Order As Received: A Foundational Virtual Order for Digital Records

Draft: AERI 2010

Patricia Galloway
School of Information, University of Texas at Austin

Abstract

Archivists pride themselves on the detection and restoration of the "original order" of the archival fonds, which represents most of the time-honored labor of archival arrangement. Digital records' affordances, however, mean that even in active use they may simultaneously lend themselves to multiple virtual orderings, none of them representing an actual physical ordering of the records on any medium. In this presentation I want to discuss experiments in archiving digital records in a formal "order as received" based upon groupings of digital files on received legacy media prepared by the donor and documented through description of the set of derivative orderings available through the original operating system environment. This practice is designed to capture and describe a specific documented state of the fonds, to provide to the potential user a representation of a stage in archival processing that is normally invisible, and to create a documented basis for other derivative orderings, including those imposed by archivists and even by individual researchers and users. I want to suggest that this kind of archival practice gets us closer to the ability to portray what is really meant by "context" in the digital environment and even to documenting the work practices of records creators and archivists that normally remain tacit.

"Original order" in archival practice

It is a fundamental article of faith among Western archivists that materials preserved in an archives must be kept in provenance-based groups representing a *fonds* and where possible should be preserved in what is referred to as "original order." According to Pearce-Moses' *Glossary*,[1] "original order" is defined as:

> The organization and sequence of records established by the creator of the records....Original order is a fundamental principle of archives. Maintaining records in original order serves two purposes. First, it preserves existing relationships and evidential significance that can be inferred from the context of the records. Second, it exploits the record creator's mechanisms to access the records, saving the archives the work of creating new access tools.

In fact, however, most bodies of documentary materials do not arrive at the archives in some perfect "original order," or at best they arrive in the order in which they were last found, which may reflect any number of events that have happened to the materials since they were created, received an arrangement, were perhaps rearranged in use, and finally were sent to the archives. This is especially true in the case of privately-created personal

---

[1] Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (SAA, 2005). Accessed from http://www.archivists.org/glossary/index.asp All citations to this work are reachable from the digital version by term.

documentation normally subject to relatively informal organizational practices by its creator.

For some reason archivists have resisted considering the actual order in which they first see the records as the original order. Instead they have taken on the responsibility of obliterating this unwanted "messy" order and *restoring* "original order," acting in accord with any perceptible filing practices to restore the assumed original order and to rationalize any minor departures from it, while fitting series of filed materials into an archivally-created hierarchical series structure based at best upon organizational charts which themselves are idealized and may never have represented a true arrangement of anything. From Pearce-Moses:

> Original order is not the same as the order in which materials were received. Items that were clearly misfiled may be refiled in their proper location. Materials may have had their original order disturbed, often during inactive use, before transfer to the archives; see **restoration of original order**.

Additionally, unfortunately, no record is created of precisely what changes were made to which records in the reordering, so it is impossible for the user to restore the "original chaos" (see below) and thereby understand more about the work practices represented by it. In the case of privately created records, special collections archivists have frequently taken more drastic measures, since such records frequently come to them upon the death of the creator and much the worse for confusion in such packing-up as they may have been subjected to by relatives. In such a case the records are arranged for the convenience of the user and often according to local and idiosyncratic categories. In arriving at this order in normal processing practice, it is also customary to "weed" duplicates and other unwanted categories of materials while this arrangement is taking place. The result is not presented precisely as original order, but nor is any public statement usually made as to what exactly has been done. The *Glossary* observes rather more acerbically:

> A collection may not have meaningful order if the creator stored items in a haphazard fashion. In such instances, archivists often impose order on the materials to facilitate arrangement and description. The principle of respect for original order does not extend to respect for original chaos.

Yet every archivist who has worked with a "disordered" collection knows that it is nearly impossible for the disorder to extend to true randomness or complete chaos: disordered paper materials are after all not shuffled as playing cards may be—in fact cannot be so shuffled because their physical characteristics do not lend themselves to such action. Instead, there are likely to be many loci of order within such a collection. Further, if the creator was a messy filer, why should we presume to turn him into a neat one?[2] These archival norms, I suggest, are the result of serving two masters—archival convenience

---

[2] Thomas Tanselle brought a needed correction to historical editing practice in the 1970s when he reminded people that correcting George Washington's spelling did nothing for the authenticity of a historical edition of his letters. See Heather MacNeil's review of the evolution of literary and historical editorial practice in "Picking our Text: Archival Description, Authenticity, and the Archivist as Editor," *American Archivist* 68(2), 2005, 264-278.

and least-common-denominator user requirements—while ignoring the needs of another: researchers who want to approach the records with as little intermediation as possible.

Order and the affordances of digital records

I would like to suggest that especially in the case of digital materials we should not and indeed are not required to take such drastic measures: we can have it all three ways. The affordances of digital records problematize the paper-based notion of order in the first place. A digital record is itself generally a congeries of fragments dictated by an underlying storage scheme optimized for efficiency of access; the records assembled and presented to the user as a whole by the operating system and/or application program or viewer.[3] Groups of digital records arranged in hierarchical directories are also a construct of the operating system in obedience to the user's choice of representation or to some default assignment of location; the records themselves are not only fragmentary as just noted, but are also not "arranged" on a medium as represented by the directory. Any ordering of digital records, original or otherwise, is a *representation* rather than a physical order (and frequently, given the affordances of the system through which the ordering is viewed, may be only one of several views, as by title, date, filename, etc.), yet representations are the only lenses through which any ordering is perceived in normal use.

