

“It was just in the air, at any place...”

The Principle of Provenance & Modern Archival Systems

Provenance

Variouly described as a “powerful guiding principle” (Dearstyne, 1993), and “the only principle” of archival theory (Horsman, 1994), the Principle of provenance distinguishes the archival profession from other information professions in its focus on a document’s context, use and meaning and is fundamentally important in characterizing an archive’s role within society. Not only does an archive save and keep all of a culture’s important records, it attempts to keep them contextually meaningful through providing structure and organization meaningful to the original purpose and function of the document. It is this intention of providing context that distinguishes an archive from a museum or library, and it is context that ultimately gives records their evidential value. The Principle of provenance has three distinct meanings (Bellardo & Bellardo, 1992). First, and generally, it refers to the “office of origin” of records, or that office, administrative entity, person, family, firm, from which records, personal papers or manuscripts originate. Second, it refers to the idea that an archival collection of a given records creator must not be intermingled with those of other records creators; in this sense, the principle is often referred to by the French expression *respect des fonds*. A corollary principle, solemnly entitled, “Principle of the Sanctity of Original Order,” states that records should be kept in the order in which they were originally arranged. Finally, the Principle of provenance refers to collecting information on successive transfers of ownership or custody of a particular paper or manuscript.

The Principle of Provenance was independently developed by early modern French and Prussian archives managers in the nineteenth century, and had its origins in necessity, both theoretical and practical. Prior to the development of the Principle, archives were arranged and described according to the “principle of pertinence,” where archives were arranged in terms of their subject content regardless of provenance and original order (Gränström, 1994). Although the Principle has been widely adopted and is, in fact, a fundamental concept within archival practice, it still generates discussion, controversy, and sometimes misunderstanding. The Principle becomes complicated when considering its inward and outward application to archival documents, and is illustrated by an argument that took place in the early

twentieth century between a Swedish archivist, Carl Weibull, and the Netherlands State Archivist R. Fruin, the only author of the *Manual of Arrangement and Description* still living (discussed in Horsman, 1994). This argument hinged on the application of the Principle to papers within a fonds. Weibull agreed with the outward application of the Principle of provenance (*respect des fonds*) but was critical of the idea that within a fonds the archivist must preserve the original order as it came from its creator (*the sanctity of original order*). That order might have served the administrative body creating the records, and it might make the archivist's job easier, but does not serve historical research. Fruin replied that Weibull was right to be concerned with the needs of researchers, but argued that re-arranging a fonds on the basis of subject oriented classification, something that a librarian might do, is a subjective enterprise, and not the archivist's responsibility or charge. The archives creating administrative body knew the best arrangement for the effective use of its records, and this arrangement would be the most objective for future research. "No, it's one thing or the other. If one accepts the Principle of provenance, then one must respect the original arrangement of the fonds. As soon as one gives it up, the whole principle is unsettled." (Fruin to Weibull, quoted in Horsman, 1994)

This argument illustrates a whole range of issues related to nature and purpose of archives themselves. Among others: Are archivists supposed to be objective and scientific? Is the Principle of provenance addressed primarily as an administrative, time-saving endeavor or because is it theoretically sound from an evidential perspective? Exactly how *is* evidence and authenticity established? Is evidence fundamentally important to archival practice? These are obviously difficult questions, and each topic has generated voluminous literature, and in once case, the objectivity of the archivist, whole new bodies of thought. But the most basic tenet of archival practice, providing context, has never been questioned. People might argue about the best way to provide context in different environments (physical versus digital), but the urge to give contextual meaning is always present. My argument is that the Principle of provenance provides the most robust and straightforward means by which archivists and researchers can deduce context from objects in any system.

In the physical world, it's relatively straightforward to identify a paper's provenance. Typically, an administrative entity can provide organization charts, manuals, and office procedures and policies. There are further contextual clues garnered from surveying the physical arrangement of the records at time of

ingest. Physical records have formal and physical characteristics like watermarks, and letterheads; and provide content-oriented clues like authors' and recipients' names, dates, and references to other documents or files. Finally, physical records have secondary information like marginalia, initials, and similar signs that reveal the information environment in which records were created, used and maintained. In this paper-based world, these physical, formal and content-oriented record elements are inseparable (see section on Diplomatics, below), and the physical location of one record in relation to another denotes how they were created and used at a given point in time. If any of these elements are not present in a record, if someone has taken the records out of their "original order," for example, the relationships and inter-relationships may be difficult, if not impossible, to clarify, and their evidential value as records will no longer exist. What happens, then, when context isn't readily available? When physical records give way to newer and more complicated electronic records, how does the Principle of Provenance work? What's involved in providing contextual information in an electronic age?

