Athena Salaba (Kent State University)

Use and users of subject authority data

Abstract
Professionals and end-users use subject authority data in a variety of ways. This paper reports on the findings of two surveys on subject authority data use by information professionals in the semantic Web environment and by information professionals in libraries and information agencies.

1: Background
The availability of subject authority data has been experiencing great changes in the Web environment. Subject authority data are defined here as the information recorded during the authority work and control process: deciding on a preferred term, linking semantically related terms, noting all resources used in the decision making, and including additional relevant information. Authority files, as well as the information included in the authority records, were originally designed for use in the construction and maintenance of subject vocabularies and the metadata creation process for resource description. In the past decade, they became widely available outside dedicated information retrieval systems. More recently, we have seen that subject authority data (e.g., the Library of Congress Subject Headings (LCSH)) may be accessed via URIs, browsed on the Web, and downloaded in a machine-processable format such as RDF/XML, N-Triples, or JSON (Library of Congress, 2009).

Potential users of subject authority data range from information professionals who create metadata and construct controlled vocabularies to those who provide reference and public services to end-users. With the variety of online resources becoming available, end-users now have more access to subject authority data and are able to use them directly in searching. In the new wave of the semantic Web, especially with large subject vocabularies becoming part of the linked data and being exposed and interconnected on the Web via dereferencable URIs, it is essential to report the results of two related studies about the users of subject authority data, the contexts in which subject authority data are used, and the characteristics of different usage scenarios.

Previous studies have examined the use of subject data in formulating search queries and in finding information resources. In a study of professional searchers, Fidel (1991) identified strategies for selecting terms from the indexing language using semantic relationships such as broader, narrower and synonymous terms. Of all the cases of term selection, 75% included indexing language consultation by the searchers. Other studies have examined end-user queries when searching online catalogs and the use of authority data displayed to users as sources for term selection (Sutcliffe et al. 2000). A number of more recent studies have focused on the effects of thesaurus-enhanced information retrieval systems and user interactions with and understanding of indexing language information (including behaviors in term selection). These studies have found that indexing languages benefit users in finding terms for query expansion and in improving search performance (Shiri et al. 2002; Salaba 2009). What is lacking from the existing body of literature are studies focusing on other uses of subject authority data, both in the traditional library environment and in the semantic Web.
2: Research in subject authority data use
The researchers felt strongly that it was necessary to conduct user studies to better understand how different user groups, not just end-users, use subject authority data. Two studies were conducted as described below.

The first study was conducted at the 2006 Semantic Technologies Conference held in San Jose, California, USA and was considered a pilot study. Surveys were made available to the conference attendees asking two main questions. First they were asked to describe their work. Most participants in this study were either creators of semantic tools (including controlled vocabularies, taxonomies, and ontologies) or developers and managers of semantic technology systems. Then, they were asked to describe the ways they use semantic tool data. Lists of categories were provided as possible answers in addition to the open-ended option (“Other, please specify”). Participants could select any number of applicable categories. The invaluable results greatly helped the researchers to extend the research to a larger user population.

For the second study, a survey was sent to information professionals throughout the world during the months of May-September 2006. Participants included authority record creators, vocabulary creators and managers, catalogers, metadata librarians, and reference librarians, among others. As with the first survey, participants were asked to describe their work. In a second question, they were asked to describe their use of subject authority data in different contexts. For each context, cataloging/ metadata creation, subject authority work, and searching or helping others search bibliographic information, a list of specified uses was provided, in addition to the open “Other” option. The results of these studies have enriched our understanding of how subject authority data is used.

2.1: Semantic Web professionals
In the first study, which was considered a pilot study, participants included creators, developers, and managers of semantic Web technologies. A number of them were semantic vocabulary creators. A total of sixty-four participants completed the survey that was distributed during the Semantic Technologies Conference. Participants were asked to describe their work and in what ways they use semantic tools such as controlled vocabularies, taxonomies, and ontologies. They were asked to select all applicable answers.

