

More than Information, other than Knowledge: the Nature of Archives in the Digital Era

LUCIANA DURANTI

PALAVRAS-CHAVE | ARQUIVOS | DOCUMENTOS ELECTRÓNICOS | INFORMAÇÃO | CONHECIMENTO | AUTENTICIDADE

A B S T R A C T

The nature of archives has been defined over time according to the use made of them. The records of an archival fonds have been considered facts, acts, sources, evidences and, in more recent times, equated with information and even knowledge. This evolution is mostly linked to the use of digital technology for creating records and the consequent blurring of the boundaries among the various types of digital objects generated within electronic systems. This trend is undermining our ability to create records that are reliable and to maintain archives so that they can be proven authentic over the long term. It is therefore essential to re-establish a concept of archives based on the circumstances of its creation, to define the characteristics that distinguish records from any other digital entity, and to protect them in trusted recordkeeping system capable of ensuring that their nature will be altered intentionally or accidentally.

R E S U M O

A natureza dos arquivos tem sido definida ao longo do tempo de acordo com a utilização que deles é feita. Os documentos de um fundo arquivístico foram considerados como factos, actos, fontes, prova e, mais recentemente, equiparados a informação e mesmo a conhecimento. Esta evolução é essencialmente tributária da utilização de tecnologias digitais para criar documentos e o consequente desvanecimento das fronteiras entre os diferentes tipos de objectos digitais gerados em sistemas electrónicos. Esta tendência tem minado a nossa capacidade de produzir documentos fidedignos e de manter arquivos de forma a preservar continuamente a sua autenticidade. É portanto essencial restabelecer o conceito de arquivo, o qual deve ser baseado nas circunstâncias da sua criação, definir as características que distinguem documentos de arquivo de qualquer outra entidade digital e garantir a sua protecção através da integração em sistemas de arquivo capazes de garantir que a sua natureza não seja alterada de forma intencional ou accidental.

«There is a powerful pair of contrasting images of the archive
– the temple of fact, objectivity and omniscience;
the factory of deceit, distortion and prejudice.
Truth-telling and fiction-making are both persistent
truths about archives!.»

Over the centuries the nature of archives has been defined more by their use and reason for preservation than by their creation and reason for maintenance. In ancient Rome, archives were *universitates rerum*, according to the juridical concept that regarded its component units, the records, either as acts or as memories of acts. As long as they were kept by a public authority in a public place, records could be trusted as statements of the facts they either put into existence or talked about. The status of archives as the repository of actions was enhanced rather than diminished by the awareness of their corruptibility, as only the object of the deepest trust would be worth the effort of falsification. Because of the trust associated with the records, archives were used from antiquity to modern times as political instruments by sovereigns and regimes of all kinds and were indeed mutilated, forged, selectively preserved to serve the interests of their creators. Regardless of the voluntary tampering with and the involuntary loss of records, those same sovereigns and regimes kept archives as a most precious asset, recognizing their inherent capacity to reveal the facts of the past. The studies on the records carried out by humanists, and the development of the diplomatic method of analysis refined the understanding and use of archives as facts and acts, and the growing attention given to the authenticity of individual records was directly related to the weight that any given record had as the substance and proof of past action.

With the development of history as a profession in the nineteenth century, archives began to be regarded as sources of historical understanding, and a great preoccupation started to arise about their integrity as bodies of records, rather than as individual records. The principles of *respect des fonds* and *respect pour l'ordre primitif* formulated in those times are an expression of the importance increasingly given to the documentary context of the records as the key to their true meaning². While diplomatists had dissected the record to assess its capacity to tell the truth, historians analyzed inventories, classifications, indexes to understand the structure of the archival aggregation and assess its capacity to reveal the cause-effect relationship between one record and another, one fact and another.

