

Provenance as a Knowledge Organization Principle[†]

Natalia Tognoli* and José Augusto Chaves Guimarães**

*Universidade Federal Fluminense (UFF), Instituto de Artes e Comunicação Social (IACS),
Rua Lara Vilela, 126, São Domingos, Niterói, RJ 24210-590, <nataliatognoli@id.uff.br>

**Universidade Estadual Paulista (UNESP), Centro de documentação histórica e universitária de Marília
(CEDHUM), Rua Iara Clube, 85, Jardim Maria Izabel, Marília -SP, 17515-330. <chaves.guimaraes@unesp.br>



Natália Bolfarini Tognoli is Professor at the Department of Information Science at Fluminense Federal University (UFF), Brazil, and at the Information Science Graduate Program (PPGCI) at the same institution. She is a member of the Brazilian ISKO Chapter. Her main research interests include archival knowledge organization, contemporary diplomatics, social epistemology, and social justice in archives.

José Augusto Chaves Guimarães is Professor at the Department of Information Science at São Paulo State University, Unesp, Brazil. He is 1-B CNPq researcher, founder and former president of the Brazilian ISKO Chapter, and member of ISKO Executive Board (2012-2016). His main research interests include knowledge organization, epistemology of information science, information ethics, juridical information, bibliometrics, and scientometrics.



Tognoli, Natalia and José Augusto Chaves-Guimarães. 2019. "Provenance as a Knowledge Organization Principle." *Knowledge Organization* 46(7): 558-568. 47 references. DOI:10.5771/0943-7444-2019-7-558.

Abstract: The principle of provenance is one of the most important milestones in archival practice and theory from the time its establishment grounded the scientific dimension of archival discipline in the nineteenth century. Since then, provenance and document context have supported the organization of archival knowledge (especially through classification and description procedures). Such relationships were gradually refined over the years and from different experiences between European archives and their classification and ordering systems. Historically, the principle of provenance is a pivotal moment in the development of archival theory, crucial to understanding the nature of records and archives. However, in archival theory, the principle of provenance still does not correspond to a single term or a single definition and scarce normalization terminology remains one of the problems of archival science, which leads to a lack of consensus about the division between the two principles of provenance and original order. Recently, the concept of provenance has been addressed by many other disciplines (law,

library and information science, computer science and visual analytics) and applied to different domains (cloud-based storage, preservation of digital records, digital evidence, digital humanities, e-science, open data, linked data, knowledge organization and indexing). As the use of provenance reaches new domains it is no longer just an organizing principle but also a means of reaching for authenticity and reliability of data and objects in digital environments or museums or to reestablish the original organic relationship in library collections.

Received: 25 January 2019; Revised: 26 August 2019; Accepted: 29 August 2019

Keywords: provenance, archival, principle, records, fonds, order

[†] Derived from the article of similar title in the ISKO Encyclopedia of Knowledge Organization, Version 1.0 published 2018-11-07. Article category: Core concepts in KO.

1.0 Introduction

Provenance is a complex concept that has been a focus of archival science for many years. We present a brief history of the principle in the archival domain and its correlated concepts and how it has recently been addressed by other disciplines and applied to different domains, especially knowledge organization (KO). We also discuss different points of view on provenance based on archival science theoretical landmarks, and how it can be represented to guarantee authenticity and reliability of data in complex and interconnected systems.

2.0 Provenance as an organizing principle: some historical remarks

The principle of provenance is one of the most important milestones in archival practice and theory from the time its establishment grounded the scientific dimension of archival discipline in the nineteenth century. Since then, provenance and document context have supported the organization of archival knowledge (especially through classification and description procedures). However, what is meant by provenance?

The International Council on Archives (ICA) (2007,11) has defined provenance as:

the relationships between records and the organizations or individuals that created, accumulated and/or maintained and use them in the conduct of personal or corporate activity. Provenance is also the relationship between records and the functions which generated the need of the records.

Such relationships were gradually refined over the years and from different experiences between European archives and their classification and ordering systems. As stated by Brenneke (1968, 78), the theoretical origin of the principle goes back to the work of Ernst Spiess, *Von Archiven*, from 1777, according to which it was necessary to “hear the voice” of the documents to know how they should be arranged and what sort of order they had in them. It was an inductive method where, in the end, each fonds would determine its content and method of organization. “Such method, to which the future belonged, emerges here for the first time as a lightning bolt: it is from this point onwards that the principle of provenance will be born.”

The opening of the archives to the citizen in his day-to-day life as well as to support historical research started in France with the creation of The National Archives in 1794, as well as discovery of primary sources, made the nineteenth century a landmark for both history and its so-called auxiliary sciences, with special emphasis on archival science. As a consequence, such an historical slant imposed on archives contributed to an internal reordering of documents that had roots in historical thought. Thus, between 1830 and 1850, from traditional arsenals of power, archives became laboratories of history, with all the consequences that this entails, especially in terms of what concerns “artificial” criteria.

