

ARCHIVES AND LINKED DATA: ARE OUR TOOLS READY TO “COMPLETE THE PICTURE”?¹ TWO CASE STUDIES

Claire Sibille – de Grimoüard
Service interministériel des Archives de France

Introduction

Archivists continually strive to provide new ways to publicize their collections and to allow patrons to connect documents and information in meaningful ways. One of the most cutting-edge trends in sharing information online is Linked Data in order to add richness to the way users experience archival materials and to present archival resources as a part of the diverse data sources on the Web. This paper will present two initiatives which were initially launched by an agency of the French ministry of Culture and Communication, the Department of French Archives (Service interministériel des archives de France) in order to meet archival needs and which ultimately will create information networks between cultural heritage institutions and beyond.

Section 1 presents a project designed to represent a thesaurus in a machine-understandable way for automating machine-assisted indexing processes and also to make it available on the Web with a straightforward navigation within it. The initial scope has since been extended in order to set up a new management environment for all scientific and technical vocabularies of the French Ministry of Culture and Communication.

In Section 2, another initiative is described. The initial objective was to produce a set of standardized forms of names as well as pattern records that local archival institutions might reuse and complete to describe specific local records creators related to the archival resources they keep. We currently wonder what identification code we should assign to the described entities both for use in business applications used by archival institutions and electronic records systems, and for use with Linked Data technologies in a global environment.

Section 3 outlines the lessons learned from both experiences, and focuses on challenges and issues for publishing Linked Data.

1. Delivering the Thesaurus for indexing local records as LOD

1.1. Background

First released in 1987, the *Thesaurus for describing and indexing local records and archives* is a controlled vocabulary which has been created by the Department of French Archives for the public archive sector in France.² Its use is mandatory for all public archival institutions at territorial level and it serves as well as an aid for users to find appropriate search terms. The *Thesaurus for indexing local records* is only available in French, including a subject heading thesaurus and 3 additional controlled lists of terms for administrative activities, types of documents and historical contexts. 1498 preferred terms and 1011 non-preferred terms contained in the subject thesaurus are grouped into 11 classes representing 11 administrative areas (agriculture, economy, justice, public works, etc.) around which terms are organized in accordance with 4 hierarchical levels. The subject thesaurus is a mono-hierarchical one, including equivalence relationships for synonyms, hierarchical relationships between broader and narrower terms, associative relationships when two terms overlap in meaning, and scope notes providing for instance guidelines on how to use a term.

Till 2008, the thesaurus was not available in machine-readable form, so it was only intended to be read by human agents having to index archival descriptions. Therefore, the first scope of the project was to publish the thesaurus in a machine-understandable format for automating machine-assisted indexing processes and for facilitating its integration into retrieval tools. Another important need was to ease the process of creating and maintaining thesauri by domain experts of the Department of French Archives. Lastly, whatever the choice made concerning the data model, all concepts represented in the thesaurus (terms, relationships between these terms, annotations, etc.) should be expressed.

1.2. Modeling methodology

When the project started in 2008, the ISO standard 25964-1 Thesauri for Information Retrieval was not yet released, and the most recent editions of ISO 2788 and ISO 5964, the international standards for monolingual and multilingual thesauri respectively, dated back from 1986 and 1985.³ ISO 2788 covered basic relationships within a thesaurus but it did not provide any formalism of expression for the implementation of these concepts. In parallel, the World Wide Web Consortium (W3C) had promoted the development of the SKOS data model, designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary (Alistair et al., 2005; Pastor-Sanchez et al., 2009). At the initial stage of the project, the SKOS model was still a work in progress but in 2009, it was announced as a standard by the W3C.

SKOS primarily focuses on describing “concepts” and their relationships. Concepts are described using the RDF language, which allows us to combine the concepts with properties (including terms). RDF is based upon the idea of making statements about resources in the form of subject-predicate-object expressions or triples:

the subject is the resource being described;

the predicate is a type of ownership of the resource;

the object is the value of the property.