Of the 64 participants, 53% were creators of semantic tools, 47% offered consultations in information architecture for information retrieval systems, and 42% were supervising metadata creation projects. In addition, 34% were metadata creators and 22% were involved in metadata quality control. Several were offering consultations on vocabulary control (22%) and metadata issues (30%). Finally, 22% were intermediaries assisting others in finding information. Among the 64 participants, 23% also selected other tasks as components of their jobs, for example, software development, research, training others on semantic tools, and offering ontology services.

When they were asked to describe how they use these semantic tools including controlled vocabularies, taxonomies, and ontologies (Fig. 1 lists categories in the order they were presented in the survey) approximately half stated that they used semantic tools in finding information resources (50%) and also in finding metadata information
An equal number of participants (44%) used these tools to find appropriate search terms, or to navigate and modify their search queries. About 39% used semantic tools during metadata creation for providing subject access to resources. Semantic tools were also used to explore topics through their semantic relationships by 38% of the participants. In addition, 25% stated other uses of semantic tools such as for building new tools and evaluating them, for data integration and interoperability, for data extraction and semantic-based data mining, for understanding complex and composite concepts for design of query services, and for developing semantic-based information architectures.

2. In what ways do you use the semantic tools?

(Percentage)

- To find/identify appropriate terms when searching for information: 28%
- To find/identify relevant information resources using specified terms: 32%
- To explore a topic through browsing synonymous and related terms: 24%
- To navigate or modify search queries through broader and narrower terms: 28%
- To access metadata information: 30%
- To assign subject terms and/or class labels in metadata records: 25%
- Other: 16%

Figure 1: Use of semantic tools by semantic Web professionals

2.2: Informational professionals

In the second study, participants included information professionals such as catalogers and other metadata creators, knowledge organization system creators and subject authority data creators, and information intermediaries such as reference librarians. A total of 798 participants from around the world completed the survey, which was available through SurveyMonkey.

Of the 798 participants, the largest group represented professionals working in the creation of metadata and cataloging records (78.8%), followed by those involved with quality control of metadata and cataloging projects (60.2%), supervising these types of projects (45.6%), creating and maintaining subject authority records (43%), and those assisting others in finding information and other reference work (approximately 30%).
When these information professionals were asked how they use subject authority data in cataloging and metadata creation, about 95% of respondents indicated that they used the data to select and verify terms for cataloging and indexing. Other uses included verifying and understanding terms and their relationships (72.3%) and exploring relationships among subject terms (69.1%). Additional uses identified by the participants included reclassification projects, correlating subject terms with classification numbers, and getting an overview of the important concepts in a field. Figure 2, below, lists the data in the order they were presented in the survey.

Among the participants who used subject authority data in subject authority work, the majority used the data to normalize and standardize terms (87.2%). Their usage also included modifying authority data (62.3%), maintaining data for future use (57.8%), and establishing and updating term relationships (52.8%). Additional uses within the subject authority work content include creating local authority records, proposing new records to a national agency, and becoming aware of new terms. Figure 3, below, lists the data in the order they were presented in the survey.

Of participants who used subject authority data in searching bibliographic information for their own needs or when helping others (i.e., end-users) to find bibliographic information, the majority used the data available to select and verify appropriate terms for use in search queries (86.7%). Other uses included finding relevant documents on a specified topic (76.6%), modifying search queries by utilizing
semantic relationships (65.3%), navigating bibliographic information by using semantic relationships (63.8%), and exploring and understanding subject areas and the relationships of domain subject terminology (61.6%). The responses also included finding local terminology, correlating subject terms to classification number and vice versa, and selecting terms for personal bibliographic files. Figure 4, below, lists the data in the order they were presented in the survey.

![Figure 3: Use of subject data in subject authority work](image)

3: Conclusion
Information professionals in the traditional library, information agencies, the more experimental Web, and especially the semantic Web environment use semantic tools, such as controlled vocabularies, taxonomies, and ontologies, in a variety of ways. A better understanding of semantic tool information use and user needs will inform future structures.