However, it was only in 1841, through a service instruction (*Instructions pour mise en ordre le classement des archives départementales et communales*), that the principle of provenance started to take shape and to be delineated theoretically from the concept of fonds. On April 24th, the instruction proposed by Natalis de Wailly (a French bureaucrat and head of the administrative section of the departmental archives of the Ministry of Interior of France), and approved by minister Duchatel suggested “to gather together by fonds, that is to unite all the deeds (i.e., all the documents) which come from a body, an establishment, a family, or an individual, and to arrange the different fonds according to a certain order” (Duchain 1983, 64). According to the principle of *respect des fonds*, all records which originated from any particular institution should be grouped and should be considered the fonds of that par-

ticular institution; records within fonds should be arranged by subject-matter groups, and each group should be assigned a definite place in relation to other groups; items within the subject matter-groups should be arranged as circumstances might dictate, either chronologically, geographically or alphabetically (Schellenberg 2003, 170). The principle determined provenance at the macro level, taking into account only the nature of the institutions which created the records, not giving importance to the administrative units or to the original order of creation and accumulation of records, as was to be stated later, by the joint approach of the principles of provenance (*Provenienzprinzip*) and the original order (*Registaturprinzip*).

In 1881, the Prussia State Archives issued a regulation, establishing a new system for the organization of records. In the document drawn up by the then state archivist Max Lehmann and approved by the director of the archives, Heinrich von Sybel, the *Provenienzprinzip* is enunciated—based on the principle of *respect des fonds*—according to which public records should be grouped following the administrative units that created them (rather than according to the nature of the institutions that created them). The main divisions within the state archives were to be formed by separating records originating with the various administrative units of the government (Schellenberg 2003, 175).

In the same document, the principle *Registaturprinzip* was announced, according to which the records of every agency should be maintained in the archival institution in the order given them by the registry office and should not be reorganized by subject-matter groups. The registry offices were government administrative units responsible for controlling the incoming and outgoing records made or received by an entity. According to Schellenberg (2003, 66) “these offices were established to handle the documentary work with the expansion of government activities after the establishment of the modern kingdoms. Today they are responsible to receive, record, distribute, and archive the official mailing, and receive different names and attributions according to the country.” Later, due to its extension to other countries, especially to The Netherlands, the *Registaturprinzip* came to be called “the principle of respect for the original order.”

It is important to say that the principle of provenance is not an isolated construction, as a solution to a particular problem. Instead, the principle is constituted from the experiences of several countries facing similar problems but of different realities. However, we should consider the importance of the French, German, and Dutch authors to the theoretical construction of the principle. As pointed out by Nesmith (2015, 87) in his entry about the principle of provenance in the *Encyclopedia of Archival Science*,

German archivist-historians refined the principle in two ways—emphasizing the importance of the relationship of a fonds to its specific administrative creator (which they called the *Provenienzprinzip*) and preserving the original order of the records within fonds (the *Registaturprinzip*). Dutch archivists made a pivotal contribution to consolidation of the principle of provenance in 1898 with the publication under the auspices of the Netherlands Association of Archivists of *The Manual for the Arrangement and Description of Archives*.

Additionally, according to Nesmith (2015), the *Manual* not only accepted the French emphasis on the origin of fonds and the two German components of the principle of provenance but also influenced theory and practice of all major Western countries, which started to apply the principle of provenance to the arrangement of their archives. The Dutch *Manual* gave the final sanction to this theory and the change in theory brought about a change in the character of archival institutions and the work of archivists (Posner 1940).

We can presume from this brief historical review that the principle of provenance is a pivotal moment in the development of archival theory. Its construction was crucial to understanding the nature of records and archives, mainly after the French Revolution, with the redefinition of the meaning of documents and its institutions.

2.1. Some terminological issues: principle of provenance, *respect des fonds*, original order, or record group?

In archival theory, the principle of provenance still does not correspond to a single term or a single definition, as pointed out by Martín-Pozuelo (1996). The scarce normalization terminology remains one of the problems of archival science, which leads to a lack of consensus about the division between the two principles of provenance and original order. For some authors (Schellenberg 1965; Evans 1991), maintaining original order is another principle, whereas, for others, it represents an extension or a second degree of provenance (Carucci 1990; Brenneke 1968).

The principle of provenance assumes different conceptual and terminological meanings in different countries, revealing different theoretical concepts and, often, unfolding in other sub-principles depending on the context. According to Horsman (1994), the archival terminology about the principle of provenance is a “Babel tower.” In Horsman’s point of view, the principle of provenance is the only principle of archival theory acting in two dimensions: an external called *respect des fonds*—to respect the archives as it was created by a person, group, or entity as a whole—and an

internal respecting of the original order given to the records by its creators. Both dimensions should act as an inseparable whole.

Considering that the epistemological status of a particular area is also conditional on the quality and clarity of its terminology, Tognoli et al. (2016) analyzed the definitions and translation equivalents in Portuguese, Spanish, Italian, French, and English in the International Council on Archives (ICA) Multilingual Archival Terminology,¹ regarding archival principles, with emphasis on provenance. The results have shown a distinct confusion between provenance, *respect des fonds*, and original order.

In some European countries, especially Italy and France, it is possible to perceive degrees of provenance regarding the producer entity or concerning the original order that was given by the records creator. In the Brazilian archival tradition, the principle of provenance is clearly separated from the principle of original order. The first one is understood as a basic principle of archival science, according to which the record produced by a collective entity, person or family must not be mixed with those of another producing entity. For the principle of original order, the Brazilian terminology understands that the archives must keep the arrangement given by the collective entity, person or family that produced it (Bellotto 2004).

In Canada, the *Rules for Archival Description (RAD)* (2008, xxiii) define *respect des fonds* as a principle composed of two parts, provenance and original order:

The principle of provenance means that the records created, accumulated and/or maintained by an individual or organization must be represented together, distinguishable from the records of any other creator. The principle of original order means that the order of the records established by the creator should be maintained by physical and/or intellectual means whenever possible to preserve existing relationships between records and the evidential value inherent in this order.