Charles Dollar (Dollar, 1992) has discussed the complexities involved in archiving electronic records. First, he points out that these records do not exist as discrete physical entities, and most of the contextual information, something as simple as the author's name, is either not visible to users, or not consistently captured. Further, the arrangement of bits and bytes that form a record seldom have any relationship to the image displayed on the monitor or printed on the page. This separation of physical and logical relations adds a level of complexity to the already complicated process of archiving. Add to this complexity the idea that instead of discrete records, future archives will be responsible for capturing and managing networked or corporate-wide databases, which "dissolve the traditional boundaries between organizations, sub-operating units, and offices that in the past provided much of the provenance based information," otherwise known as the record's "office of origin." Dollar recognizes the magnitude of this problem, argues for the importance of stringent metadata standards, and gives recommendations that archivists should follow to make capturing provenancial information possible. Basically, he believes that archivists should identify and articulate the archival requirements for capturing provenancial information, then intervene in the development of software and systems applications that would ensure the capture and preservation of this information.

David Bearman is another scholar interested in the question of provenance of electronic records. He has written quite a few articles on this subject, but his argument boils down to a few points: the Principle of Provenance typically dictates that records ought to be “understood with reference to their origins in activity” (Bearman, 1994). In trying to simplify the process, archivists tend to equate the provenance of records with the organization in which records were created or received, thus the focus on the “office of origin.” Bearman believes, however, that provenance is better understood by reference to the *function* of which the records are evidence, and to the record system in which they were created, stored, preserved, and accessed by the organization. He argues that, among other elemental attributes (Bearman, 1989); archivists should recognize that function, not organizational setting, is the “locus of the provenancial meaning.” With regards to electronic records, the problem then becomes not focused so much on the contextual and administrative information, which Dollar is interested in capturing, but on elucidating the purposes for which records were created. This significantly simplifies matters for Bearman, who argues that it is easier to capture the functional provenance of records either explicitly, recorded as data within the record by the creator or system; or implicitly, in the system design, which would reveal through analysis “the structural relations between data instances, or discovered by links to the originating activity, which is represented by the source of the records, or more exactly by knowledge of the transaction communication path.” Boiling down, once again, to articulating functional requirements for capturing provenance information, and working with systems designers and engineers to incorporate those requirements into working systems.

Provenance and Diplomatics

The idea of providing requirements and standards by which documents could be objectively judged is not a new concept. Scholars have always been concerned with objectivity of their original source material. They need to be able to establish what really took place, and to do that, they felt that the written sources should be maintained in their original order, and not rearranged. Related to the scholarly standards, the Principle of Provenance also held with medieval diplomatic procedures, which were concerned with defining and evaluating records based on their authenticity and evidential, primarily legal, value.

Originally developed in the Seventeenth century to identify, evaluate, and communicate the “true nature” of archival documents, diplomatics was developed as a measure to ensure the authenticity of legally binding documents. This is achieved by studying a document’s origins, forms, and transmission, as well as the relationships between a document and the facts presented, and between a document and its creators. More recently the term *diplomatics* has come to represent a broader concept, which covers the range of skills and methods required in the study and interpretation of the historical significance of documents. Very simply put, it is “the study of documents” (Peter Herde, quoted in Duranti, 1998). The modern use of the term was developed by Jean Mabillon in the late seventeenth century in his still influential study, *De re diplomatica*. Its subtitle, “whatever pertains to ancient instruments, their age, material, writing and style; whatever pertains to seals, monograms, signatures, or chronological notes; whatever pertains to the study of antiquity, history, and legal force of such things,” is, in effect, a definition of the field that still holds true. Of particular interest to Mabillon was the distinction between “book hands,” and “charter [or court] hands,” which corresponds to the fundamental distinction between two forms of document: those with “real effects,” like deeds to land, and those without “real effects,” like works of literature. In this current discussion of provenance, this primary distinction is interesting in that the original functional provenance, as well as that of the administering body is given tremendous weight over the character and subsequent treatment of any document.

In the modern period, Father Leonard Boyle, in his essay on “Diplomatics,” (Boyle, 1976) resisted the urge to limit diplomatics to the study of legal documents and supported those who wanted to view the discipline as the application of principles of literary criticism and interpretation to all documents of historical interest (Reimer, 1998). For him, diplomatics is the study of all kinds of historical documents, and diplomatics is “an art by which written records from any age and of any kind are made to speak again with a full distinct voice.” (Boyle, 1976) His primary concern is the definition of an effective approach to those documents, and he suggests that when studying historical documents, a scholar must establish the: *who, what, when, where, and why* of the document. In terms of the “who,” a scholar must try to identify, as far as possible, all of the people involved in the creation of the document, either primarily or secondarily: the “actor,” the recipient, the witnesses, the notaries, and the scribes. The “what” of a document includes a consideration of its quality (related to its authenticity and reliability), and its physical makeup. This facet

also requires a full and accurate transcription of the document, which will necessitate paleographical skill, linguistic expertise, and knowledge of conventional topoi and styles in the period and document type in question. Additionally, the scholar should also try to determine any kind of changes that might have been made to the document, compare it with any other copies, and never forget to look on the reverse side of the document to find any marks, which might indicate the date of receipt, any notarial commentary, or summaries. The “when” (dating) and “where,” (localizing) of the document are key factors in determining its significance and effect – and all of these factors contribute to the “why” of the document: “what purpose did it serve?” “What action did it perform?” (Reimer, 1998)

