymous. However, if we think about how they're used, I perceive a difference. I suspect that it is the way we *use* a document that distinguishes it as a record. A document is information that has been fixed in some manner, usually through writing. When we use that document as an extension of our memory, when we look to it as a window into the past, then the document functions as a record. That use may be intended when the document is created; a record is made consciously. Or, it may happen after the fact when some old document is discovered to have value for the information it preserves. Working from this distinction, it is possible that a document's recordness changes over time. A document may be created to record information for one purpose, say to track accounts for the auditor; once the audit is complete, the document has no further value as an accounting record. However, others may discover that the information in those financial records can be used as a record of many other things.[22]

We want something neat and tidy in a very messy realm, but we are facing the fallacy of the beard. When are a few whiskers a beard? We are dealing with things that may be seen as one thing or another. The trick is to know the essential character of each so that we can understand it as a hybrid.

RELIABILITY AND AUTHENTICITY

§   n., –adj., reliability ~ 1. The quality of being sufficiently accurate and authentic to serve as the basis for a decision or action; worthy of trust. – 2. (Diplomatics) Created with appropriate authority according to established processes and complete in all its formal elements.

–v., rely ~ To have confidence in; to believe in.

Notes: Reliability is a relative concept based on the qualities of authenticity, accuracy, sufficiency, completeness, and integrity. If a record's authenticity cannot be demonstrated, its reliability may be discounted. Similarly, questions about the accuracy, sufficiency, completeness, or integrity reduces a record's reliability.

§   n. ~ 1. The judgment that something is genuine, based on internal and external evidence, including its physical characteristics, structure, content, and context.

–authentic, adj. ~ Genuine; bona fide; an original.

Notes: Authenticity is closely associated with the creator (or creators) of a record. First and foremost, an authentic record must have been created by the individual represented as the creator. The presence of a signature can provide fundamental test for authenticity; the signature identifies the creator and establishes the relationship between the creator and the record.
        Authenticity can also be verified by testing physical and formal characteristics of a record. The ink used to write a document must be contemporaneous with the document's purported date. The style and language of the document must be consistent with other, related documents accepted authentic.
        Authenticity does not automatically imply that the content of a record is reliable.

---

22.   See Christopher Densmore, "Understanding and Using Early Nineteenth Century Account Books," *Midwestern Archivist* 5:1 (1980), p. 5-19. Republished in *Archival Issues* 25:1-2 (2000), p. 77-89.

The authenticity of records and documents is usually presumed, rather than needing to be affirmed. U.S. judicial rules of evidence stipulate that to be presumed authentic records and documents must be created in the "regular practice" of business and that there be no overt reason to suspect the trustworthiness of the record (Uniform Rules of Evidence, as approved July 1999).

If any words will be my downfall, it may be authenticity and reliability. As I noted above, they are often used as meaningless buzzwords or cant. These words are particularly difficult because they are closely related and often defined in terms of each other. I'm not the only one who has found these words difficult. Eun G. Park noted that "The language used by practitioners to express issues of authenticity differs significantly from the language used by the most prominent research projects." She observed that practitioners tend to understand authenticity as a concept related to accuracy, originality, and verification but that that researchers studying electronic records understood authenticity in terms of evidence, reliability, genuineness, warrant, integrity, and auditability.[23]

Because the underlying concepts are fundamental to recordkeeping, it's particularly important to have a clear understanding of what they mean. Struggling to parse these two key aspects of recordness led me to think through many other aspects of recordness.

Although a great deal of attention has been given to authenticity, I believe that reliability may be the cornerstone of recordness. Authenticity is only one factor that influences reliability, and I think it may be over emphasized. Is the record complete? Is it sufficient? Is it accurate? Has it changed? How would one know?

A record of a confession may be unquestionably authentic. But does it have all the elements to be able to demonstrate its authenticity (a signature, a date, a witness)? It may be complete – I killed Cock Robin. Does it have enough information to convict? Could additional information such as DNA evidence discount the confession? A photograph made by a surveillance camera showing that the murder was actually self-defense or that the confession was made under duress? Is the confession accurate? Would a prosecutor use my confession if I claimed to have used poison, but the victim died of a gun shot?

We have an intuitive understanding of these concepts based on our work with paper records. We grew up around paper records, and like language, we have a casual sense of their meaning. For most purposes, we can spot problems with paper records. We rely on physical clues, such as erasures, inks, papers, type, or handwriting. We rely on our own knowledge of the facts. Because electronic records are so different, we can no longer trust our casual knowledge. We need to think critically about the nature of records so that we can learn new ways to determine when records are reliable in an electronic environment. I've learned a great deal about records and recordkeeping by studying this new form of electronic records in much the same way that I learned English grammar by studying German.

---

23. "Understanding 'Authenticity' in Records and Information Management: Analyzing Practitioner Constructs," *American Archivist* 64:2 (Fall/Winter 2001), p. 270-291.

n. ~ Data describing or documenting the management, nature, or use of information resources.

Notes: Metadata is typically organized into distinct, related categories and relies on conventions to establish the values for each category. Metadata is typically used to locate or manage information resources by abstracting or classifying those resources, or by capturing information not inherent in the resource. For example, administrative metadata may include the date and source of acquisition, disposal date, and disposal method. Descriptive metadata may include information about the content and form of the materials. Preservation metadata may record activities to protect or extend the life, such as reformatting. Structural metadata may indicate the interrelationships between discrete information resources, such as page numbers.

