

## Authentication of Digital Objects Lessons from a Historian's Research

---

CHARLES T. CULLEN

The issues stemming from authenticating digital objects are quite similar, and in some cases identical, to those relating to holographs or printed books. Everyone dealing with important material in any form should approach it with a bit of skepticism, but scholars especially need to question what it is they are using. In other words, they need to authenticate all documentation they use in the processes of learning and of creating new scholarship. An authentic object is one whose integrity is intact—one that is and can be proven or accepted to be what its owners say it is. It matters little whether the object is handwritten, printed, or in digital form.

Over time, we have established various measures of authenticity for analog forms that we trust almost without question. Our trust is, however, much greater for printed books than for handwritten objects. In fact, handwritten objects raise many of the same questions of authenticity as digital objects do. The difference is that in the case of the former, the answers may be more easily found. Take Thomas Jefferson's manuscript «Report on the Navigation of the Mississippi», for example. Could he have written it? Is it his handwriting? Is the paper watermarked, and from the appropriate time period? Is the ink contemporary? Do other copies of the manuscript exist? Has its recipient or any other contemporary endorsed it? Is there other internal evidence? Who has described it for us? Has it been identified by a trusted third party?

Is a book authentic? Who published it, and who wrote it? Can they be trusted (are they worthy of one's research time?) Is the rare book what it purports to be? Is the manuscript correspondence actu-

ally by the person to whom it is attributed, and is its date accurate? These questions are now being asked more openly of objects that originate in digital form because we have not yet adopted practices or standards for providing ready answers to them. When objects are presented digitally, deciding what is required to authenticate them may be informed from past practices with non-digital objects.

Two experiences with paper objects inform my views of this subject. The first is a multi-page autograph document that lies in the John Marshall Papers at the Virginia State Library. It is labeled in the hand that wrote the entire piece, «John Marshall's Notes on Evidence in Commonwealth v. Randolph, 1796.» Although the title itself might raise some question about who penned it (How often does an author—even an eighteenth-century author—use his own name in a title of one of his documents?), this document has been used for decades for the source of historical articles and at least one full-length book on the investigation of Richard Randolph for murder. Randolph, a member of the famed Randolph family of Virginia, was related to Marshall and to Thomas Jefferson and to many other members of Virginia's «first families.»

Examination of the writing by those familiar with John Marshall's hand, however, quickly reveals that he did not pen this document. Knowing who did write it is important, but does not help make it more authentic as a Marshall document. The possibility that someone in possession of Marshall's holograph could have copied the document raises new questions, not the least of which assigns significant importance to the value of the original of a document, regardless of its form. Internal evidence, obtained by a close reading of this document, reveals that it might be a partial transcript of a hearing in Cumberland County Court where witnesses are questioned by attorneys, and it has been used by historians as a partial record of Randolph's «trial.» But Marshall's name never appears as one of the questioners, and a knowledge of Virginia law at the time would reveal that whatever was taking place could not be an actual trial, because white men could not be tried for felonies at the county court level during that period. In short, efforts to authenticate this document raise more questions than they answer. At the least, such efforts reveal that the document may not be what many had long thought it to be, and that it may not be even what its title says it is.

This example is somewhat esoteric, to be sure, because it is unlikely that one would use a similar digital document without asking the questions that eventually were asked of the document attributed to Marshall. But the questions asked of it suggest attributes that

must be held by a holograph as well as by a digital object that is to be regarded as authentic. Is it the author's work or a copy? Is it what its title purports it to be? What tests can be applied to answer these questions convincingly?

A second example is more to the point. In the collection of Thomas Jefferson's papers at the Library of Congress is a document that appears to be a list of letters written and received between 1791 and 1793, a period of time during which Jefferson was Secretary of State. An examination of the handwriting reveals that it is most likely Jefferson's, but the list is unlike his other journals of letters sent and received. A close look at the original document suggests that it was written in only one or two sittings (the ink changes only once or twice), rather than over three years. The most significant evidence relating to this document's authenticity lies in the paper itself. Holding it before a light source reveals a watermark that indicates the paper was manufactured in 1804. The document, therefore, could not be an authentic 1791-1793 document.