In 1998 Luciana Duranti revived (Richardson, 1983) the study of diplomatics with the publication of her book: *Diplomatics: New Uses for an Old Science*, and her subsequent formation of the international InterPARES work group, dedicated to using the principles of diplomatics to help define and preserve authentic digital documents. Duranti’s definition of diplomatics, “the key to understanding the action in which the document participates, and the document itself.” Her breadth of diplomatic knowledge is great, but she has made three specific distinctions, upon which I’d like to elaborate: her distinction between authenticity, reliability and genuineness, her definition of the intrinsic and extrinsic elements of documentary form, and her discussion of the nature and purpose of diplomatics.

Authenticity, Reliability & Genuineness

Duranti has proposed two somewhat conflicting definitions of authenticity, reliability and genuineness. In a 1995 article dedicated to the themes of reliability and authenticity, she states that *reliability* refers to the “authority and trustworthiness of the records as evidence,” and “the ability [of the document] to stand for the facts they are about.” For an example, a reliable certificate of citizenship can be treated as a fact that the person holding the certificate, the person in question, is in fact a citizen. A document is reliable when its form is complete (discussed below, elements of documentary form), and it “...possesses all the elements required by the system in which it was created for it to generate the consequences recognized by the system itself” (Duranti, 1995). Reliability is also enhanced when the body of rules surrounding the procedure of creation is robust and established. Some of these rules refer to appropriate persons responsible for making records, and other rules refer to correct routing and handling in the course of their “compilation, completion, and their filing, as this operation determines the

record's documentary context" (Duranti, 1995) The more rigorous and detailed the rules, the more reliable are the resulting records.

Authenticity is a separate notion from reliability, and is related to a document's transmission – providing authentic copies, for example.

... a record is reliable when it can be treated as a fact of which it is evidence. By contrast, a record is authentic when it is the document that it claims to be. Proving a record's authenticity does not make it more reliable than it was when created. It only warrants that the record does not result from any manipulation, substitution, or falsification occurring after the completion of its procedure of creation, and that it is therefore what it purports to be.

Here is where the competing definitions of authenticity, reliability and genuineness come into play:

"Genuineness is the closest concept to truthfulness...documents that are trustworthy (that is, *reliable*) because of their completeness and controlled procedure of creation, and which are guaranteed to be intact and what they purport to be (that is, *authentic*) by controlled procedures of transmission and preservation can be presumed to be truthful (that is, *genuine*) as to their content. Thus, to those who make and preserve records, the two key concepts remain reliability and authenticity, as genuineness is embedded in them." (Duranti, 1995)

but three years later, in *Diplomatics: New Uses for an Old Science*,

"A Document is 'authentic' when it presents all the elements which are designed to provide it with authenticity. A document is 'genuine' when it is truly what it purports to be." (Duranti, 1998)

It is not in the scope of this paper to determine what the subtle differences between authenticity, reliability and genuineness might be. For the duration of this paper, I will be concerned with determining whether a system "presents all of the [documentary] elements which are designed to provide [a document] with authenticity [or reliability]", and whether a given system's procedures of creation and transmission are reliably documented and robust, providing a degree of authenticity or genuineness to the documents stored therein. How does this relate to provenance? I will hope to tie this all together by the end of the paper.

Intrinsic and Extrinsic Elements of Documentary Form

A document's form reveals and affects the function it serves, and close scrutiny of a document's elements allows for an understanding of the administrative actions that generated it. The diplomatic methodology is based on the assumption that, notwithstanding the differences in nature, provenance, or date, all documents present structures that are similar enough that it is possible to envision one typical,

ideal documentary form, making it possible to examine all of the elements within any given document. Once the diplomaticist identifies the presence or absence of all of the elements, and their variations, s/he can discern the administrative function of the studied documents, and make some distinctions between authentic or reliable documents..

In diplomatic parlance, “form” is the network of “rules of representation used to convey a message...the characteristics of a document which can be separated from the particular subjects, persons, or places which it concerns” (Duranti, 1998) The documentary form is both physical, referring to the physical structure of the document – known as the *extrinsic* elements of documentary form; and intellectual, referring to the document’s internal articulation, referred to the *intrinsic* elements of documentary form. Intrinsic elements make a document complete, and extrinsic elements are those that make it *perfect*, that is, capable of accomplishing its purpose.