A more detailed account of the findings from these two studies, enriched with additional tables and charts will be included in the presentation. In addition, a discussion of the implication of the findings and future directions will help better understanding of subject authority use and further development of subject authority data.
References


Web documents have been accessed 1 September 2009.
Eduardo Ismael Murguia  (University of São Paulo, Ribeirão Preto)

Collecting and knowledge organization
A theoretical approach from the material culture studies

Abstract
This project is structured considering a discussion on the basic concepts upon which a specialized practice is
built in librarianship, archivistic and museology: the collection and the fund. Afterwards, the project is
justified through a discussion about information science leading to the emergence of the concept of document.
Yet, it is believed to find a theoretical place for such concept, reconciliatory or not, through which some
concepts of collection and funds could be reviewed. Finally, we point out our hypothesis of work.

1: The bibliographic collection
That is the first approach to bibliographic hoarding headed to bibliophilia, by
librarianship. Records of attitudes related to attachment to books are found in the
universal literature at several spells, and they are also related to several distinct
situations. To name a few, there is one made in 1752 – when Denis Diderot and Jean
D’Alembert issued the first edition of The Encyclopaedia – about bibliomania, defined
as “n.f. fury to possess and collect books” (“s.f. furor de possuir os livros e de os
colentar” apud Silveira 2004, 21).

Following the path of collectors in the pursuit of notable books for the collections,
Almeida (2001) highlights some significant positions such as: fetish-object, through
worship of readers, most of them bibliophilos, i.e. those who are addicted to books, they
go even further at associating the book with the possibility of joy due to the possibility
of physical contact, through the senses. Still, the possibility of the book being object of
care was developed by Gustave Flaubert in his first tale, issued in 1836, entitled
Bibliomania. A bibliophile is generally a great reader who started to get also interested
by the characteristics of the book itself: “Besides the content, owning a book may also
be appreciated as an object of art, for the ilustation, layout, paper, typography, or
binding” (Mindlin 2004, 16). The bibliophile is one who searches intellectual and
artistic pleasure by possessing a book.

In the last decades, librarianship turned to the studies of development of collections,
following the trends of the studies of planning and management (Vergueiro 1989).
Armed the foundation originated from management, the studies of librarianship about
collection were dedicated to the planning of its development. In this context, the
models, methods and techniques to preview the growth of a certain collection, among
whose, the methods originated in the study of customers and of bibliometry, pursuing in
the librarian’s need of having clear objectives, not to have a disordered, messy, unfixed
asset (Andrade & Vergueiro 1996).

Setting up criteria to learn where and how to allot the constant lean budget was also a
preoccupation in this study, as well as the problems of professional awareness at the
moment of the selection, and of the discard (avoiding personal interference in the choice
of the items). When librarianship incorporated the reflections and administration
studies, especially those of business administration, it favored the corporative libraries.
This way, the rules of accountability, at the lowest possible cost became a must, making
room for turning libraries into cold, impersonal places. That made the interest of
librarianship turned around institutional libraries exclusively, neglecting private libraries.

Due to this management influence, bibliographic collections were studied in a non-historical way, as if they had always existed in this format, without considering their origins and path. That is interesting to learn first show the collections were formed in order to conceive the objectives of a library, museum or even the policies of the formation of their collection. That is exactly the dynamics of those collections which bear their policies and objectives. Thus librarianship got less interested in the formation of collections, without noticing that the reason which leads someone to collect is what will later determine the objectives of the collection.

2: The fund
The archival bibliography matches the fact mentioned above concerning the archives given birth in writing. At a certain moment, tied to the power, administered and controlled (Derrida 1995), afterwards under the influence of 19th century historiography, archivistics and bibliography are related to the construction of the past, and finally, supplying information about multiple activities kept in the documents, they are offered to scientists and public in general (Fuster 1999).

However, archivistics appears as a discipline in the 19th century, due to an administrative order of the French Ministry of the Interior (1841), establishing criteria to ordering the archives, trying to standardize the manner they were treated. That note, according to Cruz Mundet (2008), puts an end to the “mess” caused by the thematic treatment of the archives. This principles were rapidly spread throughout the European countries with a consequent acceptance. What was put into discussion was the need of setting up rules and objective principles to guide the adopted procedures by the professionals who work with archives. As a consequence of that panorama relevant, totally normative and practical issues show up: the manuals. Therefore, archivistics is born through practice, and rules, neglecting theory and reflection. At least during the first decades of its discipline formation.

