TELLING THE DIFFERENCE: PERCEPTIONS OF UNIQUENESS AND IDENTITY IN ARCHIVES

Geoffrey Yeo

This is a revised and shortened version of a keynote paper presented at the XVI Congresso Brasileiro de Arquivologia in Santos, São Paulo, Brazil, in August 2010.

This paper seeks to examine archivists' understandings of the concept of uniqueness. One of the things we know – or think that we know – about records and archives is that they are, in some sense, unique. This characteristic distinguishes them from printed library materials, which exist in multiple copies. And much of our practice is driven by this understanding: the reason that preservation and security have a higher profile in records management and archival work than they usually have in libraries, is at least partly because we feel that our holdings are unique, and that if they are lost or damaged we cannot simply go to the bookstore and buy replacements. Our descriptive practices are much more labour intensive than the cataloguing practices of librarians, who can buy catalogue entries ready made from a centralised cataloguing service. We cannot do this; instead we have to construct descriptions ourselves, often very laboriously; and the explanation for this, we argue, is that our holdings are unique and descriptions of them cannot be compiled by a consortium and then resold and reused. And I would suggest that all this is largely correct; in the pre-digital world, the world of paper records, our holdings and our priorities differ from those of librarians in precisely these ways.

The idea of uniqueness is intimately connected with the belief that, at least in certain fields of endeavour, there can be only *one* original of a given work. The archetype of the unique original is the painting; only the original is the work of (for example) Rembrandt, and it stands apart from any copies, whether forgeries or 'legitimate' reproductions. Archivists have traditionally perceived archives and records in this way; there can only be one original of any record, it is unique, and everything else is a copy.

No-one would have disputed this in the days when all records had to be written by hand. But the invention of printing techniques and mechanical replication devices long ago led archivists to wonder whether they perhaps needed to reconsider this idea.

Many archivists have accepted that items produced by a printing press can sometimes have the status of records, and have found themselves driven to consider the possibility that individual records need not be unique. And if we were driven to this re-examination because of the invention of printing 500 years ago, the invention of modern copying devices such as the duplicating machine or the photocopier must lead us to question traditional notions of uniqueness even more forcefully.

In 1992 Terry Eastwood wrote that 'the uniqueness of archives derives from the place each document has in the structure of the whole, from the fact of its position in relation to other documents'. Eastwood argued that each document is not necessarily the only copy in existence, but that it is unique in its context, unique by virtue of its relationship to other documents in the series or fonds in which it is located. The idea of uniqueness in context has been widely adopted by archivists and writers about archives. Even if individual documents (individual records at item level) may not be unique, the file, the series and the fonds are assumed to be unique. No two series, no two fonds are supposed to be the same.[1]

But, more recently, some have claimed that the computer revolution requires us to rethink our ideas once again; that uniqueness *of any kind* is impossible in the digital world. Firstly, in the world of computers we encounter multiple instantiations on a scale previously unparalleled. Consider the transmission of e-mail. When we send an e-mail message, instantiations of it are created more or less automatically at various locations as the message travels from its sender to its recipient. A message is likely to be sent from one mail routing server to another and then to another any number of times before reaching its destination, and each routing server makes a copy of it. A typical e-mail message is replicated many, many times; and often neither the sender nor the recipient knows how many such replications are created or on which computers they are held.

---

[1] It can also be argued that these understandings affect, or should affect, archivists' attitudes to the disposal of alleged "duplicates"; i.e., that a printed or photocopied document in one series cannot be seen merely as a redundant duplicate of an apparently identical document in another series, because it has a different documentary context.

The rules also appear to change at series or fonds level in the digital world. If we have a series, or a fonds, that consists entirely of digital objects, we can duplicate the entire series, or the entire fonds, perhaps with no more effort than a couple of clicks of a mouse. When we have done this, we appear to have two instantiations of the same series, or the same fonds, each apparently indistinguishable from the other. What then happens to our claim that item-level records are unique in the context of their documentary relationships? If two identical records can have identical documentary contexts, if each can be located in what appears to be an identical series, does this mean that we have to abandon our ideas of uniqueness altogether?

I want to suggest that, in order to understand the notion of uniqueness, we need to examine the concept of identity, or sameness. What does it mean to claim that two objects are identical? Can we ever claim with any degree of confidence that one item-level record, or one series, is the same as another? This paper is entitled 'telling the difference', and I want to explore what it might mean to tell the difference between things that might or might not be the same.

In the 1970s, the philosopher Nelson Goodman asked whether performances of Beethoven's Fifth Symphony with a single wrong note can still count as performances of Beethoven's work. If they do, how many wrong notes can we have before a performance ceases to be the Fifth Symphony? Goodman believed that even a single wrong note disqualifies it. And he said that works of literature also operate like this; that we can have two genuine instances of the same novel or the same play, provided the *words* are exactly the same ones.