This is the time when the concept of archival bond was first theorized as the necessary link that each record has with the records belonging in the same aggregation, arising at the moment of creation and determined by the function that the record is meant to fulfill. This incremental network of relationships gives meaning to a record, identifies it, and authenticates it by placing it in perpetuity in a unique location among the records of the archival fonds³.

Undoubtedly, the prevalence of the concept of archives as sources over that of archives as facts was linked to the growing production of records caused by the historical developments that accompanied the arising of large bureaucracies and the industrial revolution, and to the consequent multiplication of record types, only few of which could be characterized as acts either *ad substantiam* or *ad probationem*⁴. In time, the concept of archives as facts was entirely replaced by that of archives as evidence, an idea similar but more nuanced. Where a fact is an occurrence, a natural event or human conduct, evidence is a relationship between a fact to be proven and the fact that proves it. The old conception of record as fact focused on the one to one relationship between the content of the record – what the record talks about – and the record itself.

In contrast, the evidentiary capacity of the record consists of the relationship that can be established between a record and an infinite number of facts through its content, its form, its archival bond and its context.

Thus, by the middle of the twentieth century, archives were regarded as sources of historical understanding and as sources of evidence. Their trustworthiness was mostly based on the circumstances of their creation. As Hilary JENKINSON put it, archives are impartial because they were not created with the needs of posterity in mind, and they are authentic because of the way of their creation and preservation⁵. Archives are the most reliable sources because they were not intended to be sources but means for carrying out activities. Because of the circumstances of their creation, archives are considered the primary means of both administrative and historical accountability⁶. Nevertheless, those who use archives as sources have recognized and still recognize that to reconstruct the past from the records requires first extracting meaning from their medium, documentary form, content, archival bond, contexts and custodial history, and then undertaking many acts of interpretation of that meaning⁷.

It was the pervasive use of digital technology that changed the general outlook on archives from sources to information. Although tracing the origin and development of this evolution and its various causes would require an in-depth study of the use of information technology in the modern office and of the professional stance of archivists with regard to such use, it may be impressionistically stated that the change did not happen immediately with the first adoption of digital technology by records creators. Digital technology was initially used to record and manipulate data, which were regarded as information that might be preserved in archives to complement the related records, but not as records itself. Later, when designing electronic document systems, information specialists were still thinking of the entities that these were supposed to create and manage as identifiable objects having a structured form, not only content – in other words, documents. The medium only existed as a neutral support, the hard-drive, a diskette or a tape, but the document as an inextricable whole of form and content only changed from an analog to a digital inscription. Although the distinction between records and documents became blurred, we still had indivisible and unchangeable units of meaning. Over time, however, the various logical components of the document began to be regarded as units existing separately, proprietary formats multiplied in uncontrollable numbers, the trend towards increased manipulability of the content diminished the importance of form, the need to update continuously the data over-rid the significance of the fixity of the document, and the requirements of preservation in spite of incompatibility and obsolescence brought about the idea that information is what really matters. As a result, many archivists started to regard archives as information and themselves as information professionals. It is hard to say how much of such choice was linked to a professional status issue, especially in the context of the current office, where specialists in information technology have been increasingly entrusted with records management responsibilities, or to a desire to facilitate communication with those who design and use digital system. Whatever the case, ironically, the terms "record", "file", "archive" and "archiving" were appropriated by computer scientists and engineers and given meanings quite different from those that originated in diplomatic and archival science, while the term "information" became the buzz-word used to refer to archival material in any form.

The consequences of this evolution were clearly observable in the course of the research conducted by the InterPARES project, directed by this author⁸. The InterPARES (International research on Permanent Authentic Records in Electronic Systems) project is a collaborative, multidisciplinary and international research endeavour aiming at developing the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records. In order to achieve such goal, the researchers considered necessary to understand whether the entities existing in a variety of digital systems had record nature, what were their function and form, what action(s) they were connected to, etc. In the cases in which the system under examination obviously did not contain records, the questions regarded whether, given the administrative or operational functions of the system, it should contain records, and whether the system itself can be considered a record (e. g. the register of native Indians is a database that can be considered a record in the making). For the purpose of these case studies, an electronic record was defined as a record created or used in electronic form⁹, and the following essential characteristics were identified:

- 1) a fixed form, meaning that the binary content must be stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside;
- 2) an unchangeable content;
- 3) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier;
- 4) an identifiable administrative context;
- 5) an author, an addressee, and a writer; and
- 6) an action, in which the record participates or which the record supports either procedurally or as part of the decision making process.