The first manual entirely devoted to the documentary treatment is the *Manual of arrangements and description of archives*, issued in 1898 by S. Muller, T. Feith and J. Fruin, more commonly known as “Manual of the Dutch archivists”. This manual was translated into more than sixty languages and was the first totally devoted to archivistics. To some authors as Duranti (1995) this manual must be considered as the first scientific treatise of archivistic. The manual consisted of one hundred rules as to how documents must be treated documentarily. It articulates such treatment from the initial idea of fund, which is based on two principles: principle of provenance and principle of original order. That is, the documentary arrangement must follow these two principles strictly.

Another manual which played an important role in the consolidation of archivistics as a discipline was Sir Hilary Jenkinson’s *Manual of archival administration*, issued in 1922, twenty four years after the Manual of the Dutch. Reflecting the positive proposals of the positivist historiography of the 19th century, and its studies on paleography and diplomacy, Jenkinson defended the use of documents as an impartial evidence, natural product of processes and transactions pertaining to administrative actions. For him, the archivist was the guardian of the truth, and should never interfere neither in the
evaluation, nor in elimination of documents, for any interference might compromise the impartiality of the fund. Such was the rule of the administrators. Obviously such documentation was chosen by the administrator, and could only be the one convenient to their interests. Therefore, there would be only an official truth, Jenkins proposes that there must be a difference between the “historical truth” and the “archival truth”. Happily, this controversial statement has been criticized by the Canadian School of Archivists.

In 1960 Tanodi (Cruz Mundet 2008) interprets the fund “as the action field” of archivistics. This is a relevant information because it brings up the role of the collection to the libraries and museums. Cruz Mundet (2008, 42) understands fund as the documental fund considered as “natural product of the activity of an entity, person, physical or moral”. The fund must not be defined by its social use, but by the internal structure of the generating entity. So, the documentary treatment in archivistics respects the primal order, based on the principle of provenance and the consequent principle of organization. This way the principle of provenance is the attitude of respect to the original order, called “natural”, in an attitude which carries on the rancidness of neutrality of the documentation of archives. Also, the principle of organization which assumes that the fund be an organic totality where the parts are related due to the prevailing ties, generated in administration.

3: The museum collection
The development of the discussion about collections is more profitable in the field of museology. The theme is inseparable from the other concepts as material culture, that is the reason why we present the topics considered relevant in conjunction with the previous aspects.

An organized collection by Appadurai (1986) narrows the concept of material culture under the perspective of social sciences. Multidisciplinary and work, does it treat themes such as goods, trade and values which are attributed to the objects. Discussed subjects either inside an advanced industrial society or inside traditional societies, including both Eastern and Western world. This world focuses on revealing the economic fetish of the object while inserted in a circuit of exchange, adding that the economic overlapping passes through anthropology and history. Thus objects must be studied inside a net of meaning and values, far beyond an economic vision, in order to bring about concrete value. However, despite the importance of cultural material for understanding the social process, the concept is still troublesome, having always prioritized the ideal aspect of culture on prejudice of its material aspect (Pearce 1994, 17).

People relate to things and objects in different ways. At the first level: direct form, prevailing the value of utility of the objects. At the second level: when some objects are made to produce or modify other objects, i.e., the case of instruments and tools. A third level is also possible when the links to the objects are established by indirect paths, symbolic mediation, notwithstanding through language or images. At last, there is a fourth level: the case of accumulation, i.e. the act of collecting objects with the simple objective of possession or exhibition. The latter is called hoarding. The phenomenon of hoarding and of collections are bought about as a multidisciplinary problem. The collections, in a broader way, draw attention of areas as sociology, for example.
However, that broadens the scope, heading beyond the collection to think its foundation over: an object. In this context, Baudrillard (1968) was one of the first to think of an object as not being isolated and decontextualized, but inserted in a surrounding environment, to soon be approaching its economic and social aspects. Inside the capitalist society, it can be said that the object of hoarding turns not only into one of exchanging but also of consuming. The author highlights previously overlooked aspects, as removal of objects from social systems, consequently giving place to repositories of objects and of collections.