Almost all these tests can be applied to digital objects, and they need to be. But because digital objects bear less evidence of authorship, provenance, originality, and other commonly accepted attributes than do analog objects, the former are subject to additional suspicion. Tests must be devised and administered to authenticate them.

In many cases, problems of authentication arising from objects that originate as digits are obvious. In trying to find solutions to those problems, however, we must carefully test all suggestions to ensure that they do not themselves open new issues that may be inherent in this medium. The problems of preserving digital objects have received more attention than have questions of authentication (people, I suppose, are less worried about authenticity than about preservation). But why preserve what is not authentic? Might the preservation of a digital object imply an endorsement of authenticity, even if nothing else is done to it? More than one archivist has stated that the only sure means of preserving a digital object is to save a printed copy. Concerns with format codes, migrations from version to version, dependence on hardware-would all be solved by printing a copy (or many copies) and putting it (them) in a safe place. Do that to a digital object before confronting the questions of authenticity, and all that is valuable may be lost. Converting a digital object from one program to another, or migrating it from version to version, could present problems of authenticity that may or may not be solved by careful attention to provenance.

A digital object must be authenticated at the time of its creation by a means that will convey a high degree of confidence to all users, including subsequent use by the originator. Clifford Lynch wrote an interesting and convincing article on the integrity of digital information, published in the December 1994 issue of the *Journal of the American Society of Information Science*. He seems to assume, from traditional experience perhaps, that readers will be responsible for authenticating copies being used on the basis of cataloging data to which they must be alert. Retrieving electronic files by title, for example, might lead one to a revised work, different from the original. The reader must exercise caution, Lynch writes, and be ready to detect signs of alteration. «The expectation should be that violations of integrity cannot be trivially accomplished», he says. Accepting this in the world of printed objects is relatively easy. It is much more difficult in the realm of electronic digital information.

Andy Hopper of Cambridge University suggests an authentication strategy that is worthy of consideration, if not adoption. In his system, the concept of a trusted third party is borrowed from the print world. According to this concept, trusted librarians help authenticate their print holdings through recognized acquisition processes, accepted cataloging procedures, and careful stewardship of their collections, especially those in manuscript form. If a special collection librarian tells us, either directly or by means of a catalog card, that the book in hand is one of two extant copies of Ariosto's *Orlando Furioso* printed on vellum in Venice in 1542, and that it was prepared for the dauphin of France, the library's and the librarian's reputation go a long way toward instilling some degree of confidence that the document is indeed authentic. Moreover, all of this information may be checked. If another librarian delivers to a reader a box of letters cataloged as Ernest Hemingway's, authentication is assumed until internal or physical evidence suggests someone has made a mistake. Knowing that the materials-hard-copy objects-have gone through a process of description and identification, if not authentication, conveys a sense of trust that they are authentic, at least until proved otherwise. Some of the problems of description that help authenticate printed special collection objects have similar, if not identical, examples in the digital world. Take one final example as evidence: in the Newberry Library's special collections is a printed copy of the classic book on rhetoric in Renaissance England, *Arte of Rhetorique* by Thomas Wilson (1525-1581). This particular copy is identified as having belonged to Elizabeth I as part of her royal library, and it is authenticated as such by its original binding, which bears the mark of the royal arms.

Book historians know that until the time of James I, the royal arms were put only on the books within the monarch's own library. (After 1603, King James allowed them to be placed on books bound for other members of court.) Elizabeth's coat of arms on the binding of this copy of Wilson's *Rhetorique* therefore marks it as authentic, as long as external evidence does not dispute it. (If it could be shown that the binding was not sixteenth century, for example, or that it resembled the work of a sixteenth-century forger, the authenticity might be questioned).

Some accepted system of similar assumption of authentication needs to exist in the digital world, but it is more difficult to achieve because digital material is more changeable, accidentally or deliberately. Andy Hopper and others suggest that some means of marking digital objects could help solve many of the problems of authentication. Hopper argues that libraries might serve as authenticators by marking digital objects by some means that would remove doubt as to their characteristics at time of origin. A method must be developed whereby a trusted third party, ideally a trusted librarian, would put a marker on a digital document—a marker that could not be predicted or devised (guessed)—that would mark the document's time and date. The marker might be a number based on sonic rays at various times during the day, a number large enough to prevent guessing (Hopper suggests 100 digits). A professor writing a paper could send the document to the librarian to be marked, and it could then be returned to and held by the author. In the future, the object could be authenticated by its marker, regardless of who held it. Any change in the document would remove the marker. This procedure would be used by librarians who receive digital objects from donors. The marker would ensure that digital objects are as authentic as analog objects at time of cataloging.