The primary methodology of analysis was comparing that which is unknown with what is known by constructing a template representing the traditional record. Following the classic diplomatic assumption that, regardless of differences in nature, provenance or date, all records are similar enough to make it possible to conceive of one typical, ideal documentary form containing all possible elements of a record, the researchers hypothesized that, while they may manifest themselves in different ways, the same elements that are present in traditional records exist either explicitly or implicitly in electronic records, and that all electronic records share the same elements. Thus, they created a template

that would allow for a systematic analysis of the electronic records contained in several different systems. «The template is a decomposition of an electronic record into its constituent elements which defines each element, explain its purpose, and indicates whether, and to what extent, that element is instrumental in verifying the record's authenticity¹⁰.» The template was used to study a significant number of cases in order to assess whether the elements identified in it were present in the records contained in the systems in question and, if so, to what extent and where, and to identify elements present in the records contained in the systems but not in the template. As the case studies proceeded, the template was refined in light of the results.

The analysis indicated that only a half of the examined systems contained records (twelve out of twenty-two), primarily because the entities identified within the other half did not appear to possess either a fixed documentary form or a stable content. When systems did contain records, these could rarely be compared with the model represented by the template, because they were not good records. For example, in most systems, there was no explicit manifestation of the archival bond (e. g. a classification code) among the records participating in the same action or transaction and, although it was easy to identify the business processes supported by the system, it was not always possible to determine how the records participated in or supported specific activities¹¹. In addition, it was often difficult to determine the significance of the presence or absence of given elements of documentary form or of annotations. The investigation also revealed that creators tend to trust unconditionally their digital systems, to rely on procedural means for protecting the authenticity of the entities they contain, and to focus on the management of the entire system rather than of the individual entities within the system. The most common means to ensure authenticity were access privileges (protected by means of passwords, user ID, user profiles), audit trails and back-up procedures.

From these case studies, which were conducted during the first phase of InterPARES, it appeared quite clear that, generally speaking, the modern office manages information instead of records, and does not take appropriate measures to ensure that records, rather than information, result from its business processes, and that such records have the capacity to stand for the facts they are about, can be used as sources of evidence and of historical understanding, and are accurate, reliable and authentic.

The consequences generated by this state of affairs, especially as it regards accountability, have begun to be felt. Concern is growing in those countries where e-government is about to be launched. But the solution does not seem to be very close. The reasons might have something to do with the fact that the idea of archives as an integrated indivisible whole of records that only have meaning in their documentary and administrative context has been lately much neglected, and that the use of digital systems has brought the focus on the individual records and, at the same time, has altered their nature by treating them as information. But, these problems would not be so significant if they were not compounded by the pervasive use of the web to conduct business. Roger T. PÉDAUQUE writes: «the Web can be described as an infinity of linked documents. Its architecture is based on three pillars: resources identified by a universal addressing scheme (identification), which are represented as a nonexclusive set of schemes (representation) and exchanged using standard protocols. This architecture assumes that documents can be accessed from anywhere, using any type of hardware and according to the specificities of the user groups¹².» The key to this access is the development of appropriate metadata. In a way, metadata have become the essential tool that substitutes registers, classifications, inventories, indexes and thesauri for the purpose of information retrieval of both material posted on the various web sites and of digital objects maintained in office applications, document management systems and recordkeeping systems. This may be a good thing as long as one does not forget that those descriptive instruments do not have only a retrieval function and should be maintained for the other essential functions that they fulfil¹³. In addition, because of their development and use in the context of the web, metadata appear to have acquired also in the context of the office systems and applications the capacity of identifying and organizing information in the records that meets the cognitive demands of the users. Thus, the concept of records as information is being replaced by the concept of records as knowledge, which integrates the idea of reasoning into information. A new professional has emerged, called "knowledge engineer", whose specialty is to develop knowledge management systems, that is, systems capable of identifying the knowledge that the user needs to access. These systems are beginning to penetrate the modern office and to be used as recordkeeping systems, which of course they are not. They are just another application, very similar in a way to Geographic Information