Extrinsic elements are those that constitute the tangible composition of the original document and its external appearance. It is possible, preferable even, to examine these elements without reading the content of the document under question. These elements are: the medium (papyrus, vellum, ink types, wax medium, etc.), the script (discerning different hands, signatures, etc.), languages (use of rhetorical style), special signs, seals, and annotations. Study of the extrinsic elements is closely allied with paleography, although its aims are different. Whereas paleography is more concerned with gaining an understanding of social issues and developments through the close scrutiny of writing, diplomatics is concerned primarily with conferring authenticity. While diplomatics uses some of paleography’s intellectual tools to examine extrinsic elements and their various components, like inks, illuminations, graphic characters and seals; diplomatics only takes advantage of certain aspects of those components, and only for specific reasons.

Intrinsic elements are those that are “integral components of intellectual articulation.” They are the mode of presentation of a document’s content, or parts of the content which determine the “tenor of the whole.” Duranti posits that all documents “present an obvious typical structure” and “an ideal analytical sub-structure.” This ideal sub-structure is made up of three sections: 1) the *protocol*, which places the document within an administrative context, indicates the people involved in the document’s creation (the actor, recipient, notaries, presiding officials, etc.), and the initial formula (“To whom it may

concern:"); 2) the *text*, which contains the action of the document, including considerations and circumstances which were present at creation. It is the central part of the document, and from an administrative or historical point of view it is usually the most important part of the document because it represents its substance, its reason for existence. To a diplomaticist, however, the text is no more or less important than the other two sections; 3) the *eschatocol*, which is a form of documentation, and contains "the context of the action...the enunciation of the means of validation, and indication of the responsibilities for documentation of the act." It also contains the final formulae ("Sincerely and with best wishes,") (Duranti, 1998).

Purpose of Diplomatics: Evidence

Duranti states that most diplomaticists define their field as "the science which critically studies the document in order to determine its value as an historical source," thereby identifying the primary purpose of diplomatics as historical in nature. The origin of diplomatics, however, is linked to bureaucratic needs to determine the authenticity of documents, for the ultimate purpose of ascertaining the reality of rights or the truthfulness of facts represented in them. There is also a distinction between legal, diplomatic and historical authenticity: Legally authentic documents "are those which bear witness on their own because of the intervention, during or after their creation, of a representative of a public authority guaranteeing their genuineness;" diplomatically authentic documents "are those which were written according to the practice of the time and place indicated in the text, and signed with the names of the persons competent to create them;" and historically authentic documents "...are those documents which attest to events that actually took place or to information that is true." These differing views of authenticity are totally independent of each other – for example a Papal brief which does not contain the expression "datum...sub anulo piscatoris," may be legally and historically authentic, but diplomatically inauthentic.

When a person conducts a diplomatic investigation into a document or object, s/he is hoping to prove or disprove something about that document's validity, authenticity, and genuine-ness. In effect, s/he is trying to prove that a given object has actually lived the life it presents, either through its documentary forms, or through it's context – "is this a real Picasso?" "Is this really a letter from Albert Einstein?" "How can I get this document to tell me its story?" Diplomatics gives the researcher the tools necessary to uncover an object's provenance and history. Without a written document tracing a document's

movements from place to place throughout its life, applying the principles of diplomatics is often the only way to truly link together events from the past, and to give contextual meaning to any orphaned document. Within the physical world, orphaned documents are relatively rare – they are usually at the very least found in a specific place, at a specific time. In the electronic world, however, virtually every document is an orphan. There is no context, and the digital objects have no way to tell their story.

Again, what does all of this have to do with provenance? Diplomatics, in a very broad reading its terms and procedures, is concerned with determining functional provenance of documents: Where did any given document come from, what channels did the document travel? Who was responsible, at what time, for transmission and copying of the document? What is the document made of? What are its parts? Furthermore, diplomatics is highly informed by historical procedures and needs, as are archives; conferring authenticity, reliability and genuineness is simply a different and perhaps more robust way of talking about evidential value; and finally, studying the elements of documentary form are related to concepts of preservation – both digital and physical. However, in the current atmosphere, digital preservation, its nuances and needs in the area of archival practice are of highest importance.

Provenance as Purveyor of Authenticity and Evidential Value

Margaret Hedstrom is concerned with articulating the nuances of archival practice in the electronic environment (Hedstrom, 1991; Hedstrom, 2002). In a 1994 paper, she deals extensively with the Principle of Provenance in relation to the concepts of integrity and preservation of electronic records (Hedstrom, 1994). She defines archives as institutions that “organize and manage records to protect their integrity as documentary evidence and to provide successive generations with materials for fresh interpretations of collective actions and consciousness.” Preservation in an archival setting includes retaining a record’s content and the ability to reproduce its structure, but also to provide “linkage between the archival document and related records, the record’s creator and recipient, the function or activity it derived from, and its place in a larger body of documentary evidence.” These linkages provide contextual or provenancial information, and hence authenticity, and are, in the physical realm, captured through the physical organization of archives. Instead of being primarily concerned with the formal and content-oriented characteristics of a record, like Dollar, or the functional roles a record might have played within an organization, like Bearman, Hedstrom is concerned primarily with protecting those characteristics of

records which establish their reliability, authenticity, and utility as evidence, both legal and scholarly. The Principle of Provenance, then, is a means to an end. Provenancial and contextual information help define authenticity; authenticity confers evidential value to a record; archives exist to provide records with evidential value; so provenancial information must be preserved. Instead of specifically discussing provenance, Hedstrom focuses on protecting the evidentiary nature of archives through the development, once again, of standards and requirements for software and hardware applications and systems.