Resources of the thesaurus are concepts that are referred to by persistent URIs (Uniform Resource Identifiers) you want to be able to trace. The main properties of the concepts are:

terms or labels, that may be preferred labels (`skos:prefLabel`) or alternative labels (`skos:altLabel`);

relationships between concepts: hierarchical (`skos:broader` and `skos:narrower`), associative (`skos:related`), and annotations (for instance, the `skos:scopeNote` property supplies some information about the intended meaning of a concept);

alignment relationships with other vocabularies or resources (the `skos:closeMatch` property indicates that two concepts are sufficiently similar; the `skos:exactMatch` denotes a higher degree of similarity).

The thesaurus source used a “term-based” approach based on the ISO 2788 standard, wherein terms can be semantically equivalent, each one expressing a single concept, in which case they are distinguished as “preferred terms” or “non-preferred terms”. On the other side, in the SKOS (concept-based) model, each concept is associated with labels, which are modeled as RDF literals, so the expression of relationships between labels is not supported. So, it was necessary to make a mapping from the term-based model with its term-to-term relationships to the concept-based model with its concept-label and concept-to-concept relationships. The key principle underlying the mapping between the *Thesaurus for indexing local records* and SKOS was that, in the term-based approach, a preferred term and its multiple related non-preferred

terms were acting as labels for a single concept. Initially, each single SKOS concept was created with a URI based on the domain name of the Department of French Archives. In 2012, the former URIs were replaced by ARK identifiers as a system of persistent identifiers was set up by the French ministry of Culture. Lastly, the thesaurus has been partially aligned with Dbpedia and RAMEAU, the French subject headings maintained by the National Library of France (RAMEAU itself has been skosified and aligned to other controlled vocabularies; see Summers *et al.*, 2008; Van der Meij *et al.*, 2010). At first, we used ThManager, an open source tool for creating and visualizing SKOS (Lacasta *et al.*, 2007). However, an important limitation of ThManager was that it only implemented a former version of SKOS. Then, the SKOS files were edited in the XML editor oXygen and consistency checks of the thesaurus were performed with scripts.

1.3. Developing a set of tools for editing and publishing SKOS vocabularies

Simultaneously, a platform for making the thesaurus available according to the recommendations of the semantic web was developed for the Department of French Archives by an IT services company. ⁴ The developer chose to use Sesame, an open source Java framework for querying and storing RDF data.

The data access interface follows recipe 6 of the W3C Best Practice Recipes for Publishing RDF Vocabularies, as the URI can be “dereferenced” in different manners according to the context. ⁵ The preferred terms are represented by individual RDF pages for every concept contained in the thesaurus. For human use, these pages will be enhanced by a search facility and a hierarchical tree suitable for browsing. Each page is linked to other pages through narrower, broader and related concepts. The URL-rewrite and content negotiation capabilities of the Apache web server are used to do the mapping of requests, e.g. for <http://data.culture.fr/thesaurus/resource/ark:/67717/T1-524> (concept represented by the preferred label “identité commerciale”). If RDF/XML is requested, a redirect is performed to the RDF Distiller with the address of the current version XHTML+RDF page as a referrer (<http://data.culture.fr/thesaurus/data/ark:/67717/T1-524>). If HTML is requested, a redirect is performed to the current version, which is resolved to a language-specific representation by content negotiation (<http://data.culture.fr/thesaurus/page/ark:/67717/T1-524>). In addition, the application provides a SPARQL endpoint, which enables users (human or other) to query a knowledge base via the SPARQL language. ⁶ The endpoint provides a web form in which you can enter a query and get back the results in a machine-readable form.