In the United States’ archival tradition, the principle of *respect des fonds* is also the basis of arrangement for records, but the fonds were replaced by record groups, defined as a hierarchical division that is sometimes equivalent to provenance, representing all the records of an agency and its subordinate divisions. However, a large agency with a lot of records may have its records subdivided into several record groups by treating the records of different divisions as separate collections rather than as a series (Pearse-Moses 2005). The record group is the primary unit for archival arrangement and description, serving to reinforce the archival concern with provenance. According to Trace (2015, 314) the concept feeds into the notion that there are

certain contexts that must be captured to ensure that bodies of records can serve as evidence of their creator’s functions and activities. According to Scott (1966, 493) “the concept of ‘group’ provides the required administrative context for archives and is used as a basis for (a) the descriptive inventory, (b) the physical arrangement of archives (i.e., the so-called shelf group), and (c) the numerical control of archives.”

In order to illustrate the problem of scarce normalization of terminology in the archival science domain regarding the discipline’s most important principle, we provide a compendious survey of definitions of provenance in different archival traditions (Table 1).

The definitions presented in Table 1 reinforce the European tradition of one principle, two degrees: provenance regarding creator and concerning original order. On the other side, the new world archival traditions—represented here by the United States, Brazil, and New Zealand—work with two organizing principles: provenance and original order, which together support the organization of archival knowledge. Canada seems to be the exception, working with two levels of the same principle.

The principle of provenance has been differently developed from one country to another because of their different

administrative cultures. Despite a terminological or conceptual confusion presented in archival literature, it is possible to say that archival documents (or records) are the products of activities developed by a person or entity in the execution of a function. Consequently, to understand this process means to perceive the network of relations between objects (records, archival documents), agents (creators either individuals or institutions), and functions (the necessary actions to the accomplishment of attribution within the scope of an administrative structure), as determinative elements to recognize the archival bond. That is why we can never conceive a record as an isolated element, especially because the recognition of its provenance allows it to be used as evidence of activities. Duchein states (1983, 67):

Consequently, to appreciate a document, it is essential to know exactly where it was created, in the framework of what process, to what end, for whom, when and how it was received by the addressee, and how it came into our hands. Such knowledge is possible only to the degree to which the whole of the documents which accompany it have been kept intact, quite separate from and without confusion with documents of different origins.

Countries	Definitions		
	Principle of Provenance	Principle of <i>Respect des fonds</i>	Principle of Original order
Brazil	Principle according to which the record created by a collective entity, person or family must not be mixed with others of another creator (Bellotto 2004)		The archives must keep the arrangement given by the collective entity, person or family that produced it
Canada	The person(s), family (families), or CORPORATE BODY (bodies) that created and/or accumulated and used RECORDS in the conduct of personal or business life; see also <i>Respect des fonds</i> (RAD 2008)		The principle that the RECORDS of a person, family or CORPORATE BODY must be kept together in their original order, if it exists or has been maintained, and not be mixed or combined with the records of another individual or corporate body. (RAD 2008)
France	Two degrees of provenance: regarding the producer entity and concerning the original order that was given by the records creator		
Italy	Also known as <i>metodo storico</i> or <i>ricostituzione dell'ordine originario</i> . Two degrees of provenance: regarding the producer entity and concerning the original order that was given by the records creator		
The United States	The principle of provenance or the respect des fonds dictates that records of different origins (provenance) be kept separate to preserve their context (Pearce-Moses 2005)		Original order (also registry principle, <i>respect for original order</i> , <i>l'ordre primitif</i> , <i>respect de l'ordre intérieur</i>), n. ~ The organization and sequence of records established by the creator of the records (Pearce-Moses 2005).
New Zealand	The agency, office, or person of origin of records, i.e. the entity which created, received, or accumulated, and used the records in the conduct of business or personal life. Also referred to as records creator (Archives New Zealand 2003).		The order in which records and archives were kept when in active use, i.e. the order of accumulation as they were created, maintained and used. The principle of original order requires that the original order be preserved or reconstructed [where possible] (Archives New Zealand 2003).

Table 1. Definitions of provenance in different archival traditions.

In other words, to appreciate a document by analyzing its provenance means recognizing its context of creation. Such a search for the contextual elements, as discussed by Guimarães and Tognoli (2015), can be considered a special dimension of domain analysis. It is important to recognize that the enunciation of the principle of provenance and the concept of *fonds* brought new perspectives for the work of the modern archivist, who became free from further interpretations, since the relationship that the document establishes with its creator lingers as the major and most significant means to analyze it, under archival procedures.

3.0 Interpretations of provenance in the archival domain

Today it is possible to say that the principle of provenance is commonly accepted as the basis for archival knowledge organization (being fundamental to arranging and describing archival documents by respecting their *raison d'être*). However, in the last thirty years, the application of the principle has been constantly reviewed and discussed in the archival literature because of the increase of information and communication technology responsible for changing the way records are created and used. Considering such a landscape, we provide a set of definitions of “provenance” in archival science from different theoretical perspectives (as physical and conceptual construct, as social-historical context, and as the creator and records history), which encompass the principle in the twentieth century, in order to introduce the views of some landmarks of archival literature.