Systems, which, depending on the way they are used, may or may not produce records, but certainly should not keep records or guide their retrieval.

What does this situation do to electronic records? Quite simply, by presenting all digital content which is identified by metadata as being information plus meaning, capable to be interpreted by the system according to the needs and the specific context of the reader, annihilates them as facts, evidence, sources, and makes a travesty of their natural attributes of impartiality and authenticity. As PÉDAUQUE remarks, through this process of interpretation, meaning is replaced by knowledge, and this «introduces the concept of personalization for a given reader or user¹⁴». True recordkeeping systems support knowledge enabling by establishing the identity of each record and maintaining the integrity of the record and of its identification, and by controlling, facilitating and documenting its routing, transmission, handling, filing, retrieval, accessibility and disposal, but they do not and should not interpret either the records, their metadata, or the data about the processes in which they participate. There is no doubt that knowledge engineering, especially as it regards the development of ontologies¹⁵, can be extremely helpful for the intellectual control of records, but its function of isolating meaningful parts within a set of interdependent elements that has a unity of meaning and can only be understood as whole can make sense only within an application used by an individual or organization for very specific and clearly detailed and personalised research purposes, certainly not in the context of a record keeping system, where records configured as knowledge might be closer to works of fiction than to sanctuaries of truths, as our ancestors wanted them. This is because the validity of the knowledge that can be derived from an interpretation of the records is based on protocol rather than on contents: in its context, the truths...of the procedures of its making and conservation... can be decoded, comprehended and, for that matter, contested¹⁶.»

Records have meaning in the context of the archival body and system in which they belong. This is because they are made of two inextricable components, the documents and their interrelationships, and the latter are determined by the actions of which the documents are the written instruments and by-product. The information contained in the archives and the knowledge derived from interpreting it by whatever means can only be validated by the authenticity of the records belonging in the archival body. This authenticity comprises

the identity of the records and their integrity. The identity of each record is revealed by its elements of form that link it to a given time and place, an author, an addressee and a writer, an action, an aggregation of interrelated records, and a provenance context. Its integrity is revealed by the administrative and documentary procedures in which it participates, by the control exercised on its accessibility, maintenance, use and long-term preservation. Knowledge engineers are studying ways of allowing users to increase the credibility or confidence that can be attributed to the information in a record or the knowledge gained by linking parts of several records through metadata by instituting some sort of certification for the process itself, but, as MABILLON put it, «it is not authority that makes for authenticity, it is authenticity that makes authority in virtue of the conception according to which authority is never but a matter of witnessing¹⁷». In other words, records, be they in paper or in digital form, are witnesses to and of human conduct and we must maintain and assess them as indivisible units of meaning in their context of creation. If we do that, the information and the knowledge derived from them can be trusted. But first we need to change our way of defining the nature of archives and move the perspective from the user to the creator, remembering that what makes a record is the why of its creation and the how of its keeping. The old saying that a record is what a creator treats as a record is now more valid than ever: we have observed that what a creator treats as information is nothing more than information.