The platform was initially made available on the portal of the Department of French Archives. However, in 2012, its scope was extended to the release as Linked Open Data (LOD) of other vocabularies of the French Ministry of Culture and Communication within the framework of the HADOC program (“Harmonization of the production of cultural data”). ⁷ The Department of French Archives is involved in this project which aims at setting up a new management environment for scientific and technical vocabularies of cultural and heritage organizations in order to ensure consistency of cultural data. Between 70 and 100 controlled vocabularies about architectural heritage, fine arts, decorative arts, archeology, ethnology, etc., are concerned. As these vocabularies are managed in various applications, a free software, GINCO, was developed for the management of vocabularies. It implements the principles defined in the ISO standard 25964-1:2011 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval.

GINCO is released under the terms of the CeCiLL v2 license. It is available on the GitHub hosting service. ⁸ Its main features include:

- design and management of multiple lists of authorities, taxonomies, thesauri;

- management of terms: status, language, unique identifier of type URI, different types of notes associated with the terms, assignment of terms to concepts, equivalence relationships between terms;
- management of concepts: status, unique identifier of type URI, different types of notes associated with concepts, hierarchical and associative relationships between concepts;
- management of thesaurus arrays (virtual relays, facets) and their use in hierarchical editions;
- management of groups of concepts (micro-thesaurus, theme, domain);
- alphabetical and hierarchical editions ;
- import and export of thesaurus in SKOS/RDF format ;
- consultation through web services;
- alignments between vocabularies and external resources;
- implementation of a reference terminology, in order to provide users with unified access to all vocabularies.

2. Delivering archival authority records as LOD

2.1. Background

French archival institutions have to prepare authority records describing the creators of the archival materials they keep, using ISAAR (CPF) in order to ensure consistency and enable data exchange. As public organizations in France are very similar in the different administrative divisions, it was decided to address this issue centrally and as part of a collaborative work. Therefore, a collaborative project aiming at developing a national authority file describing local public organizations since 1800 was set up by the Association of French Archivists in partnership with the Department of French Archives. The project's scope consists of producing lists of standardized authorized forms of names and standardized descriptions of records creators for each individual institution (records creator) at a local level. For this purpose, patterns would be proposed centrally to serve as basic documentary materials, easily reusable by different archival institutions to describe their own records creators.⁹

Thus, a list of authorized forms of names was developed and made available on the website of the Association of French Archivists. This list is organized according to a thematic division between thirteen major administrative areas: General Administration; Public Finance and Taxation; Police and civil protection; Agriculture; Economy and Industry; Education and Research; etc. ¹⁰ In addition to listing authorized forms of names, the project participants are producing standardized EAC-CPF authority records using ICA-AtoM, the free, open-source software developed by the Canadian company Artefactual Systems in collaboration with the ICA Program Commission (PCOM) and a growing network of international partners. ¹¹

Three years after the project's launching, time has come to evaluate how much success participants have had in creating records for local and regional bodies from those high-level pattern records. It seems that the process of data re-use is still in early stages and that, at the moment, archivists re-use only those pattern records as management tools. Although archival authority records by nature constitute connection points between different information resources, many French archivists still consider EAC-CPF as a mere technical choice for encoding contextual information rather than as an opportunity to link archival authority records to other resources retrievable on the Internet. Each authority record may, through its area

relationships, be related to other governmental organizations and agencies at local or central level. For example, relationships can be established between administrative courts at the local level and the Council of State (in French: “Conseil d’Etat”), which is the higher level of administrative jurisdiction, and thus can also be found in the information system of the National Archives. However, there is currently no connection between the generic patterns produced for local archives and the authority files of the National Archives. Neither are the generic patterns connected with the authority records created for individual institutions by territorial archival institutions. Therefore, all these projects should be interrelated, coordinated, and federated in order to share, exchange, and interconnect information.

2.2. Modeling methodology

Though the authority records created within the framework of this project do not describe individual institutions but categories of institutions, there is a strong rationale to expose them as such in the web of data. If actionable links by means of persistent identifiers are established between them and each of the authority records for individual institution of the same category, described locally in a distributed way in the respective archival systems, these reference records are likely to become federating points for all the authority records of individual institutions. Therefore, the Department of French Archives launched a research project to convert this set of authority records to the representation model of the Semantic Web. This project perfectly fits with the objectives identified by the roadmap “Cultural metadata and transition to a Web 3.0” developed in 2013 by the Ministry of Culture and Communication. ¹² This roadmap lays the groundwork for the use of Semantic Web technologies by the producers of cultural data: setting up an infrastructure for permanent identifiers of cultural resources, creating and maintaining a semantic interconnection of cultural resources, etc.