But what if the words are similar, yet not precisely identical? What if we have two *versions* of a novel or a play? If we follow Goodman's argument, must we deny *any* kind of sameness between them?

What if we have the holograph manuscript written by the novelist or playwright and also an exact transcript made by a later copyist? In this case, the words *are* the same, but one instantiation of the text was created before the other. One was created by a writer's inspiration, the other by means of more or less mechanical copying.

Or what if we have two copies of the transcript? Are they the same as each other? Their wording and their circumstances of creation are perhaps identical, but each transcript occupies a different physical space. One of them was on my desk on a particular day in the year 2010, the other was not. So, in this instance, what we might call their *history of use* is different. Does this mean the two transcripts are not the same as each other?

These are not easy questions to answer. There is a long scholarly debate about identity and sameness. The Greek philosopher Heraclitus asked how he could say that the *river* he bathes in today is the same river he bathed in yesterday, given that yesterday's water has now flowed on towards the sea. Much more recently, the Swiss linguist Ferdinand de Saussure wrote about the *train* that leaves Geneva for Paris every evening; how is it that, although the rolling stock and the crew are different, we are likely to say that the 'same' train leaves each evening?

Even in the case of so-called *'identical' twins*, although each of them will have many individual characteristics, we know that many people are likely to see their differences as insignificant. The question is: what does it mean to perceive a form of identity between two or more objects that are arguably distinct?

I want to suggest that a large part of the answer lies in our use of the word (*in*)*significant*. Or, if we prefer, (*un*)*important*. We often consider that some of the differences between two objects are significant or important, and that others are trivial.

Many archivists will be familiar with Thomas Jefferson's polygraph machine that could be used to create two 'identical' records in a single act of writing. It can be seen at Jefferson's house, Monticello, in Virginia, where it is on display. Jefferson invented this machine so that he could write a letter and create two copies of it simultaneously.

Now, strictly speaking, even two letters created using Jefferson's machine are not identical, since the molecules of paper and ink are different; from this perspective each letter can be considered unique. However, if their distinct material properties

are considered unimportant, the two letters may perhaps be treated as interchangeable.

Let me give another example. When English kings in the medieval era wanted to write to rulers in other countries, they sometimes sent two 'identical' messages. Each instantiation of the king's message was sent by a different route in case the dangers of medieval road journeys prevented one of them from reaching its addressee. Both copies of the message served the same purpose; both bore the king's seal; both were written on the same date and contained the same text. And yet each copy was written on a separate sheet of parchment; one copy travelled by one route, the other by a totally different route. Are they truly identical? If we think that only the dating and the content of the message is significant, we can say that the two messages are identical, and that neither is unique. But if we think that their physical separateness is significant, if we think it is important that each message travelled by a different route, then we find ourselves obliged to say that in some sense each is a unique message.

I would now like to explore how these issues operate in the digital world. We are often told that the digital revolution changes all the rules, and that concepts we applied in the paper world have been shown to be outmoded by the march of technology. Many scholars and computer scientists claim that the world of computers allows us to create multiple instantiations of documents that are exactly identical. They argue that I can create a document on my computer, and if I send it to you, then your computer has a copy that is in every respect the same as mine. And I can also make back-up copies on a portable disc, I can copy the document from one disc to another, and every copy is identical to every other. The idea of a unique document – or so many commentators have told us – is all but dead in the digital environment.

But I would like to suggest that, in fact, the computer world is no different in this respect to the world of paper. Once again, it is a question of what we think is significant. If we think that only the *content* or the *structure* of the document are significant, then we can say that all the copies are identical, that none of them is unique.

5

But we must not assume that everyone's estimate of what is important will be limited to content or structure. Take my example of the e-mail message. Some people may consider the copy of the message retained by the sender to be identical to the copy received by the addressee, and to the copies retained by the various mail servers en route. But others may insist on their difference, even though all the copies represent the same activity, the same communicative event, and their textual content is indistinguishable. They are representations of the same activity but they are not the same representation. Each copy of the e-mail is physically separate and has its own contexts of creation, maintenance and use. Each serves a different purpose, and each will carry slightly different metadata, because every intermediate e-mail server adds a small amount of extra metadata when it retransmits the message. Moreover, each copy may be rendered differently when we view it on different occasions, depending on the technical configuration we use to view it.

What I am talking about here is closely connected with what literary scholars such as Matthew Kirschenbaum call 'digital materiality': the notion that even digital objects have a physical existence. Explorers of the digital world often seem reluctant to allow any space for consideration of physical entities; their notions of virtuality appear to suggest that digital objects are immaterial and have no physical status. This is what Kirschenbaum calls 'the tactile fallacy': a presumption that such objects must lack materiality because 'we cannot reach out and touch them'.