In the late 1980s, the archivist Hugh Taylor announced a paradigm shift in the archival field calling the discipline to rethink its theoretical basis and traditional practices aiming to survive in this new context of knowledge production. Although its importance and precision have never been contested, we could say that facing new perspectives of creating, organizing, and representing processes, the principle also goes through changes that should be considered in the moment of archival knowledge organization. Duffy and Harris (2002, 268) wrote:

In modern bureaucracies, it is common for the same records to be created, accumulated, and used by numerous, different, successor or parallel agencies. Records emanate from business activities and in turn are used to support and carry out other business activities. This reality has led numerous archivists to suggest that the multi-faceted aspects of provenance are eroded when archival practice dictates the creation of fonds-level description and credits the creation of the records (and thus provenance) to one, and only one, individual or organization.

Therefore, there is a need to connect the record to all of its contexts by focusing on all of the relations that were established between it and the entities that use it in such a way that the network between objects, agents and functions must be represented entirely. This conception enlarges significantly the notion of context of a record, going much beyond the mere conception of physical production to reach other instances, e.g., the user.

Although the rediscovery of provenance can be considered a recent debate, its origins date back fifty years to the time when Peter Scott (1966) defended, in a seminal article entitled “The Record Group Concept: A Case for Abandonment,” the idea that provenance should encompass more than just the physical arrangement of the *fonds*, by reaching the representation of the many-to-many relationships of records. According to Cunningham (2016), while all archivists agree that provenance is a defining feature of archives, the reality of provenance is poorly understood, once based on a simplistic assumption that there is a mere one-to-one relationship between a given provenance entity and a given body of archives. Still, according to Cunningham (50) the responsibility for such a “narrow view of provenance” is the Dutch trio of archivists, Muller, Feith and Fruin, whose “rigid adoption of the rules for arrangement and description led archivists to believe stubbornly that a given body of archives could only ever have one provenance—a belief that simply does not reflect reality.”

In 1964, Scott suggested to Australian archivists that they might abandon the record group as the locus of intellectual control, advocating instead in favor of function-based series as a means of controlling records (497): “An obvious solution is to abandon the record group as the primary category of classification and to base the physical arrangement of archives on the records series as an independent element not bound to the administrative context.” To Scott, this approach was the most efficient way to record the complex nature of provenance. However, the description system of the Australian National Archives since the publication of Scott’s articles about the abandonment of the record group concept in 1964, ignores the record group concept of describing records series as the primary level of classification and the item as the secondary level. According to Scott there are some problems presented in the application of the record group concept to archives (e.g., when the record group is created from records that have been transferred from another agency, there is a risk of losing the creation original order. A parallel situation can occur when similar series are created by different agencies in the same organization, and that can only be solved with the application of the series as a primary category).

Scott’s ideas about multiple provenance echo today in the essays of the postmodern archival science approach represented by Terry Cook (1993), Peter Horsman (1994),

Sue McKemmish (1998), David Bearman (1999), Eric Ketelaar (2000), Laura Millar (2002), Tom Nesmith (2015) and others. From these authors' perspectives, the principle of provenance assumes multiple characteristics in a context of contemporary production, where a record is created with the purpose of supporting various activities through various moments of creation in different entities. In this context, the rediscovery of provenance, as mentioned before, was recognized and widely accepted by the postmodern approach of the Canadians by the end of the 1980s.

In archival classification procedures, the fonds are the broadest and most essential units of record, followed by the record series and the item. However, over the last decades, the application of the record group concept on archival description faced a practical problem that derives, as pointed out by Cook (1993, 24) "from viewing it exclusively as a physical entity rather than as a conceptual principle." This view reflects, according to Bearman (1999), a record-oriented, descriptive cataloging tradition, as opposed to the context-oriented life-cycle data-management approach. In other words, the archivist must focus on the context rather than on the record to understand provenance as a virtual and dynamic principle that rules practical activities and the establishment of physical series. According to McKemmish (1998, 192) "the physical reconstruction of the *fonds* in a record group, while providing one view of what is a multiple reality, obscures or obliterates other views." As she understands the record as a "continuum," to establish a one and only creator or fond would be too limiting.

Although Duchein cannot be properly considered an author of the postmodern approach, in the article "Theoretical Principles and Practical Problems of Respect des Fonds in Archival Science," he writes (1983, 64) that the principle of provenance "like many principles ... is easier to state than to define and easier to define than to put into practice." According to him, five problems can occur when applying the principle:

- 1) Definition of fonds according to the hierarchy of creating agencies—when most of agencies have a complex functional organization making it difficult to establish a single division;
- 2) Changes of jurisdiction of archival creating agencies—when some jurisdictions are abolished, created, or mixed;
- 3) Extent of the meaning of the provenance of a fond—which agencies should the archivist consider to name the fond, the creator or the receiver? How does one decide?
- 4) The tension between open and closed fonds—in modern administration, it is difficult to say when exactly an

agency stops existing or when its name is changed or placed in the administrative system;

- 5) The relation between *respect des fonds* and respect for the original internal arrangement—should the archivist consider them as two different principles or two levels of the same principle? Should the archivist trust that the agencies are in fact respecting their principle of original order in arranging their archives?

The author gives us answers and possible solutions to all of these questions, but we do not intend to discuss them in this article.

Duchein's article is also one of the important works responsible for the revival of the principle of provenance in the English-speaking archival community. The concept of archival fonds and the principle of *respect des fonds*, as discussed by Duchein and other Canadian archival scholars, became the theoretical foundation for the Canadian standardization of archival description (Nesmith 2015, 290):

The Canadian rediscovery of respect des fonds explores the boundary and structures of archival fonds. It reiterates the importance of fonds as the highest level of archival control and develops strategies to deal with the complexity of modern bureaucracies. The structure view of the archival fonds, that is, the study of its external and internal dimension, highlights the organic connection of authorities, functions, activities and records—the essence of the principle of respect des fonds. This journey of rediscovery also examined a notion of archival fonds capable of expressing a multiple, dynamic, interconnection between records and their creators.