Under the headings of semiology, the work of Pearce (1992, 66) discusses and interprets this objects in their meaningful contexts. For him, the objects are also significant elements, though different meanings can be attributed to them according to the time of the production and interpretation, and to the places they are stored or exhibited. Some places, such as museums are likely to become environments in which the objects might be reorganized meaningfully, not only for being in sites that have a meaning themselves, but also for being arranged inside the museum (e.g. highlighting an object). The author also discusses the issue of collection formation. They are also cast or elections made by these objects for different purposes, e.g. to be possessed or exhibited, as can be seen later. There you have the cycle closed, since it is in the arrangements of the collections that the meanings of the collections find their vitality.

Not all collections belong to their creator, i.e., collector. Many collections survive, are modified by sales, heritage, etc., and become part of or are constituted themselves as museums. The example for that is the pioneer work of French art historian Bazin (s.d.), when tracing a long path of museums, in the sense we mentioned earlier; he follows the Western history entirely.

Recent works are intended to clarify the birth of museums, as Hooper-Greehill (2003), for those who consider museums and modern collections as to serve the formation of knowledge. In the Renaissance, the scenario of the collector is constituted by a new social and economic arrangement, and a new kind of questioning and scientific knowledge is translated in the picture of the prince. Afterwards, we notice that the changes in the scenario, due to the constant changes, give place to museological collections created by the actual society, with a new scientific, ethnographic spirit, until we reach the multidisciplinary modern museum, in the vision of the author. Another modern concept of museum is offered by Bennet (2005), who analyses different aspects under the vision of a Foucaultian interpretation. Following the presentation on the origin of museums and their exhibitions, and the political rationalism upon which it is sustained, the author emphasizes the narrative aspect of the museum through the objects.

If the thematic of collections hoarding have been objects of interdisciplinary nature of investigation, noticed in sociology, history, art history and semiology, the investigation of how this thematic is applied to the universe of museology, the area in which the study of collections have been related to directly, is justified. In relation to museology, especially to the issue of formation of collections, Pearce (1992) clarifies the relation between the collection and the collector. For the author, the relationship is established at the moment of selection; however, the reasons which lead to a collection are obscure, but vital. Those reasons are part of the characteristics of the human being, because they elucidate the relationship between the man and the object, when hoarding
must always been taken to consideration. The author goes on, stating that in contemporary capitalist society hoarding is considered an entertainment, a hobby or a leisure activity, if not playful. Due to this fact, competition among collectors is quite common. Hoarding is a voluntary act which leads to the construction of a collection, never thought in parts, but inseparable as a whole. The collections always start in a spontaneous way, and in this sense they exist just for the simple collector’s will, although many of them have been built in order to get social prestige. In many countries, collectors have expressive means and meeting places for their activities.

The collections can differ in many ways according to the aim of the collector. Pearce (1992) points out the souvenirs in first place, for it is the most common form of collection, and common to almost everybody. This kind of collection can be constituted by all kinds of objects, which can only archive unity if associated to someone’s or a group’s life. Though, they can reach a museum if they mean anything upon a spell or person, or other information can be subtracted. Second place, differently from the souvenirs, the fetishist collections recall not the objects, but the collector him/herself. The difference between these two modes are very subtle, because both are enrooted in an irrational process, through which an affective relationship between the object and its collector exists. But in the case of souvenirs another universe is involved, and construed from the removal of the object of a network, in which it might have existed in a different manner. In third place, we have the systematic collections, which appear in the Renaissance, and are rationalistically construed, according to the articulated principle of the collected objects, under one or several categories.

4: Information science
Among the changes due to technology, one deserved especial attention: the possibility of broadcast and reproduction of knowledge, the so called “information explosion”, which is nothing but the consequence of the means of information as a facilitators to circulation of information. Still, the concept of information itself appears as a response to the explanations of the phenomenon, which despite having been announced in the 19th century, gets astounding proportions in the last century.