The first deliverable of the project led by the Department of French Archives is an overview of initiatives underway as regards modeling archival authority data. Two models in particular have been evaluated and compared: the ontology developed by the IBC Institute for Cultural, Natural, and Artistic Heritage in Emilia Romagna (Istituto per i beni artistici culturali e naturali della Regione Emilia-Romagna or IBC), which derives from the EAC-CPF original XML schema (Mazzini and Ricci, 2011); and the model developed for actors/creators of cultural resources developed by the French Ministry of Culture and Communication within the framework of the above mentioned HADOC program.¹³

The first result of the IBC initiative was an ontology describing all the elements and the attributes of the EAC-CPF schema and referring to tag library and XML diagram for technical specification. However, this first ontology could not be used to open the authority records codified with EAC-CPF to the world of LOD. So, the developers changed the point of view of the model and moved from the description of the XSD schema in RDF to the definition of a new model based on the schema. As far as possible, they tried to make use of other existing most popular RDF vocabularies: Dublin Core, OWL, FOAF (Friend of a friend), Geonames, Biographical Vocabulary, and VIAF (Virtual International Authority File). However, though the model is intended to apply to every type of entity (corporate body, person, family), it focuses on the description of persons. In addition, as other RDF vocabularies were used, the model includes irrelevant classes such as MediaType, Size or Duration, RightsStatement, etc.

Though its approach is more generic, the HADOC model developed for actors/creators of cultural data seems to better meet our needs. This model is a component of a more global model which defines semantic data required to represent the “identity card” of a cultural property. This “identity card” models all the information required (classes, properties) for identifying a cultural object. The cultural property (class “Cultural Property”) is characterized by one or many identifiers (class Identifier), one or many names (Appellation class), but also by its

custodial institution (class “Organization”), by one or many events to which it is related (class “Event”) and by one or many actors/creators (class “Actor”). The component “Actor” is defined on the same concepts of the global model “Cultural property” to which it is semantically linked. The specific model “Actor” is a kind of specialization of the general model “Cultural property”. The Actor is defined by identifiers (class “Identifier Component”), by one or many activities (class “Activity”), by related events (class “Event”), and by roles (class “roleActor”). The Actor is characterized by the network of his relationships with other actors or other historical resources (association class relationshipActor).

Thus, although the HADOC model for Actors is not based on EAC-CPF, the solution that has been adopted is to use and enrich it with some relevant EAC-CPF elements and to take advantage from the ontology developed by the IBC Institute to develop our own model. This solution could have the benefit of reusing an internally developed model without expensive modifications. Another advantage is to facilitate a smooth integration of the specific project led by the department of French Archives to a more overall project led by the French Ministry of Culture and Communication. Lastly, it allows us to link up the project with the strategic roadmap on cultural metadata and the transition to Web 3.0.

The EAC-CPF abstract model proposes 90 elements and around 30 attributes for contextual description of individuals, families and corporate bodies. But only 60 elements are used to produce authority records in our specific project. Therefore, we are identifying the more relevant elements to be expressed in the target model being developed.

2.3. Identifiers for authority records and described entities

One of the ISAAR mandatory elements is “Authority record identifier” to trace the provenance of an authority record and to identify it uniquely within the context in which it will be used. In addition, a fundamental requirement for using authority records as reference resources on the web, is the assignment of persistent and unique identifiers. However, ICA-AtoM does not control identifiers. It is said in the official documentation: “Record a unique description identifier in accordance with local and/or national conventions”. So, for the moment only “local” identifiers are assigned to generic authority records. Archival institutions, which will export and re-use generic authority records from ICA-AtoM, in order to create specific authority records for each individual institution, will also need to implement web persistent identifiers. For this purpose the Department of French Archives recommends the use of ARK identifiers. Assigning ARK identifiers to these specific authority records would allow to federate them and to link them with other contextual data.