If we think of the record as a means of carrying out an activity that is kept by its creator or legitimate successor for action or reference, rather than as fact, source, evidence, information or knowledge, then we know that its authenticity depends on identifying it through metadata related to its creation, indicating the time, persons, action or matter and archival bond (be they in a register or in a record profile or a topic map), and maintaining it, with its metadata, in a trusted recordkeeping system under the responsibility of a trusted custodian¹⁸. In fact, as STARN writes, «the mostly tedious and formulaic quality of recordkeeping is a better reason for trust than the transports of sublimity. Boredom and redundancy are friends of truth in the archives because they mitigate against hasty judgments and improvisation¹⁹».

The nature of archives has not changed with their form. If anything, their new fragility, manipulability and vulnerability, due to the characteristics

of their medium and of the technology generating them as well as to the incompatibility and obsolescence of systems and hardware, has made essential to represent such nature in the utmost detail and to protect it through a chain of preservation activity that begins at the moment of their creation. Only if we explicitly identify the digital records of an archives²⁰ as distinct from any other digital object that may accompany them, and are able to ensure their continuing authenticity through time, we will be able to continue to use them as facts, sources, evidence, and as reliable information, although certainly not as knowledge. There is no doubt that the most trustworthy form of knowledge is one based on the study and interpretation of archives, but archives are no more knowledge than they are truth: this is one truthful piece of knowledge to keep in mind when thinking about them.

NOTAS

¹ Randolph STARN – "Truths in the Archives". *Common Knowledge* 8:2 Duke University Press, 2002, p. 387.

² Luciana DURANTI – "Archival Science". *Encyclopedia of Library and Information Science* New York, Basel, Hong Kong: Marcel Dekker, Inc., 1997, p. 1-4.

³ Luciana DURANTI – "The Archival Bond". *Archives and Museum Informatics* 11, 1997, p. 215-217.

⁴ Luciana DURANTI – *Diplomatics. New Uses for an Old Science*. Chicago: Scarecrow Press, SAA and ACA, 1998, p. 65-66.

⁵ Hilary JENKINSON – *A Manual of Archive Administration*. Oxford: The Clarendon Press, 1922, p. 12-13.

⁶ Terry EASTWOOD – "Reflections on the Goal of Archival Appraisal in Democratic Societies". *Archivaria* 54 (Fall 2002), p. 66-67.

⁷ Thomas TANSELLE – "The World as Archive". *Common Knowledge*, cit. p. 402-403.

⁸ The first phase of InterPARES, dealing with databases and document management systems, was concluded in 2001, and the second phase, begun in 2002, will be concluded in 2007. For further information, see the project's website www.interpares.org.

⁹ A record was defined according to archival tradition as any document made or received and set aside for further action or reference by a physical or juridical person in the course of a practical activity as an instrument and by-product of it. A document was defined as recorded information, information as a message intended for communication across space or time, and data as the smallest meaningful piece of information.

¹⁰ Heather MACNEIL – "Providing grounds for trust: developing conceptual requirements for the long-term preservation of authentic electronic records". *Archivaria* 50 (2000): 56. This article describes in detail all the elements included in the template and the way it was used in the case studies. The template itself can be found on the InterPARES website at www.interpares.org <resources> <reports> "The Template for Analysis".

¹¹ The absence of proper classification systems, mostly due to the blind trust of records creators in the retrieval capacities of the powerful search engines of increasingly complex technologies, and the tendency to keep records by application rather than by archival aggregations, appear to be the main cause of unretrievability of records.

¹² Roger T. PÉDAUQUE – "Document: Form, Sign and Medium, as Reformulated

for Electronic Documents", Version 3 (July 8, 2003), 8. Electronically available on http://archivesic.ccsd.cnrs.fr/documents/archives/0/00/00/05/94/index_fr.html.

¹³ For example, registers attest to the fact that a record has been sent or received at a certain time and assigned a certain classification; classifications show the archival bond among the records of a funds, the office competent and the office of primary responsibility for each record aggregation, etc., inventories detail the contexts of the records and their custodial history and represent the collective authentication of the archival funds as they make explicit and perpetuate the relationships among the records and their aggregations, indexes are keys to register numbers, classification codes, etc., and thesauri show the preferred term, the authority file, etc.