Douglas (2017), in the article "Origins and Beyond: The Ongoing Evolution of Archival Ideas about Provenance," presents three distinctive ways the principle is understood in archival theory: as an organizing principle, as a physical and conceptual construct, and as a social-historical context. The first one is linked to the original meaning of the principle, in the nineteenth century, as a basis to the arrangement and description; the second refers to the rediscovery of the principle by Canadian and Australian archivists, who observed a multiple and multi-faceted provenance due to the changes in records digitally created and; the last one is related to the importance of knowing the record creation history as well as its use and accumulation histories in order to define all of its contexts.

In addition to all of these contexts, it is also important to consider the concept of societal provenance, given by Nesmith (1999, 146), based on the idea that the societal dimensions of record creation and archiving should have a greater impact in archival theory and practice: "The prov-

enance of a given record or body of records consists of the societal and technical processes of the records' inscription, transmission, contextualization, and interpretation, which account for its existence, characteristics, and continuing history." This notion of societal provenance asks archivists to recognize that record creators do not act alone, since they are members of society, making and archiving records in social settings, and for social purposes. In this sense, the societal dimension infuses all layers of provenance (Nesmith 2007).

Looking beyond the narrow view of provenance as a mere principle for arrangement and description, it is important to dive into deeper waters to understand provenance in a more complex context that tells us much more about the social environment where records are created and used than about its creator only. Laura Millar (2002) proposes what she has called the death of the fonds and the resurrection of provenance, advocating for the abolition of the concept of fonds, and defending what she called "*respect de provenance*," a concept encompassing three related components of provenance (12): a) creator history—a history that should focus on the creator, not the records. "The emphasis needs to be placed on who, not what. This is our existing archival provenance, enhanced to accommodate organizational and functional changes over time. The creator would not be defined in a single word or phrase;" b) records history—story of the physical management and movement of the records over time. The record-keeping history would be the story of the archives themselves; and (13), c) custodian history—the explanation of the transfer of ownership or custody of the records from the creator or custodian to the archival institution. Postmodern archival ideas about multiple provenance support some other interesting notions about the concept in archival literature. According to Douglas (2017), earlier studies by Nordland (2004), Wurl (2005) and Bastian (2006) examined various impacts of collectivities in records creations and provenance.

Although we aimed in this paper to discuss provenance as a concept fundamental to archival organization, it is important to mention its use also as a means of retrieval of information from archival records. Bearman and Lytle (1985) view provenance as a method of archival retrieval, based on principles of archives administration and reference practices of archivists. Lytle states (1980, 64):

Subject retrieval in the Provenance Method proceeds by linking subject queries with provenance information contained in administrative histories or biographies, thereby producing leads to files which are searched by using their internal structures. Information in the pure or theoretically defined P Method derives only from what is known about the file- the

activities of the creating person or organization and the structure or organizing principles of the file itself.

Another conception of provenance is proposed by Guimarães and Tognoli (2015, 567) who consider provenance as a specific approach of domain analysis in the archival context: "the base for this discussion is the concept of fond, which is based on a network of structures, functions and activities of the producing entity that will give origin to the archival bond of the documentary set." This idea arises from the fact that provenance studies have specific procedures that comprise: a) the study of the structure of the entity or person that created the record; and, b) the study of the functions of the mentioned entity or person. "After those two complimentary procedures it is therefore possible to determine the fonds or records groups, as well as the arrangement and classification schemes."

4.0 Different perspectives of provenance in different domains

Recently, the concept of provenance has been addressed by many other disciplines (such as law, library and information science (LIS), computer science and visual analytics) and applied to different domains (cloud-based storage, preservation of digital records, digital evidence, digital humanities, e-science, open data, linked data, knowledge organization and indexing), as we can observe in Lemieux's and the imProvenance Group publication on provenance entitled "Building Trust in Information: Perspectives on the Frontiers of Provenance," issued in 2016, aiming to respond to what prompted rising interest in the concept of provenance by all of these domains. The group organized a workshop to discuss the different concepts and uses of provenance from a multidisciplinary and interdisciplinary perspective (4):

The aim was to create cross-disciplinary bridges of understanding with a view to arriving at a deeper and clearer perspective on the different facets of provenance and how traditional definitions and applications may be enriched and expanded via a multidisciplinary and interdisciplinary synthesis.

Although disciplines may have different points of view regarding the concept of provenance, a common use of the concept is identified among them; the role of provenance in the determination of trust, especially due to the increasing use of information and communication technology.

In the law domain, the provenance of how a piece of evidence came into the hands of investigative authorities is a fundamental aspect of the admissibility and weight to be given to such evidence (Cohen 2010). This approach

helps to understand the concepts of reliable and non-reliable evidence in law procedures. According to Lemieux and the imProvenance Group (6), “provenance as a means of assigning attribution is not only important in terms of establishing the degree of trust that can be placed in information, but also in terms of assigning rights, such as intellectual properties [sic] rights.”

In the computer science domain, provenance has been defined in the W3 Provenance dictionary (World Wide Web Consortium 2013) as the “pedigree” of data, which means the description of people, institutions, entities and activities involved in producing, influencing or delivering a piece of data or a thing. The concept of provenance to the domain of visual analytics, as the science of analytical reasoning facilitated by interactive visual interfaces, is explored by Jankun-Kelly (2011) on a basis similar to that of the computer science domain. The concept comprises the information about entities, activities and people involved in producing a piece of data or a thing, which can be used to form an assessment about its quality, reliability, or trustworthiness.