In addition to the issue of implementing web identifiers for records as web resources, archival institutions have become aware of the need for another level of identification: that of uniquely identifying the described entities themselves. Archival institutions wonder what identification code is the most appropriate to adopt for the described entities. In France, the National Institute of Statistics and Economic Studies (INSEE) issues codes that are compliant with ISO 6523:1984 – Information technology – Structure for the identification of organizations and organization parts. But as these codes are managed only at the French level, they cannot be interoperable at an international level. Issues raised with the identification system proposed by INSEE led the department of French Archives to consider the Register for the *International Standard Name Identifier* or ISNI.¹⁴ ISNI identifiers are international, unique, web-compliant (resolvable as http URIs), centrally managed codes, and the central database is already publicly available on the web. Moreover, France is strongly involved in the ISNI governance system as well as in the quality control of the ISNI database, and the National Library of France has just become an ISNI registration agency on its own which plays a leading role on ISNI in France.

Nevertheless, questions still remain about the principles we should adopt for the assignment of this identifier, for example, the issue of the granularity level of an institution's subdivisions to which identifiers should be assigned. Another pending issue is that of relationships between these subdivisions and how can ISNI address these relationships.

3. Lessons to be learned from both projects

3.1. A better understanding of the challenges of Semantic Web

Both projects highlight the advantages for archival institutions to be involved in the LOD movement. Our gradual acquisition of the different LOD technological components, and standards allowed us to better understand the issues of Semantic Web.

Sharing structured and trust data is part of the missions of archival institutions, which already use protocols such as OAI-PMH to improve data interoperability and interchange. The Web of data represents a new opportunity to disseminate and encourage the reuse of archival descriptive metadata. It promotes:

- data transparency and referencing on the Web, by providing easier data access to search engines and users;
- data interoperability, by breaking down silos of data, whose nature, source and structure may be very different;
- data reliability, by tracing the source of resources through URIs; so archival data could be linked to reliable data sets and archival institutions could be positioned as trusted third parties;
- flexible reuse by third parties, by making these re-users able to recover and reprocess the relevant data sets to crosscheck them with external or local data sets.

In addition, Linked Data enables archivists to use and incorporate information from other Linked Data providers into their local description, making their descriptions more complete, and more value-added.

Moreover, after librarians (with the FRBR and FRAD models) and museums curators (with the CIDOC-CRM model), archivists are aware of the pressing need for a comprehensive model.¹⁵ Indeed, for various historical reasons, it seems a bit backward to have developed standards and exchange formats, and then archival software for producing archival descriptions, with no comprehensive overview of entities described in archival descriptions and without explaining the relationships between these entities.

Thus, traditional archival descriptions should evolve to adapt to the evolving technological environment and be part of the Web. This involves enhancing archival data that are much used, not only by the usual users but by other communities, and adapting to changing research practices and Web 3.0 use.

3.2. Promoting our expertise in LOD technologies

At first, there was no direct connection between the two projects. Then, they have converged to larger-scale projects funded by the French Ministry of Culture and Communication, involving other cultural heritage domains. As the Department of French Archives developed its own expertise, it could easily be involved in these cross-domain projects. For instance, we participate in the development of the roadmap "Cultural metadata and transition to a Web 3.0" driven by the General Secretary's departments of the Ministry of Culture and Communication. We currently participate or lead 4 of the 9 working groups charged to define policies as regards implementing persistent identifiers for cultural resources, experimenting inter-institutional interconnection of cultural data, positioning the Ministry of Culture and Communication as a leader in traceability of data, etc.