¹⁴ *Ibid.*, p. 14.

¹⁵ Ontologies are representations of a domain, focusing on its vocabulary and related meaning as understood by human beings and by machines. They have both a semantic and a lexical dimension, and a formal structure.

¹⁶ Randolph STARN – "Truths in the Archives". *Common Knowledge*, cit. p. 393.

¹⁷ Dom Jean Mabillon in Randolph STARN – "Truths in the Archives". *Common Knowledge*, cit. p. 399.

¹⁸ A trusted record-keeping system is one that controls what records are included in the system, who can include, retrieve, modify, delete or remove them from the system, and how the records are included, maintained, retrieved, deleted or removed from the system. A trusted custodian is a person entrusted with the responsibility of preserving the records, having demonstrated that it has no reason to alter the records entrusted to its care or to allow others to do so, and is capable of implementing the necessary measures for the physical and intellectual protection of the records. The requirements for authenticity developed by the InterPARES project can be found in the appendix to the book posted on its website: <http://www.interpares.org/book/interpares_book_k_app02.pdf>

¹⁹ STARN, *Ibid.*, p. 392.

²⁰ The records of archives are not all in digital

form. It has not happened yet and it will probably never happen. This is why I prefer to avoid the term digital archives and to use the expression "an archives' digital records".



O SISTEMA DE GESTÃO E AUTOMATIZAÇÃO DE BIBLIOTECAS

PROCURA INTEGRIDADE E QUALIDADE?

O ALEPH500 permite-lhe gerir com QUALIDADE, FLEXIBILIDADE e SEGURANÇA.

O sistema mais flexível, aberto e eficiente para a **Gestão de Bibliotecas**, com mais de 700 instalações em 48 países.

Fácil de utilizar e de actualizar, o ALEPH500 pode ser personalizado conforme as necessidades. O conjunto de tabelas de configuração de que dispõe permite ao gestor da biblioteca controlar e modificar o ambiente a qualquer momento. O ALEPH500 está disponível em 20 idiomas, incluindo o Português.

A DATINFOR dispõe ainda dos seguintes produtos para uma melhor gestão de Bibliotecas:

• **METALIB** (Virtual Library Services): a solução perfeita para gerir ambientes híbridos em bibliotecas, como portal de pesquisa de recursos de informação. Metalib é uma verdadeira entrada virtual para chegar a todo o tipo de informação, gratuita ou licenciada, disponível em modo local ou remoto.



MetaLib

• **SFX** (Context Sensitive Linking): uma ferramenta única e revolucionária para navegação e pesquisa. Possibilita a interligação entre todos os aspectos das coleções digitais de uma biblioteca, incluindo repositórios em texto integral; bases de dados sobre resumos, indexação e bases de dados de citações; catálogos "on line"; e referências extraídas de artigos de investigação, bases de dados de conteúdos, arquivos "e-print" e outros recursos da web.



S.F.X

• **DIGITOOL** (Digital Asset Management): um sistema de gestão de recursos digitais que disponibiliza todas as ferramentas necessárias para criar e gerir coleções digitais de acordo com uma estratégia de gestão digital global. Com capacidade para gerir grandes quantidades de informação e ao mesmo tempo disponibilizar rápida e eficazmente resposta a pesquisas direccionadas, a DigiTool proporciona a resposta adequada para as necessidades actuais de qualquer biblioteca.



DigiTool

DATINFOR
INFORMATION, ANALYTICS & SERVICES, S.A.

S4Y solutions for → YOU

Urbanização Quinta do Património, Lote 5 - R/C,
2685-005 Sacavém
Tel.: 219 426 600 Fax: 219 426 615

Rua Dr. Carlos Felgueiras 98 - 4.º D,
4470-157 Maia
Tel. 229 439 900 Fax: 229 439 910