In LIS, provenance enables understanding of semantic changes to classification and order of things over time, as pointed out by Tennis (2016). In information retrieval, the provenance is also one of the key access points in libraries. Tennis (94) also argues that in the field of KO the concepts of ontogeny (a term borrowed from biology to discuss the history of a concept through revisions on indexing languages) and provenance are related:

If provenance is defined as the chronology of custody and context (in the physical world often signaled by physical location) of some material, then we can see how revisions of indexing languages could change the context of a concept. With the change in context, the concept may change its meaning, and it is the meaning of the concept, in relation to other concepts and the documents they index that we care about in knowledge organization.

The principle of provenance is essential to guarantee the trust of a record, once it consists of the record’s identity from the establishment of its creator. In this sense, it is not the case that documentary organization in LIS only occurs on a piece-by-piece basis, because the concepts of book series and journal collections (as a starting point for the organization of collections) are provenance-oriented. In those cases, the editor, the title of the series and the journal title can be considered as fonds that ought to be respected by recognizing—and representing—the original organic relationship between the records that compose them: the books or the journal issues. In other words, the concept of provenance could be enlarged in knowledge organization

systems (KOSs) to every access point which presupposes (and depends on) historicity and context.

Such an approach gives a new dimension to the author as an access point, for instance, since it is not only an intellectual (and sometimes material) creator of the document but also (and mainly) a representative of an institutional context and of a subject domain in which the author has a certain position and integrates an epistemic community (Meyer; Molineux-Hodgson 2010). An example is the author Birger Hjørland, in the subject “domain analysis” whose provenance embodies his institution, his academic background, his theoretical views, his research themes, and the organic relationships (theoretical dialogues, convergences, divergences) that he establishes with other researchers that integrate the epistemic community of domain analysis (e.g., Tennis, Smiraglia, Mai, etc.) as well as the organic bond of his papers on domain analysis.

The preservation of digital objects is a challenge that requires solid information on provenance to ensure their reliability and authenticity as records (Lemieux 2015). Metadata models and standards must address provenance as a guiding concept of respect and be pursued to ensure the long-term preservation and accessibility of digital objects. According to Dappert and Enders (2010), provenance information is one of the administrative metadata, along with descriptive metadata, structural metadata, and technical metadata, usually referred to as preservation metadata. They define (6) provenance as “information about who has cared for the digital object and what preservation actions have been performed on it, as well as rights and permission information that specifies.” Michetti (2016) pointed out the importance of provenance to digital preservation and trust, defining the concept as a crucial factor of evaluation when assessing the credibility of records on the internet. In the digital domain, as in the traditional one, keeping objects alongside the context that provides meaning to them is the main goal of archival studies. “In addition, provenance of digital objects is itself a digital object that also requires preservation. Both provenance and provenance of provenance are fundamental aspects in any preservation model, theory and practice” (61).

In 2004, The Dublin Core Metadata Initiative (DCMI), through its Collection Description Working Group (DCCDWG), defined provenance as a term for metadata schemes as “a statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.” Still, according to the working group, “the statement may include a description of any changes successive custodians made to the resource.” Hence, from the DCMI perspective, provenance in a metadata domain regards custodial history, which, in the *International Standard for Archival Description (ISAD (G))* is a related term to the elements “archival history” and “immediate source of description,”

which in turn provide information on the history of described units (fonds, groups, items) and about the source of acquisition. The DCMII perspective of provenance on custodial history can also be seen as reminiscent of the postmodern interpretations of archival provenance, especially regarding Millar's proposal of the three related components of provenance: creator history, records history, and custodial history. However, it is important to note that DCMII only addresses the latter two components, while the creator view on provenance is an exclusive concern of the archival perspective.

Provenance is a concept also discussed in the museum domain. Since the 1970s, after the United Nations Educational, Scientific and Cultural Organization (UNESCO) Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, researchers and museum professionals have focused on the importance of studying provenance within the museum context. In the museum domain, provenance is the history and circumstances of an object's ownership or possession, and to determine it, some key elements must be identified such as the place and time of an object's origin through a maker, patron or culture or from the time of an object's discovery at a findspot, through any subsequent points of rest and any transactions, reaching the place and time of that same object's current or last known owner or possessor (White 2017). The rare books domain also shares the museum viewpoint about provenance as the history of an object's ownership or possession. According to Reed (2017, 3766) "a rare book's provenance is the history of its ownership, often elaborating its associations with libraries, private owners, or collections." Still, "in its broad contemporary sense, provenance concerns the original owners and all subsequent persons and institutions that acquired books by purchase, gift, inheritance, or legacy up to the present."

It is also relevant to point out the growing importance of the concept of provenance in the domain of linked open data (LOD), once it "allows supporting trust mechanisms, access control, and privacy policies, digital rights management, quality management and assessment" (Sakr et al. 2018, 181). In the context of LOD, practitioners are resorting to N-quads (subject, predicate, object, source institution) to include provenance in assertions about objects in the semantic web to guarantee reliability and accountability of data sources. About provenance on the web, Moreau's (2010) work illustrates the importance of provenance to the reproducibility of scientific results, to tracking attribution and credit in curated databases and to making trust judgments about the information used over the semantic web. Moreau states (53):

The Web has become a global information space where the contents of databases are increasingly exposed directly. The Semantic Web facilitates the annotations of these data sets with RDF metadata, forming a global web of Linked Data. Technologies such as mashups, tweets and RSS feeds integrate data from multiple sources, providing users with information customized for their needs. In this context, tracking provenance is perceived as a critical issue since it helps determine the quality of and trust one can put into data.