In addition, the term "Linked Open Data" itself implies linking with external parties and therefore good relations with other key data owners should be established. So, the cross-organizational working groups set up in the framework of the HADOC program were very useful for discussing and harmonizing the use of semantic resources, the creation of LOD, and so on. These organizational links also ensure that we will not be alone in the process with the risk of no one being able or willing to reuse and link to our data.

Lastly, we will be able to reuse in our own projects the expertise we gained. For instance, last year, we launched a project of a national archival portal which is intended to serve as an aggregator for the Archives Europe Portal (APE). In such a project, Semantic Web technologies should be taken into account, with objectives such as:

- multidimensional search by means of rich domain ontology and multiple aggregations views of the same data would be possible;
- data providers could publish data in reusable form that could be incorporated into multiple portals but updates would remain under their control;
- information structure could be directly machine accessible to facilitate cross-portal integration.

3.3. Applying Link Data methods to archival data

The approach taken in both projects was a very pragmatic one. For instance, in the first case, the specification of underlying concepts and semantic relationships was originally intended to meet a specific need (provide archival institutions with a reusable and structured form of the *Thesaurus for indexing local records*), but it also helped us to consider other re-uses of our vocabulary by other communities. We moved from a traditional approach (a vocabulary for indexing archival descriptions) to an ontology-based approach (concepts represented by preferred and non-preferred terms). This pragmatic "step by step" approach enabled us to identify the following requirements for making our data ready for LOD:

- setting objectives, i.e. defining the content to be published as LOD, identifying organizations to collaborate with for linking open data; and having an idea how our LOD will be reused and what external resources we could connect it to;
- analyzing the quality of the data and cleaning data;
- creating an RDF data model by selecting the most useful elements in the source data as a starting point to be converted into RDF, and by identifying the triples in our data;
- mapping the RDF data model to source data;
- selecting ontologies in order to enrich the RDF model with links to ontologies;
- selecting the most appropriate method to convert data to RDF;
- making the data openly available on the web for everyone to use.¹⁶

For our second project about authority records, the main difficulty will be to select the most appropriate namespace and elements for our model. We realized that there was no single RDF model available that could cover all our needs. Therefore, we will have to combine multiple namespaces which cover only part of the needs and at the same time have significant overlaps with each other.

Lastly, both projects also led us to ask specific questions regarding the alignment of our data with other resources. Why link to them? How good are they? What is the quality of the LOD we want to use and its relevance to our data? Furthermore, an important lesson from the alignment of our Thesaurus for local records to Dbpedia and RAMEAU is that domain experts should regularly review vocabulary alignments and that information extraction can be hardly automated. The concepts of our thesaurus were linked to Dbpedia and Rameau concepts with `skos:exactMatch` (“equivalent term”), (“approaching term”) `skos:closeMatch` or `foaf:focus` (“equivalent resource”) relationships. However, Dbpedia often evolves as Wikipedia changes, and changes may be important. Regarding the publication, at first, triples related to the aligned concepts were extracted and integrated in a static way into our own triple-store (the properties: term, description, webpage), enabling us to enrich the description. However, over time, we observed differences between the extracted information and the information in Dbpedia such as: identifiers that no longer exist, or ambiguous or unreliable properties, concepts that may be represented by irrelevant terms. So, in the current version, only the types of alignment (“Concept equivalent” or “Concept approaching” or “Resource associated”) and the URIs are displayed, and the user will have to deduce the identity of the vocabulary from the URIs.

Conclusion

Publishing and using linked data does not represent a change in the definition of archival description, but it does represent an evolution of how archival description is accomplished. For example, linked data is not about generating a document such as an EAD finding aid or an EAC authority record. Instead it is about asserting sets of statements about archival things or entities (corporate bodies, persons or families), and then allowing those statements to be brought together in any number of ways for any number of purposes. An archival finding aid or an authority record is one such purpose.

In addition, both examples show that a lot can be done to inter-relate existing archival data with other data from other cultural heritage domains by using existing tools and by remodeling existing datasets. They also show that creation and maintenance of linked data are difficult because they require skills of different types from several people, such as metadata specialists, archivists, computer programmers., etc. A team is all but necessary: like any new technology component, the use and management of RDF/XML and triple stores as well as writing SPARQL queries require new skills and know-how.