Research being carried out currently in several places will probably shortly establish as standard the use of digital forensic techniques to view more directly the distribution of magnetic fluxes on a magnetic medium, so that the archivist can see the physical distribution both of the files shown on a directory listing and of wholly or partially "erased" files whose space has been marked as free. This will permit an even fuller view of the creator's work processes, since during early use of any magnetic medium it is rarely cheap enough not to reuse and few users to date have known that patterns of former use remain on the medium partly written over by more recent files, or have cared enough to erase the medium more thoroughly through reformatting, which is usually time-consuming. It will be interesting to see what effect this has on recovering "order." This order is at a remove from the user's grasp of it and there are discussions still to be had about the ethics of recovering information the creator did not intend be seen. I think it likely that from these considerations will emerge a more complex donor agreement addressing these affordances and the archives' right to access to them that will begin to resemble informed consent agreements. As a practical matter, should the creator agree, it will most likely become standard to archive a disk image of the mediium in question along with the manifest materials copied forensically from the medium and a specification of their order.

Receipt of digital files by an archives

We must then consider how digital records are received for preservation, whether on removable storage media or devices or still stored on an integral disk internal to the computer itself. In the former case, storage media or devices may represent backup activities carried out by the creator or may have been prepared specifically for

---

[3] This is true of now dominant random-access media (magnetic or optical disks or flash memory units) but is not generally true of formerly dominant sequential tape media.

transmission to the archives. In the latter case we may consider that the records are most likely to be in some kind of "untouched" or original state, providing a contextual snapshot of multiple activities in various stages of completion, especially if we know the circumstances under which the computer became inactive before coming into posession of the archives. There is one further option, not yet much in evidence though likely to become more common over time, in which the record creator has specifically deposited individual records at or near the time of creation, whether through some kind of record management application designed to effect regular capture and storage or through a specific action of deposit; in this case, the order found will probably be some predetermined order into which the materials are made to fit, and it may or may not have been designed through consultation with the creator.[4]

Backup media may have been created using specific backup utility programs and can represent the whole contents of a computer or some specified subset. Quite frequently, however, backup media may represent a creator's specific work practices of versioning, as for example when she makes use of removable storage to perform periodic backup of the different versions of a particular document undergoing intensive work. Some computer users are conscientious in backing up at regular temporal intervals, whereas others will habitually back up groups of files undergoing active work at smaller intervals—usually representing the amount of work they would rather not lose.[5]

As we all know, however, backup media may be prepared especially for archival deposit, whether by the donor, others acting for the donor (including dealers, assistants, or relatives), or acquiring repository archivists working with the donor, and all of these groups represent different motivations that may be manifest in the order imposed on the media content.[6] In all of these cases the order as received undoubtedly has meaning (which may be manifest, for example, in the contrast between the dates listed for the individual items by the operating system as contrasted with the dates of those same items as made available in the internal metadata of the individual files), even though archivists may not be interested in the meaning overlaid on the materials by what may be seen as third parties. This meaning, furthermore, may in fact capture the creator's original order in some way, depending on how and when it was carried out. In any case, loss of any of these orderings loses one state of the archival bond, one view of the records' relationships, and a potential opportunity for contrasting this ordering with other related orderings (as e.g. contrasting an automated backup of an entire computer with the actual computer's contents).

An alternative approach for arranging digital materials

Considering that the order as received of digital objects represents a more or less intentional order, it is clear that in preserving it we are preserving something of the work

---

[4] So far the only self-deposit that is widely solicited by archives is the deposit of materials by academics into institutional repositories.

[5] Most recently, users may safeguard versions by emailing them to an external mailbox that may be captured elsewhere.

[6] Cite Lucie Paquette, the Paradigm project, and the dealers at the 2008 Flair symposium. It would be interesting to ascertain whether creators whose collections are sold to collecting repositories keep copies of their deposited materials for themselves, especially considering the fact that as yet few if any collecting archives accord creators online access to their own materials.

practices of the creator or of people around him, and as the media in question serve as the interface between the creator and the archives, they also represent at least part of the transfer process. If this aspect of archival materials is of interest, the affordances discussed above provide a workable possibility for us to preserve it and at the same time to expose the professional work of archivists directly: rather than devising some kind of archivist-created arrangement, we can ingest digital materials into a digital archival repository just as they come to us, according to whatever directory tree may have been on the medium used to transfer the materials to the archives, creating a *virtual* "order as received" that does not make any changes to whatever relationships may have existed among the files and persisted on the transfer medium, and not presuming that there is no meaning at all in such an order.[7] Since users can in any case potentially search through all these materials without reference to any ordering, or can use digital tools to order materials according to date, author, or any number of other available attributes, one may adhere to the ideal of "more product less process" and do no more "arranging" at all, thereby preserving one more step in the lives of the records received.[8]