All the different approaches brought together here are far from exhausting the discussions about the principle of provenance and its possible application to the archival field. Yet, they also are far from exhausting the discussions in the knowledge organization field. If we consider the constant changes in the creation, organization and use of records, we can say that provenance is an ever-evolving concept.

5.0 Some recommendations and conclusions

When the concept of provenance emerged in the archival science context in the nineteenth century it had a pragmatic goal: to organize the set of records whose organic relationships with its creators had been destroyed due to a thematic classification. Such an approach led the archivists to apply it as a physical organizing principle, concerned about regrouping the records belonging to the same *fond*. The view of provenance as a static concept within the principle of *respect des fonds* lasted until the end of the twentieth century. When information and communication technologies changed the way records were created and used, the fundamental principle of archival knowledge organization had to be reviewed. In such a context, archivists from many parts of the world responded to the call of rethinking provenance as an intellectual construct. The rediscovery of provenance was the answer given by Australian and Canadian archivists, enabling the creation of reliable records in dynamic and virtual record creation systems in the digital environment, with no rigid structures and multiple provenances. Such records, as a consequence, started to be considered as evidence of activities (which is the whole purpose of an archival document or record).

When the use of provenance reaches new domains (e.g., knowledge organization, computer science, law, library and information science, etc.), provenance is no longer just an organizing principle but also a means of reaching for authenticity and reliability of data and objects in digital environments or museums or to reestablish the original organic relationship in library collections. In this sense, new approaches of provenance can and should arise from a multi-

disciplinary and interdisciplinary dialogue between the interactions of different domains. Especially concerning KO, we believe that the presence and impact of provenance in KO procedures, tools, and products present a very prominent scenario that needs to be better explored.

Note

1. "From 2010, the International Council on Archives (ICA) embraced the challenge of creating an archival terminology database, resulting in the identification of 320 terms in English. This initiative provided elements so that a group of scholars and practitioners worked with the definitions of these terms as well as their terminology correspondent and definitions in fifteen other languages, resulting in the 2013 edition of the *Multilingual Archival Terminology*, available with an interactive function, in which the base offers the opportunity to add languages, terms and definitions" (Tognoli et al. 2016, 118).

References

- Archives New Zealand. 2003. "Glossary of Archives and Recordkeeping Terms." <https://archives.govt.nz/resources-and-guides/>
- Bastian, Jeannette Allis. 2006. "Reading Colonial Records Through an Archival Lens: The Provenance of Place, Space and Creation." *Archival Science* 6: 267-84.
- Bearman, David and Richard Lytle. 1985. "The Power of the Principle of Provenance" *Archivaria* 21: 14-27.
- Bearman, David. 1999. "Documenting Documentation." *Archivaria* 34: 33-49.
- Brenneke, Adolf. 1968. *Archivistica: Contributo alla teoria ed alla storia archivistica Europea*, trans. Renato Perrella. Milano: Giuffrè. Translation of *Archivkunde*.
- Canadian Committee on Archival Description. 2008. *Rules for Archival Description*. 2nd ed. Ottawa: Bureau of Canadian Archivists.
- Carucci, Paola. 1990. *Le Fonti archivistiche: Ordinamento e conservazione*. Roma: Nuova Italia Scientifica.
- Cohen, Fred. 2010. *Digital Forensics Evidence Examination*. 2nd ed. Livermore, CA: Fred Cohen & Associates.
- Cook, Terry. 1993. "The Concept of the Archival Fonds in the Post-Custodial Era: Theory, Problems and Solutions." *Archivaria* 35: 24-37.
- Cunningham, Adrian. 2016. "Describing Archives in Context: Peter J Scott and the Australian 'Series' System." In *Building Trust in Information: Perspectives on the Frontiers of Provenance*, ed. Victoria L. Lemieux. Switzerland: Springer International, 49-57.
- Dappert, Angela and Markus Enders. 2010. "Digital Preservation Metadata Standards." *Information Standards Quarterly* 22, no. 2: 4-13.
- Douglas, Jennifer. 2017. "Origins and Beyond: The Ongoing Evolution of Archival Ideas About Provenance." In *Currents of Archival Thinking*, ed. Terry Eastwood and Heather MacNeil. 2nd ed. Santa Barbara, CA: Libraries Unlimited, 25-52.
- Dublin Core Collection Description Working Group (DCCDWG). 2004. "Dublin Core Collection Description Proposed Term: Provenance." <http://www.uko.ln.ac.uk/metadata/dcmi/collection-provenance/2004-02-10/>
- Duchain, Michel. 1983. "Theoretical Principles and Practical Problems of Respect des Fonds in Archival Science." *Archivaria* 16: 64-82.
- Duff, Wendy and Verne Harris. 2002. "Stories and Names: Archival Description as Narrating Records and Constructing Meanings." *Archival Science* 2: 263-85.
- Evans, Frank B. 1991. "Records and Administrative Processes: Retrospects and Prospects." *Janus* 1: 29-37.
- Guimarães, José Augusto Chaves and Natalia Bolfarini Tognoli. 2015. "Provenance as a Domain Analysis Approach in Archival Knowledge Organization." *Knowledge Organization* 42: 562-9.
- Horsman, Peter. 1994. "Taming the Elephant: An Orthodox Approach to the Principle of Provenance." In *The Principle of Provenance: Report from the First Stockholm Conference on the Archival of Principle of Provenance*, ed. Kerstin Abukhanfusa and Jan Sydbeck. Stockholm: Swedish National Archives, 51-63.
- International Council on Archives (ICA). 2007. *ISAD(G): International Standard for Archival Description*. Paris: International Council on Archives.
- International Council on Archives (ICA). 2015. "Multilingual Archival Terminology." <http://www.ciscra.org/mat/>
- Jankun-Kelly, T.J. 2011. "The Case for Visual Analytics Provenance Cases." In *Workshop on Analytic Provenance: Process + Interaction + Insight*. China.
- Ketelaar, Eric. 2000. "Archivist Research Saving the Profession." *The American Archivist* 63: 322-40.
- Lemieux, Victoria L. and the imProvenance Group. 2016. "Provenance: Past, Present and Future in Interdisciplinary and Multidisciplinary Perspective." In *Building Trust in Information*, ed. Victoria L. Lemieux. Switzerland: Springer International, 3-45.
- Lytle, Richard H. 1980. "Intellectual Access to Archives: I. Provenance and Content Indexing Methods of Subject Retrieval." *American Archivist* 43: 64-7.
- Martín-Pozuelo, Mari Paz Campillo. 1996. *La construcción teórica en archivística: El principio de procedencia*. Madrid: Universidad Carlos III de Madrid.
- McKemmish, Sue. 1998. "Are Records Ever Actual?" In *The Records Continuum*, ed. Sue McKemmish, and Michael Piggot. Clayton, Victoria: Ancora Press.