In terms of archival collections, MARC format and EAD are metadata standards. Dublin Core is another metadata standard intended for description of Web resources.

In terms of information technology, metadata includes the documentation of data architecture, properties, and methods necessary to store, retrieve, and use the data in a meaningful manner. To the extent that data is a record, it may also include administrative, descriptive, preservation, and structural information.

Metadata is not really a hard word. It has, in my opinion, become very nearly a useless term. Not that there isn't need for a term that describes information about information. Rather, people have used the poor word so many different ways that it is now effectively meaningless. At a recent conference, I heard several people use metadata as a verb. I have no clue what 'to metadata' means. Ken Thibodeau mentioned to me that "Metadata used to have a perfectly nice meaning in engineering, and that information technologists' appropriation of the term 'ontology' was payback for librarians' corruption of 'metadata'."[24]

I see metadata much the same way that Handy saw post-modernism. People don't want to use the word 'cataloging' because it connotes sensible shoes and a bun, drawers of three-by-five cards, and a compulsive attention to lots and lots of rules. Metadata identifies users with the new world of the information economy; metadata is neither dusty or musty. There is some truth in the aphorism, 'Metadata is cataloging for men.'

What I find frustrating about the abuse of metadata is a desire to gloss over complex concepts. What troubles me is that people tend to disregard differences that are important. For some time, many within the library community tended to use metadata synonymously with Dublin Core, as though the only information about our collections we needed to track was descriptive information. I was pleased to see that OAIS described many different kinds metadata relating to other functions.

---

24.   Personal correspondence.

The Future of Words

In the end, words are slippery. Meaning is elusive. Language is clear only when there is nothing to argue about. It's a vicious circle. As we parley ideas, our understanding of the ideas changes, and as the ideas change, so does the meaning of the words we use to represent those ideas.

Writing the glossary has helped me understand and engage in the larger dialog of the archival profession. I've been forced to think about the words more carefully. The process has encouraged me to think, in Frank Burke's terms, more theoretically about the profession.

John Roberts summarized archivy in eleven words, "We save what is historically valuable–*there*; that is the theory."[25] Where Roberts sees simplicity, I see confusion and complexity. Roberts sees reflection as so much useless navel gazing. I see lots of questions surrounding almost every word in his statement. The way I understand those questions will have an immediate, practical impact on how I do my job. Notice that I said "understand those questions." I don't think the questions have answers, per se. Like words and definitions, questions and our understanding of them will change over time.

›   The first question Roberts' statement raises is, Who is 'we'? In context, it's clear that he is referring to archivists. In the age of electronic records, it will take more than archivists to preserve records. We must partner with information technologists and others, and those collaborations will blur the boundaries of who is an archivist.

›   'Save' raises the question of how the process of saving records has changed. Because the process of migrating records to avoid technological obsolescence often results in changes to the records, we must think about how records' reliability is affected.

›   'What' refers to the object of this process. Dynamic information systems, hyperdocuments, and virtual reality challenge our understanding of recordness.

›   Roberts' final phrase "historically valuable" begs any number of questions. What is valuable? To whom? For what purpose?

Roberts' statement becomes more complex when we consider not *what* the words mean, but the *why* underlying the words.

I believe that the fundamental function and essential characteristics of a record remains the same. People have needed a way to fix memory for future reference for centuries, and that need has not gone away. We must come to understand how people are using new forms of records. We must also consider how the characteristics of those new forms map to established characteristics and the significance of new characteristics.

The recordkeeping process has changed over the centuries, and we must come to some understanding of the impact of current changes in that process. In a paper environment, paper records were an unintended

---

25.   "Archival Theory: Much Ado about Shelving," *American Archivist* 50:1 (Winter 1987), p. 66-74.

byproduct of other activities; records just happened. At the University of Texas, Harry Ransom built his reputation by collecting the many drafts of authors' works; he saw those drafts as evidence of the creative process, giving a richer understanding of the final work. In the digital environment, records don't just happen. With word processing, many – if not most – of the drafts Ransom would have wanted are lost. Each time a document is opened, revised, then saved, the previous versions disappear. Few database systems are designed to be able to rollback data so that it is possible to see the state of the data at any given point in the past; data is added, changed, and deleted, with no thought to preserving older data for future reference. The challenge of paper documents is an excess of irrelevant memory captured in piles of paper. The challenge of electronic records is incremental amnesia.

As we come to a more complete understanding of records and recordkeeping, we return to the thread of professional identity that runs through this paper. How will our understanding of archivist change?

I believe our job will continue to involve the traditional functions of appraisal, acquisition, processing, preservation, and reference. I would add to that list the need to work with records creators to ensure records' reliability.

While our work remains fundamentally the same, I believe – and this is not a novel concept – that our character must change. We have often been seen as custodians of the past. We are seen as the keepers of old things. While we will remain the custodians of old records, we must change our emphasis from the preservation of culture as a thing of the past. We must become advocates for future users of current information by ensuring the transmission of culture. What of the present needs to be remembered for future users? I believe that our knowledge of what has enduring value and how researchers use materials is the thing we do best. It is that knowledge that can enable us to help records creators know when to commit intentional acts of memory, to know what to save. And, to do that, we must be able to speak clearly.