Bibliography

ALISTAIR, Miles, Matthews, Brian, Beckett, Dave, Brickley, Dan, Wilson, Michael, & Rogers, Nikki (2005). SKOS: A language to describe simple knowledge structures for the web. <http://epubs.cclrc.ac.uk/bitstream/685/SKOS-XTech2005.pdf> Accessed 24 July 2014.

ANGJELI, Anila. Archives and Linked Open Data: are our tools ready to “complete” the picture? Some observations from a metadata and standards specialist, a librarian collaborating with archivists to interconnect our data. Annual Congress of the Society of American Archivists, San Diego, 2012. <http://files.archivists.org/conference/sandiego2012/401-Angjeli.pdf>

ASSOCIATION DES ARCHIVISTES FRANÇAIS, Service interministériel des archives de France. Notices d'autorité pour l'administration territoriale de 1800 à nos jours. <http://aaf.ica-atom.org/> Accessed 24 July 2014.

INTERNATIONAL COUNCIL ON ARCHIVES. ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition, 2004. <http://www.ica.org/10206/standards/standards-list.html> Accessed 24 July 2014.

ISO 25964-1:2011 Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: International Organization for Standards, 2011.

ISO 25964-2:2013. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. Geneva: International Organization for Standards, 2013.

ISO 2788:1986 Documentation -- Guidelines for the establishment and development of monolingual thesauri. Geneva: International Organization for Standards, 1986. [replaced by ISO 25964-1:2011]

ISO 5964:1985 Documentation -- Guidelines for the establishment and development of multilingual thesauri. Geneva: International Organization for Standards, 1985. [replaced by ISO 25964-1:2011]

LACASTA, Javier, Lopez-Pellicer, Francisco J., Muro- Medrano, Pedro R., Nogueras-Iso, Javier, & Zarazaga-Sori, F. Javier (2007). ThManager: An Open Source Tool for Creating and Visualizing SKOS. *Information Technology and Libraries*. 26.3. <http://www.ala.org/lita/ital/sites/ala.org.lita.ital/files/content/26/3/lcst.pdf> Accessed 24 July 2014.

Linked Archival Metadata planning project (LiAM). *Linked Archival Metadata: A Guidebook* (version 0.99). <http://sites.tufts.edu/liam/deliverables/prospectus-for-linked-archival-metadata-a-guidebook/> Accessed 24 July 2014

MAZZINI, Sylvia, RICCI, Francesca (2011). EAC-CPF Ontology and Linked Archival Data. Berlin, 1st International Workshop on Semantic Digital Archives (SDA 2011). http://www-e.uni-magdeburg.de/predoiu/sda2011/sda2011_05.pdf Accessed 24 July 2014.

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION (France). Feuille de route stratégique Métadonnées culturelles et transition Web 3.0. http://www.culturecommunication.gouv.fr/var/culture/storage/pub/feuille_de_route__metadonnees_culturelles_et_transition_web_3_0_janvier_2014/files/docs/all.pdf Accessed 24 July 2014.

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION (France). Présentation du programme HADOC (Harmonisation de la production des données culturelles). <http://www.culturecommunication.gouv.fr/Ressources/HADOC> Accessed 24 July 2014.

Ministère de la Culture et de la Communication (France). Les vocabulaires du ministère de la Culture et de la Communication. <http://data.culture.fr/thesaurus/> Accessed 24 July 2014.

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION (France). Wiki de l'outil GINCO. <https://github.com/culturecommunication/ginco/wiki> Accessed 24 July 2014.

PASTOR-SANCHEZ, Juan Antonio, Martínez Mendez, Francisco Javier, & Rodríguez-Muñoz, José Vicente (2009). Advantages of thesauri representation with the simple knowledge organization system (SKOS) compared with other proposed alternatives for the design of a web-based thesauri management system. *Information Research*, 14(4). <http://informationr.net/ir/14-4/paper422.html> Accessed 24 July 2014.