But in addition, if time permits, the archivist may add value by creating a virtual ordering adherent to local archival series arrangements by mapping digital objects from the "order as received" representation into a "local archival order" representation. Further, where living authors may be interested in becoming involved with arrangement, an "authorially reconsidered order" representation may also be constructed; and other virtual orderings suggested by users might also be shared.[9] The point is that in the digital environment, where multiple virtual orderings can be made available, and where there can be great interest in seeing any orderings that represent some part of the process by which the materials reached the archives and indeed by which they *became* archival (especially on the chance that they may reflect the ordering through which the materials were used by their creator or primary user), I think it is finally advisable to abandon the too frequently silent emendation by archivists of "order as received" and the recognition of such an order as not just a troublesome mess to be easily swept away but as a source of information that should not be discarded.

Implementing order as received in the digital repository

There are several ways to deal with digital records' order as received. The closest one can come to preserving the "actual" order, the frequently fragmented order as it actually exists on the medium as distributed by the native operating system, is to capture a disk image or clone of the medium via a bitwise copy captured without creating any metadata recording the fact of capture; this is about as close to the archival concept of "fixed to a medium" as one can get in the digital context. A disk image permits the re-representation of the file arrangement according to all the affordances of the native operating system when mounted as a disk in that environment—to all archival intents and

---

[7] Note that in the world of personal computers, floppy disk storage (and indeed hard-disk storage) did not provide hierarchical directory structures under early disk operating systems, including PDDOS/MSDOS.

[8] Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68(2), 2005: 208-263.

[9] See Yakel et al.

purposes it is the same disk.[10] Further, files captured in this way are true clones, identical to the files on the original media (even in being scattered throughout the image), instead of copies of individual files created within an operating system, which is designed to add new time-stamp metadata when the copy function is used. It may for some purposes continue to be acceptable to use non-proprietary "archiving" utilities like cpio and tar in the UNIX environment, which are designed to preserve overt files without change and with the data needed to restore their relationships in directories and their permissions and other behavioral characteristics within the native system.

Archivists are now discussing the use of disk imaging as the gold standard for file capture, both for the opportunity to learn more about the creator's work practices and for the ability to obtain clones to transfer to more modern media. There is an aspect of this practice that raises questions now under discussion by archivists, which is that a disk image also captures fragments of files marked as "erased" with reference to the native operating system but not yet overwritten (covert files), and to date donor agreements have not provided for the informed consent of donors to the viewing of any covert files that may be on the transfer media.

Alternatively one could simply archive the entire overt contents of the medium as one bundled object that can be unpacked by appropriate software; in the UNIX environment, it has long been the practice to use the cpio and tar utilities to bundle and unbundle contents as ordered by directory trees, also preserving ownership and access permissions, last modification datestamp, and size—in other words, things about the files known to the system. Outputs of both of these open utilities can be compressed for storage without loss of data.

Once archived, however, the question arises of how to represent the captured files and relationships for the archive user. To achieve a representation of order as received, there are at present a couple of options. The simplest is to create a collection for each media unit, to contain the image or file bundle plus the individual files extracted from it for convenience of user access. This would allow a user able to mount or unbundle the image or bundle in the relevant environment to access, with the relevant system tools, anything recovered from the media unit. But where the user only wanted to access individual files, that would be straightforward. In the case of hierarchival directory trees where the user wanted to see some kind of schematic representation of the relationships among the files, overt files obtained from the image or bundle could be mapped to a virtual structure that would present them in the archival environment in something approximating their relationships on the original media unit. Interestingly, once one goes beyond permitting the user to interact with the image or bundle directly, we are back to discussing the degree of distance the potential user is likely to want.

---

[10] It is possible to go further by making a copy of the magnetic flux patterns on a disk, so that the underlying formatting of the original disk is captured, but it is then necessary to recover the information known to the native operating system using statistical software. This kind of imaging is done in order to recover unknown disk formats so that an emulated controller can extract the disk image. It may also be used in the adversarial forensic environment to seek evidence that may have been hidden purposefully.

Summary: Copy methods and ordering representations
- disk image cloning (dd software on UNIX): preserves bit patterns exactly as found on media; requires additional processing to extract files singly or recover directory structure, so additional utility programs are needed (e.g. disktype on UNIX).
- cpio or tar copy: preserves overt files only but also preserves system-created metadata about files, including directory structure; individual files can be extracted directly from the cpio or tar "archive" using the same software that created it, and some have argued that such an archive, supplied with a listing of its contents as metadata, is an adequate representation and access format.
- ordinary copy: preserves overt files but changes system-created metadata about files, recording the fact of copying

When disk image cloning or cpio/tar copy is used, the single file obtained should be processed to create fixity data and archived before individual files are extracted. It is important to consider the native operating system and to preserve somehow (as a listing, for example) the representation of the file system within it as seen by default within the native operating system. In addition, if the native operating system permits simple reorderings of file listings (alphabetically, by date order, etc.), these affordances need to be noted in the documentation of the archival materials from this environment. This represents another reason to keep digital materials in the context-related order as received.