Of course, digital objects can easily be moved from one storage device to another, but at any one moment, the bits and bytes that comprise a digital object are somewhere real and physical; they exist on a particular storage device at a particular point in time. The same can be said of digital *collections*: transient though such collections may sometimes be, at any given moment their components must have a material existence.

Digital materiality often seems counterintuitive, because the computer environment is engineered to present an illusion that digital objects are *not* material. But if one accepts their materiality, it follows that, even in digital environments, no two objects are exactly alike; none is identical to any other. There are always differences between objects, however small those differences may be.

The set of electrical signals that I have on my computer is necessarily discrete from the set that you have on yours. One set of signals was generated or copied at a certain date and was stored on a particular medium at that time, and this set of signals was used by me last week; another set may present very similar content when viewed on screen, but was generated or copied at a different date and was used by you this morning; a third set was perhaps stored somewhere else but has not been used by anyone at all.

The principle of individualization – the principle that each object is different – underlies the discipline of *digital forensics*, which uses the knowledge that no two objects are ever exactly alike to uncover the distinct histories of individual digital objects with the aim of detecting computer crime. I suggest that we need to use this same knowledge in our own discipline, and to be sensitive to the uniqueness of digital documents at item level.

And I believe that the same applies to the series or fonds. Even if we can have multiple instantiations of a digital series or a fonds that are indistinguishable in terms of content or structure, each of these instantiations has its own materiality and its own contexts.

Earlier in this paper I mentioned the philosopher Nelson Goodman. Goodman made a distinction between what he called 'autographic' and 'allographic' works. Autographic works are works such as paintings, where we can identify originals and distinguish them from reproductions or forgeries. Allographic works include novels, poems and symphonies, which (if we accept Goodman's argument) can be repeated as often as we wish. To quote Kirschenbaum once more: 'A copy of the *Mona Lisa* is just that – a copy of an acknowledged original – while a copy of Mary Shelley's novel *Frankenstein* is a perfectly valid way of experiencing the work (you don't have to ... sit down with the holograph manuscript to ... claim to have read *Frankenstein*).' In Goodman's view, each printing of the novel or each performance of the symphony gives us a genuine instance of the work, and novels and symphonies cannot be forged in the way that paintings can.

But Goodman's analysis is open to debate. Where so-called allographic works are concerned, we still need criteria on which to base our judgement that one performance of a symphony is as genuine as another. To put it another way, we need criteria for determining the identity (the sameness) of two performances that purport to be identical instances of the symphony in question. To resolve this, Goodman claimed that what he called 'properties constitutive of the work' can be distinguished from properties that are merely contingent. In the case of symphonies, he argued that strict adherence to the composer's score is the property that guarantees a 'genuine instance' and, as I noted earlier, that even a single wrong note disqualifies it. But many critics have questioned this argument and have disputed his assumption that other properties, such as the context in which a performance takes place, can always be dismissed as irrelevant.

Goodman's concern was with literature, music, paintings, and other works of art, but comparable issues arise with records. In a recent article, Heather MacNeil and Bonnie Mak suggested that most digital records might be perceived as allographic. But, again, it depends on what characteristics of records we think important or significant. Archivists are likely to believe that their *context* is important as well as their content and structure. Each instantiation of a digital resource has its own contexts and, in this sense, each has an autographic dimension that cannot easily be denied.

In this way we can rediscover the belief that each record is unique. It is not merely documentary context that distinguishes one resource from another, as earlier writers sometimes claimed, but contexts of production, contexts of maintenance, and contexts of use. In the world of digital records we do not have to forego the idea that no object or set of objects is identical to another. We may need to rethink our understanding of the concept of uniqueness, but we certainly should not abandon it.

Bibliography

Eastwood, T, 'Towards a social theory of appraisal'. In: Craig B L (ed.) *The archival imagination: essays in honour of Hugh A. Taylor*. Association of Canadian Archivists, Ottawa (1992), pp.71-89

Goodman, N, *Languages of art*, 2nd edn. Hackett Publishing Co., Indianapolis (1976)

Kirschenbaum, M, 'Editing the interface: textual studies and first generation electronic objects'. *Text: An Interdisciplinary Annual of Textual Studies* 14 (2002), pp.15-51

Kirschenbaum, M, *Mechanisms: new media and the forensic imagination.* MIT Press, Cambridge, MA (2008)

MacNeil, H, and Mak, B, 'Constructions of authenticity'. *Library Trends* 56 (2007), pp.26-52

O'Toole, J M, 'On the idea of uniqueness'. *American Archivist* 57 (1994), pp.632-58

Paskin, N, 'On making and identifying a "copy"'. *D-Lib Magazine* 9 no.1 (2003), at <http://www.dlib.org/dlib/january03/paskin/01paskin.html>

Schwartz, H, *The culture of the copy*. Zone Books, New York (1996)

Yeo, G, '"Nothing is the same as something else": significant properties and notions of identity and originality'. *Archival Science* 10 (2010), pp.85-116