However, not all birth to science or socially productive activity is easy. In fact it is all about the construction of history, built and modified according to social changes. So, at the first moment information was seen as an object associated to transmission technology, through where the messages could be decoded in impulses, which allowed their circulation. From this perspective information should be qualified as proposed by Shanon and Weaver. In a pioneer article, Shera (1968) sets up the ignition of information science, highlighting the detachment of the proposed documentation by Otlet and librarianship. In the same perspective Borko (1968) puts the bases of information science systematizing its object, importance and application. Yet, Saracević (1996) emphasizes the multidisciplinary character of this new science. I mention the three articles among so many others, due to the repercussion as pioneering articles in the pursuit of the object in the area. This is a major issue because it shows the pursuit of a definition sustained by the differentiation of librarianship and documentation, despite the intrinsic relationship among them.

Recently, several scholars, since the placement of information science, began to make an accurate visit to the French tradition documentation. Otlet and Briet occupy,
more clearly, the center of concerns (Rayward 1995). This new conceptual trend is represented by what we call the “theory of the document”. Proposals as those of Frohmann (2004; 2009) involve the pursuit of a reflection prioritizing the epistemological question about the document and the social overlapping of the used objects in a common form, notwithstanding determining the social life of the individual. The proposals of this author also rebounds in the traditionally adverse areas to this kind of discussion as archivist, which also starts to concern about the document in the search of the theoretical and professional redefinition (Schwartz & Cook 2002); or the issue of the social determination that the documents cause, analysed in concrete cases.

A special case among the recent authors is Buckland, who designs a very interesting path throughout the years. Although this is not done in a strict chronological order, we notice that the path starts with the preoccupation upon the recovery of information and a systematic approach of management of the information resources, to soon go over the history of information science, as far as the concern about the document, and the European documents. In a relevant article, Buckland (1991) differentiates three concepts of information: information as process (communication action), information as knowledge (understood as non-action, but as object agent of changes). These two concepts have a common characteristic: they are intangible. However there is a third concept which treats information as a thing, that is tangible. Information as a thing is neither a process, nor knowledge itself, but their physical representation.

Let’s consider Buckland’s proposal as a response to a weary information model. The influence of new technologies in which concerns the broadcast possibilities and perception of reality called virtual led to overvalue of new material realities, represented by the French school of philosophy in the 1990s. Following this thought, Buckland’s reading of Pearce is fundamental to contextualize his answer.

According to our argumentation, we see that until the 1990s information science kept a dialogue closer to librarianship than archivistics; while the latter gets a disciplinary status in the 19th century, although it was born as a bureaucratic disposition. After some time, but still in 19th century, several manuals, practical and normative, come up to guide the documentary job in archives.

Curiously, in 1956, four years after Briet’s article, Schellenberg (2002) publishes his book Modern archives... Despite lacking elements to state the influence of Briet upon Schellenberg at this moment, he singles out the archives of records proposing a specific treatment to this type of documents, very similar to the conception of French documentation. Also getting close to the separation between librarianship (Chicago School) and information science (e.g., special libraries).

The detachment among the traditional archives and the records management proposal by this author will be at charge of reconciliation, by the so called Canadian School at the end of the 1980s and notoriously in the 1990s, all around the issues of the magazine Archivaria. Such issue serves as the scenario to new ideas in archivistics, represented by a profound reflection on information. So, the archival document, which at a certain moment had an administrative value and then historical, in this new age will have informative value, closer to information science. In parallel, Frohmann’s proposals also reached archivistics, according to our previous clarifications.

We went through this path proposed by information science and archivistics with the objective of achieving the proposal of the existence of a current concern about the
document of both areas, focusing on material value. A return to the immateriality of information. This object also works as a substrate to understand two principles upon which, in this case, librarianship and archivistics are supported: collection and fund, respectively. We call collection and fund the principles due to all social activity. Moreover, activities related to knowledge and culture need objects, places and agents. And it is also in the relation among these activities, in this case areas, that they can think of themselves. Therefore, explaining the changes that focused on the preoccupation about the object document, we can do a new reading of information science, archivistics and librarianship, through the comparison of the collection of fund.

References
Bazin G., s.d., El tiempo de los museos, Daimon, Madrid.
Bellardo T., Buckland M., 1998, Historical studies in information science, ASIS, Medford.
Duranti L., 1995, Ciencia archivistica, s.n., Córdoba.
Rayward W.B., Arnan R., Hasta la documentación electronica, Mundanau, Madrid.