This idea of protecting evidential quality of records through careful preservation of context is an idea, like the Principle of Provenance in the late Eighteenth Century, whose time has apparently come. One of the most powerful new ideas in the archival community, The Open Archival Information System Reference Model (OAIS), uses the idea of context and provenance as integral parts of its scheme. The remainder of this paper will provide a brief review of the OAIS model, discuss the use of context and provenance within the schema, and elucidate the relationships between OAIS and traditional archival theory, as exemplified by diplomatics.

The Open Archival Information System (OAIS)

History

The OAIS reference model provides a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term. It was developed by an initiative spearheaded by NASA and its Consultative Committee for Space Data Systems (CCSDS), which was created in 1982 to collaboratively develop standards for data supporting space research. In 1990, the CCSDS became the working body for ISO Technical Committee 20 (TC 20 – Aircraft and Space Vehicles) – Subcommittee 13 (SC 13 – Space Data and Information Transfer Systems), with the intention that its recommendations could undergo normal ISO review and eventually become true ISO standards. At the time of this merger, the ISO suggested that TC 20 / SC 13 should develop archive standards that would address data used in conjunction with space missions, as well as immediate and long-term storage of this digital data (Sawyer & Reich, 2003).

So at its inception, the TC 20 / SC 13 was compelled to develop an archival standard for highly complex, inter-dependent data. Their first response was to acknowledge that: 1) there was no existing, widely recognized framework for developing specific digital archive standards; 2) Broad participation

among institutions was desirable – including both archival and non-archival institutions as well as space related and non-space related organizations. “All participation is welcome!”; 3) Development of a “Reference Model” to establish common terms and concepts would be the first order of business. This reference model would identify common terms and concepts; provide a framework for clarifying the significant relationships among entities in the archival system; and serve as the foundation for the development of standards supporting the archival environment. Finally, this framework would focus on data in electronic forms, recognizing that other forms exist in most archives. This framework, the OAIS Reference Model, was released as a draft recommendation in May 1999 (Lavoie, 2000).

System Basics

Any organization seeking to preserve and make accessible information over the long term will benefit from the OAIS model. The reference model presents the conceptual design for an OAIS archive, including primary components and associated functions and relationships, which support the requirements set forth for OAIS institutions; that is, organizations wishing to be recognized as OAIS compliant must agree on and enforce rigorous metadata standards; their systems must be robust and sustainable; they have to understand the scope and needs of their “Designated Community,” which includes knowing their users’ information strengths and weaknesses; and these organizations must provide to their users and follow themselves documented policies and procedures ensuring that information is preserved against reasonable unforeseen disasters – which will also ensure that the delivered data will be authentic and traceable to the original (Lavoie, 2002).

The system environment is based on the interaction of four entities: the information producers, the information consumers, the information managers and the archive itself. Information is defined as “any form of knowledge that can be exchanged,” and it comes in two forms – physical or digital – both labeled the “data object.” System users interpret the data objects as meaningful or appropriate through the interplay between the community’s inherent knowledge base, and the representation information associated with each data object. The “representation information” is that supplemental information necessary for the user to understand the archived information (Lavoie, 2002).

Each data object in an OAIS system is called an “information package” and is made up of four types of “information objects”: 1) *Content Information* is defined as the primary information of interest, and

made up of the data object and its associated representation information; 2) *Preservation Description Information* is that information necessary to preserve its associated Content Information, specifically including information related to provenance; 3) *Packaging Information*, which brings all of the various components together into an identifiable and unique entity; and 4) *Descriptive Information*, which is information related to search and retrieval by the user community. There are also three types of information packages in this system: 1) The Submission Information Package (SIP), which comes from the information producer to the archive; 2) the Archive Information Package (AIP), which is that information package actually stored by the archive, and; 3) the Dissemination Information Package (DIP), which is the information package transferred from the archive in response to a user request (Lavoie, 2002).