- Meyer, Morgan and Susan Molyneux-Hodgson. 2010. "Introduction: The Dynamics of Epistemic Communities." *Sociological Research Online* 15: 1-7.
- Michetti, Giovanni. 2016. "Provenance: and Archival Perspective." In *Building Trust in Information*, ed. Victoria L. Lemieux. Switzerland: Springer International Publishing, 59-68.
- Millar, Laura. 2012. "The Death of the Fonds and the Resurrection of Provenance: Archival Context in Space and Time." *Archivaria* 53: 2-15.
- Moreau, Luc. 2010. "The Foundations for Provenance on the Web." *Foundations and Trends in Web Science* 2: 99-241.
- Nesmith, Tom. 1999. "Still Fuzzy, but More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory." *Archivaria* 47: 136-50.
- Nesmith, Tom. 2007. "The Concept of Societal Provenance and Records of Nineteenth-Century Aboriginal-European Relations in Western Canada: Implications for Archival Theory and Practice." *Archival Science* 6: 351-60.
- Nesmith, Tom. 2015. "Principle of Provenance." In *Encyclopedia of Archival Science*, ed. Luciana Duranti and Patricia C. Franks. Lanham, MD: Rowman & Littlefield, 284-8.
- Nordland, Lori Podolovski. 2004. "The Concept of 'Secondary Provenance': Re-interpreting Ac ko mok ki's Map as Evolving Text." *Archivaria* 58: 147-59.
- Pearce-Moses, Richard. 2005. *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists.
- Posner, Ernst. 1940. "Some Aspects of Archival Development since the French Revolution." *The American Archivist* 3: 159-72.
- Reed, Marcia. 2017. "Provenance of Rare Books." In *Encyclopedia of Library and Information Sciences*. 4th ed., ed. John D. McDonald and Michael Levine-Clark. Boca Raton, FL: CRC Press, 3766-73.
- Sakr, Sherif, Marcin Wylot, Raghava Mutharaju, Dan Le Phuoc and Irini Fundulaki. 2018. *Linked Data: Storing, Querying and Reasoning*. Cham: Springer.
- Schellenberg, Theodor Roosevelt. 1965. *The Management of Archives*. New York: Columbia University Press.
- Schellenberg, Theodor Roosevelt. 2003. *Modern Archives: Principles & Techniques*. Archival Classics Series. Chicago: Society of American Archivists.
- Scott, Peter. 1966. "The Record Group Concept: A Case for Abandonment." *American Archivist* 29: 493-504.
- Tennis, Joseph Timothy. 2016. "Conceptual Provenance in Indexing Languages." In *Building Trust in Information*, ed. Victoria L. Lemieux. Switzerland: Springer International, 93-9.
- Tognoli, Natália Bolfarini, José Augusto Chaves Guimarães and Gilberto Gomes Cândido. 2016. "The Terminological Dimension of Provenance Description in the Multilingual Archival Terminology – ICA: Some Translation Problems." *Aidainformazioni* 34: 117-25.
- Trace, Ciaran. 2015. "Record Group." In *Encyclopedia of Archival Science*, ed. Luciana Duranti and Patricia C. Franks. Lanham, MD: Rowman & Littlefield, 312-5.
- White, Layna. 2017. "Provenance of Museum Objects." In *Encyclopedia of Library and Information Sciences*. 4th ed., ed. John D. McDonald and Michael Levine-Clark. Boca Raton, FL: CRC Press, 3756-65.
- World Wide Web Consortium. 2013. "PROV-O: the PROV Ontology." <https://www.w3.org/TR/provdictionary>
- Wurl, Joel. 2005. "Ethnicity as Provenance: in Search of Values and Principles for Documenting the Immigrant Experience." *Archival Issues* 29: 65-76.