Despite its science-fiction flavor, such a method seems to meet accepted tests of authenticity. A trusted third party can claim nothing more about an object, analog or digital, than what can be cataloged, and that information derives largely from physical evidence. Identifying an object in a catalog record or a collection description puts a marker on it that most of us use as the first step in the process of authentication. Is the document what it purports to be or what its owner claims it to be? Scholars often require means to test the cataloger, and the physical attributes of analog objects offer more opportunities to do such testing than do those of digital objects. Handwriting, publishing history, bindings, watermarks, inks, and various forms of internal evidence provide answers to questions of authen-

ticity in analog objects that are lacking in digital objects. Digital objects have attributes that can be used to help with authentication, but none is sufficiently trustworthy or stable to be acceptable unless a workable system of certain marking can be devised.

Certifying that a digital object is the product of its author is difficult when the object originates in electronic form. Without a deliberate and distinctive marking caused by the author that could not be guessed by another or altered by anyone, it seems impossible to authenticate an electronic document beyond doubt. Authors of files or images must take steps to establish authorship of their work; if not, our only option is to accept the assertions of others. Electronic files left behind by someone who has not taken action to establish authorship are subject to suspicion if authorship is asserted by anyone else at the time of «cataloging.» This leaves us where we have been all along—at the mercy of catalogers. But, in the case of a digital object, we are actually worse off than we would be if we were dealing with an analog object. This is because we lack the physical evidence provided by analog objects—evidence that offers the means to test the cataloger. This ability to test both reassures the user and helps keep the cataloger honest. I find no corollary in the digital object realm.

The concern over authenticating digital copies of analog objects is almost as important as that relating to objects that originate in digital form. Scholars are keenly interested in having access to documentary evidence in digital form, and librarians have begun to consider digitization a desirable means of preservation, in spite of the recognized problems inherent in it. Those who hope to use this material, once it has been digitized, must be able to rely on its authenticity, just as they have become accustomed to do in all the forms currently available. Documentary editors, as well as librarians, have new responsibilities as they publish and provide access to their materials in digital form with all the value they have added intact. The work of documentary editors offers some insight into the questions raised over authenticity of digital objects, especially those that derive from analog or holographic objects.

The first task of a documentary editor who is working on an edition of a subject's papers is to locate all the objects that have ever existed as part of the *corpus*, incoming as well as outgoing. This sometimes requires reliance on copies of papers that evidence suggests once existed in original form but which have not been found. Once the collection is organized, each item must be dealt with separately. That is the first stage for authentication tests, starting with the question of whether the item is what it appears to be. All the available phys-

ical attributes assist in answering these questions, but sometimes only internal evidence leads to a final answer (as in the Marshall and Jefferson examples described earlier). The editor is obliged to share these findings with readers and to describe the item in such a way that few, if any, questions remain about the document as object. Not unimportant in this description is all available information about other copies of an item, be they photocopies, carbons, letterpress, polygraph, drafts, or additional holographs. Knowing as much as the editor about all copies is the only sure way for other readers to test the «cataloger's» description, and only by having this information available can a reader have full confidence that all questions of authenticity have been asked. In preparing digital files of historic documents, editors begin their publication by attaching a full document description to a transcript. This is the scholar's seal of authenticity, as it were, or at least as much of a seal as a scholarly editor can provide.