Finally, within each OAIS entity, there are five “functional entities:” 1) The **Ingest Function** receives information from producers and prepares it for storage in the archive. It accepts information in the form of SIPs, performs checks on the SIP to ensure data standard compliance, generates an AIP from one or more of the SIPs, and extracts Descriptive Information – metadata for search and retrieval, thumbnail images for browsing, etc. – from the AIPs, and finally transfers the newly created AIP to the Archival Storage Function and the associated Descriptive Information to Data Management Function. 2) The **Archival Storage Function** is responsible for storage and maintenance of AIPs. These responsibilities include migration to new media as required, error checking, implementing disaster recovery strategies, and providing copies of requested AIPs to the Access Function. 3) The **Data Management Function** is primarily responsible for keeping the system up and running, coordinating the descriptive information tying together the archive’s AIPs, executing query requests and generating result sets for the “Access Function,” and creates reports and performs updates. 4) The **Administration Function** manages day to day operation of the archive, which includes auditing incoming SIPs to assess compliance with agreed upon standards. It also serves as the interface between the archive and the two components of the OAIS environment – archive management and the designated community. 5) The **Access Function** assists users in identification and acquisition of relevant information in the archive, and delivers that information from the archive to the user. This requires a single user interface to the archive’s holdings for both search and retrieval purposes, it must generate a DIP in response to a user request and

obtain copies of the appropriate AIPs from Archival Storage, get the appropriate Descriptive Information from Data Management Function, and deliver the results to users (Lavoie, 2002).

Use of Provenance in OAIS

In an OAIS system, provenance information is encased in the Preservation Information Object, which is that information “necessary to adequately preserve the particular Content Information with which it is associated. It is specifically focused on describing the past and present states of the Content Information, ensuring it is uniquely identifiable, and ensuring that it has not been [knowingly or] unknowingly altered.” It is the second major component of the information package, and focuses on the information necessary to manage the perpetual existence of the data object and its content over a long period of time. Whereas the Content Information records the static properties of an archived object, the Preservation Information, which also includes unchanging properties, emphasizes the temporal characteristics of the object, extending from its creation to its ingest into the digital archive, to its retention in the “archival store,” or long term storage (Lavoie, 2002).

Preservation Description Information is composed of four categories of information. 1) **Reference** describes the identification systems and the means by which an object is uniquely identified both within a system and externally to users; 2) **Context** documents the relationships of the Content Information with its environment, including the reasons for the object’s creation and its relationships with other Content Information objects; 3) **Provenance** documents the history of the Content Information, including its origin, changes to the object or its content over time, and its chain of custody; 4) **Fixity** provides the Data Integrity checks and Validation/Verification keys used to ensure that the particular Content Information object has not been altered in an undocumented way. All in all, the Preservation Description Information records the identity, relationships, history, and integrity of the archived Content Data Object (Lavoie, 2002).

Context Information

Context Information records two separate pieces of information: First, the system requires information regarding the motivation or rationale for creation of the Content Data Object; and second, it documents significant relationships between the primary Data Object and other Data Objects in the system. This information can take the form of two broad categories – 1) either other manifestations of the

Content Data Object, such as alternate software formats (html, .pdf, and Word versions of the same document) or additional versions of the document in the same software format (Word 6.0 > Word 97 > Word 2000 > Word XP – this is called **Manifestation Content**); or 2) Content Data Objects whose intellectual content is related to that of the primary Object. This includes those Data objects that either form a well-defined series or collection, or whose intellectual content describes, elaborates on, critiques that of the primary Data Object. It would also include a set of objects whose content, in aggregate, forms a simple complex Object at some higher level of abstraction (like a set of .pdf documents forming the chapter of a book – this distinction is referred to as **Intellectual Content**) (Lavoie, 2002).

Recording the context information is likely to be a subjective exercise, dependent on the needs and requirements of the Archive and/or the Designated Community. Additionally, the framework states that this information is for management of the preservation process only, and specifically NOT aimed at facilitating understanding or interpretation of the Content Data Object's intellectual content. This functionality is addressed by the metadata elements within the Data Object's Representation Information. Finally, the documentation of these relationships should be considered critical to success of the system. Successful linking of Manifestation or Intellectual Content groups will promote simultaneous migration, and dissemination as a unit (Lavoie, 2002).

Provenance Information

Provenance Information “documents the history of the Content Information. This tells the origin of the source of the Content Information, any changes that may have taken place since it was originated, and who has had that custody of it since it originated.” It specifically addresses the temporal characteristics of the archived Data Object, beginning with the Object's creation and extending to its current state as an object in the archive. Objects are considered as dynamic entities with the current archived state seen as the “culmination of an evolutionary process,” of which the time spent in archival retention might only be a small part. In addition to recording the chronological life of the Data Object, Recording the particulars of events in the life cycle of the Data Object like creation, transfer of ownership, ingest into the system, and migration from one format to another; and comprehending the impact that these events had on the Data Object is an additional function of Provenance Information.

There are five sub fields within the Provenance Information category: 1) Origin: which describes the process by which the Data Object was created, i.e., digital scanning, digital recording of audio material, born-digital, or etc.; 2) Pre-Ingest: describes the circumstances surrounding transfer of the intellectual property rights associated with the Object from one entity to another; 3) Ingest: describes the processes associated with preparing an Object for inclusion in the “archival store;” 4) Archival Retention: describes the steps taken to translate the Object from one digital format to another (this is known as *transformation*, in the terminology of the OAIS system); and 5) Rights Management: documents the archive’s current ability to make the Object available to the archive’s users, including any restrictions regarding access and use of the Object by users.

As discussed above, Provenance Information captures temporal, dynamic characteristics of the Object and has two hierarchical structures. First the system requires definition of the major stages of a Data Objects life cycle (creation, pre-ingest, ingest, etc.); and second, the system wants documentation of specific events within those stages, for example, migration of an object from one format to another within the Archival Retention Stage, or “archival retention decision” within the Ingest phase. An related issue is the fact that an Event like migration could result in the generation of a new Content Data Object, itself preserved in the archival store, and the archival retention decision might include information regarding rights management and hence have some bearing on that phase of the Object’s life-cycle.

In this system, Provenance Information becomes most complex and challenging when considering the network of metadata necessary to represent objects logically and consistently. The Framework document suggests that the best way to consider the issue is to think of Provenance Information metadata as, “‘transcending’ specific physical manifestations of a Content Data Object (e.g. in alternate formats), and instead applying at something like a “work” level to the entire collection of related Objects...such an approach, however would require a clear definition of a “work,” perhaps something similar to what is proposed in the FRBR model.” (Lavoie, 2002) (Note: the FRBR model is a very interesting one, with distinct references and interest in providing clues to context and provenance of a work. Save it for a later day.)

A Comparison: The Use of Provenance in Diplomatics and the OAIS Model

Ascertaining the provenance of any given object is paramount in determining its authenticity, and authenticity is absolutely necessary if an object is going to be used as evidence for an argument, or in legal proceedings, for example. Proving authenticity, or in this sense, genuineness, is also important when transacting business – proving that a given object *is* what it purports to be, either the object being bought or the currency with which it is being paid for. The medieval practice of diplomatics defined a programmatic schema for determining authenticity, outlined above (Section 2). This schema is so robust and influential, that the properties used therein – trying to date and place the object irrespective of obvious physical clues, have become so ingrained in our culture as harbingers of authenticity, that they are used widely and usefully for a number of purposes: Literary critics for example use facets of diplomatic practice to date an author's works; forgery specialists use the scientific application of paleography and dating techniques to track down fakes and forgeries; and historians use the principles of diplomatics to place an object squarely in its historical context.

Whether they knew it or not, the authors of the OAIS System were using diplomatic concepts when developing their reference model. Not surprisingly, they placed the information modules related to provenance in the “preservation information” section – because to genuinely preserve an object is to preserve its context, intellectual meaning, and authenticity.

By breaking down the idea of context into two sections, “manifestation information” and “intellectual information,” the framework authors were, in effect, dividing the object's characteristics into intrinsic and extrinsic qualities. This is, I'll admit, taking a very broad view of diplomatic principles. Whereas diplomatics is concerned with determining the authenticity of one given document, the OAIS system is concerned with preserving hundreds or millions of documents in a trustworthy manner. There are of course differences, mostly having to do with diplomatics' focus on an object versus the OAIS focus on an Object's representation. However, generally speaking: in diplomatics, determining a document's intrinsic and extrinsic qualities allows the scholar to place an object within a continuum, or within a context. If the document purports to be a family deed, but does not conform to “family deed” strictures of the period in which it was written (intrinsic qualities) – that document is either an anomaly or a fake – but the difference has been noted, and the relationship between a “typical” document and the “given”

document has been assessed. The scholar is one step closer to understanding the context of that document's creation. In the OAIS model, context is provided in the context module, and both intrinsic qualities, which are related to intellectual information; and extrinsic qualities, which are related to manifestation information, can be recorded, and the relationships between documents can be determined at the outset. For example, if in the year 2050 a scholar is using an OAIS system to look at a ".pdfx10" of a document authored in the year 1988 – a year that ".pdfx10" wasn't in existence – will be able to review the Object's "manifestation information" from the Context Information module to determine the Object's original medium, or software application. This module will show the migration events in the document's life cycle, and the scholar will be closer to determining the context of the document's creation, use, and storage. Further, the context information module is also concerned with providing an intellectual context to the document. This is dealt with by the "intellectual information module," which could list intellectually related works, critiques, and addendums. This is related to the diplomatic idea of the extrinsic qualities of determining a document's language, medium, seals, and annotations.