Preparing a digital transcript of a historic object introduces new problems to the issue of authentication. How do we know the transcription is accurate and that it is exactly what the editor prepared originally? The method of providing access to journal articles adopted by JSTOR may offer the best answer for authenticating modern digital transcripts of manuscripts or printed material that originated in analog form. They provide the user with a digital transcription of the text, which is fully searchable and otherwise subject to all the vagaries of digital files. They also provide an image of the original text. If both copies could carry some form of marking that could not be manipulated, the problem of authentication would be solved. This system should work quite well for documentary editors and the readers of their digital publications. Providing an image of the document that is transcribed would be an important improvement over present forms of presentation, because it would permit easy verification of transcriptions. Inaccurate transcriptions are the downfall of documentary editors (as they should be), and mistakes often go undetected. The reader, who may have a high level of confidence in the scholarly work of the editor, is left to assume that the transcription is accurate and authentic. Having a means of testing this assumption would be a great improvement.

Related problems that arise from considerations of authenticity seem to offer little to assist us in answering the primary question. Creating a digital file, and even marking it in such a way that will ensure authenticating it as my own, will mean little if the file itself cannot be read at any point in the future. If the file cannot be read, it cannot be authenticated as mine. (It would be even more mad-

dening if the file could be authenticated but not read.) The same can be said for provenance. If a file can be marked in such a way that its authenticity is assured, issues of its subsequent provenance might not matter in questioning its authenticity. But if a file cannot be read, its provenance will mean little, even if it can be tracked over a long period. Without a marker of authenticity, provenance of a digital object would be of limited use in establishing authenticity. It would help test the cataloger, but the current technology would render uncertain any assertions of authenticity. The instability of software alone would introduce questions that would challenge any claims of authenticity suggested by a trusted provenance.

Paul Conway (1999) says the existence of digital objects moves challenges of preservation from guaranteeing the physical integrity of objects to assuring their intellectual integrity, including their authenticity. He adds that librarians can control this by «authenticating access procedures and documenting successive modifications» to digital files. Authenticating access procedures may affect provenance more than the integrity of the digital object itself, but it would be difficult to guarantee authentication with only this control. It seems that, in this argument, the alteration of an original record is acceptable as long as it is documented. Acceptance of changes with documentation is unreasonable over time and places unnecessary burdens on users. In this case, as in others, preservation without authentication results in a loss of intellectual integrity.

We are not close to having a means of marking digital documents that cannot be challenged – a means that would establish authenticity. Absent such a technique, we are left to consider what other attributes, if any, might approach the establishment of authenticity. Few suggest any high degree of confidence that would come close to what we have for analog materials, but consideration of the problem raises some issues that relate to other concepts that bear on the problem. How confident can one be when an object whose authentication is crucial depends on electricity for its existence? Surely there are higher degrees of confidence in some cases than in others, but something more than provenance or traditional testing methods established for analog objects is needed. I believe it is easier to describe the characteristics of an authentic digital object than to support the authentication beyond a reasonable doubt. My definition is conditional; it depends on an object's capability of being proved to be authentic. Establishing a method of authentication of digital objects that would be unconditional may be possible. At the least, we must agree on some means of testing the authentication of digital objects. The consequences of not doing so are dire.



REFERENCE

CONWAY, Paul.

1999. The Relevance of Preservation in a Digital World. Technical Leaflet, Section 5, Leaflet 5, p. 8, Andover Mass.: Northeast Document Center. Available from <http://www.nedec.org/plam3/tleaf55.htm>.

---

CHARLES T. CULLEN is president and librarian of the Newberry Library in Chicago. A legal historian, Mr. Cullen earned his Ph. D. degree from the University of Virginia. He has taught at Princeton University, the College of William and Mary, and Averett College. While at the College of William and Mary, he worked on the Papers of John Marshall, and became editor of that project in 1976. At Princeton, he was editor of the Papers of Thomas Jefferson.

A charter member of the Association for Documentary Editing, Mr. Cullen served as its president in 1982-1983. He currently represents that organization on the National Historical Publications and Records Commission. He has served on the New Jersey Historical Commission, and was vice-chairman of the board of the Founding Fathers Papers, Inc., from 1986 until 1992. He is currently vice-president of the board of the Modern Poetry Association, and chairs the board of the Heartland Literary Society.

Mr. Cullen has published or contributed to 30 books and articles. He has lectured widely on subjects relating to the Age of Jefferson, the scholarly use of computers, and the role of humanities research libraries. His scholarly computing expertise was recognized by the Association for Documentary Editing, which in 1987 awarded him its Distinguished Service Award «for outstanding contributions to the field of documentary editing through the use of computers.