The provenance information in the OAIS system is generally concerned with providing a time-line of events in the Object's life. These events will often feed into the context information module, providing even more context in the birth, use, and storage of any Object within the system, and is most closely allied to the tertiary definition of provenance in the archival sense – "collecting information on successive transfers of ownership or custody of a particular paper or manuscript." In fact, the second phase of an Object's life cycle is "pre-ingest," is defined as "the circumstances surrounding transfer of the intellectual property rights associated with the Object from one entity to another." This is straight archival provenance, and the idea that the events don't end when the Object enters the system is also interesting. In the OAIS model, an object is never stagnant or fixed, but is being acted upon continuously. It's a vibrant and real-life system, which records the events of vibrant Objects.

After reading the literature on diplomatics and the OAIS reference model, I came to view the OAIS model as a kind of automated diplomatics, not in the sense that the OAIS is calling for the recording of specific diplomatic information, but in the sense that the OAIS authors were using diplomatics as a kind of conceptual model that would allow their system to confer authenticity on any given object. Additionally, the intermingling of provenance and context, that the provenance information feeds into the context

information, conferring conceptual weight, and hence authenticity on an Object, is attractive and is reminiscent of the complexity of diplomatics.

Further Research Possibilities

This promises to be the first of many papers on the subjects of provenance, context, authenticity (or reliability, or genuineness), and digital preservation. There are numerous topics I would have liked to explore but could not, within the scope of this paper. They are:

- A more detailed look at Luciana Duranti's definitions of diplomatics and her changing definitions of authenticity, reliability and genuineness. I am sure there is some reason for her changing vocabulary but at the moment can't see it.
- A review of diplomatics in relation to the InterPARES proposal
- A more complete review of diplomatics in relation to the OAIS model (I was too focused on the Principle of provenance)
- A comparison of the InterPARES and OAIS.
- The use of FRBR (Functional Requirements of Bibliographic Records) as a bibliographic standard – how FRBR uses the dictums of provenance to confer representative completeness
- The development of the EAC (Haworth, 2001) standard, and its relationship to diplomatics and provenance.
- Comparison of the ideas of providing “context,” and recording “provenance.” Sometimes these terms are interchangeable, sometimes there's an unspoken distinction.

References

- Bearman, D. (1989). *Archival Methods: Archives and Museum Informatics Technical Report*. Pittsburgh: Archives and Museum Informatics.
- Bearman, D. (1994). *Electronic Evidence: Strategies for managing records in contemporary organizations*. Pittsburgh: Archives and Museum Informatics.
- Bellardo, L. J., & Bellardo, L. L. (1992). *A glossary for archivists, manuscript curators, and records managers*. Chicago: Society of American Archivists.
- Boyle, F. L. (1976). Diplomatics. in J. M. Powell (editor), *Medieval studies: An introduction*. Syracuse: Syracuse University Press.
- Dearstyne, B. W. (1993). *The archival enterprise: Modern archival principles, practices, and management techniques*. Chicago, IL: American Library Association.
- Dollar, C. M. (1992). *Archival Theory and Information technologies: The impact of information technologies on archival principles and methods*. Macerata: University of Macerata Press.
- Duranti, L. (1995). "Reliability and Authenticity: the Concepts and Their Implications." *Archivaria*, 39, 5-10.
- Duranti, L. (1998). *Diplomatics: New uses for an old science*. Lanham, Md.: Scarecrow Press.
- Gränström, C. (1994). "The Janus syndrome." *The Principle of Provenance*. Stockholm: Swedish National Archives.
- Haworth, K. M. (2001). "Archival description: Content and context in search of structure." in D. V. Pitti, & W. M. Duff (editors), *Encoded Archival Description on the Internet* (pp. 7-26). Binghamton, NY: Haworth Press.
- Hedstrom, M. (1991). "Understanding electronic incunabula: A framework for research on electronic records." *American Archivist*, 54(Summer), 334-354.
- Hedstrom, M. (1994). "Electronic archives: Integrity and access in the network environment." in *Networking in the Humanities: Proceedings of the Second Conference on Scholarship and Technology in the Humanities* (pp. 77-95). London: Bowker-Saur.
- Hedstrom, M. (2002). "Archives, memory and interfaces with the past." *Archival Science*, 2, 21-43.
- Horsman, P. (1994). "Taming the elephant: An orthodox approach to the Principle of Provenance." *The Principle of Provenance*. Stockholm: Swedish National Archives.
- Lavoie, B. (2000). "Meeting the challenges of digital preservation: The OAIS reference model." *OCLC Newsletter*, (January / February), (2002). Dublin, OH.: OCLC Online Computer Library, Inc.pp. 26-30.
- Reimer, S. R. (1998) *Manuscript Studies : Medieval and Early Modern* [Web Page]. URL <http://www.ualberta.ca/~sreimer/ms-course.htm> [2003, December].
- Sawyer, D. & L. Reich (2003). "ISO Reference Model for an Open Archival Information System (OAIS) Tutorial Presentation." http://ssdoo.gsfc.nasa.gov/nost/isoas/presentations/oais_tutorial_200306.ppt : NASA.