SERVICE INTERMINISTÉRIEL DES ARCHIVES DE FRANCE. Thésaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines. <http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/thesaurus/> Accessed 24 July 2014.

SOCIETY OF AMERICAN ARCHIVISTS, EAC(CPF): Encoded Archival Description (Corporate bodies, Persons, and Families), 1st Edition, March 2010. <http://eac.staatsbibliothek-berlin.de/> Accessed 24 July 2014

SUMMERS, ED, Isaac, Antoine, Redding, Clay, & Krech, Dan (2008). LCSH, skos and linked data. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC 2008)*, 25–33.

VAN DER MEIJ, Lourens, Isaac, Antoine, Zinn, Claus (2010). A web-based repository service for vocabularies and alignments in the cultural heritage domain. *Proceedings of the 7th EUROPEAN SEMANTIC WEB CONFERENCE (ESWC)*, volume 6088, 394-409. <http://www.few.vu.nl/~aisaac/papers/STITCH-Repository-ESWC10.pdf> Accessed 24 July 2014

W3C RECOMMENDATION, SKOS Simple Knowledge Organization System Reference. W3C, August 18, 2009. <http://www.w3.org/TR/skos-reference/> Accessed 24 July 2014.

W3C WORKING GROUP NOTE, Best Practice Recipes for Publishing RDF Vocabularies. W3C, August 28, 2008. <http://www.w3.org/TR/swbp-vocab-pub/> Accessed 24 July 2014.

WILL, Leonard (2012). The ISO 25964 data model for the structure of an information retrieval thesaurus. *Bulletin of the American Society for Information Science and Technology*, 38(4):48–51

Notes

¹ This title is inspired from Anila Angjeli's excellent paper: "Archives and Linked Open Data: are our tools ready to "complete the picture"?", Annual Congress of the Society of American Archivists, San Diego, 2012: <http://files.archivists.org/conference/sandiego2012/401-Angjeli.pdf>

² For more information, see: <http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/thesaurus/>

³ ISO 2788:1986 *Documentation -- Guidelines for the establishment and development of monolingual thesauri* and ISO 5964:1985 *Documentation -- Guidelines for the establishment and development of multilingual thesauri*.

⁴ Available at: <http://data.culture.fr/thesaurus/>

⁵ The W3C Best Practice Recipes for Publishing RDF Vocabularies are available at: <http://www.w3.org/TR/swbp-vocab-pub/>

⁶ The SPARQL endpoint is available at: <http://data.culture.fr/thesaurus/sparql>

⁷ A detailed presentation of the HADOC program, including the conceptual model for cultural properties, is available at: <http://www.culturecommunication.gouv.fr/Ressources/HADOC/Presentation-du-Programme/Presentation-du-programme>

⁸ See: <http://github.com/culturecommunication/ginco>

⁹ For more information about the project launched by the Association of French archivists in partnership with the Department of French Archives, see: <http://www.archivistes.org/Notices-d-autorite-producteurs-1781>

¹⁰ The list of authorized form of names is available at: <http://www.archivistes.org/Notices-d-autorite-producteurs-1781#le-referentiel-national-des-formes>

¹¹ The authority records describing local records creators can be downloaded from: <http://aaf.ica-atom.org/> The ICA-AtoM official website is at: <https://www.ica-atom.org/> The current release of ICA-AtoM is 1.3.1. However, the company Artefactual systems, that developed ICA-AtoM, launched an Enterprise version (AtoM) released in September 2013 (see: <https://www.accesstomemory.org/en/>).

¹² The roadmap on "Cultural Metadata and transition to Web 3.0" can be accessed at:

http://www.culturecommunication.gouv.fr/var/culture/storage/pub/feuille_de_route__metadonnees_culture_lles_et_transition_web_3_0_janvier_2014/files/docs/all.pdf or

http://www.culturecommunication.gouv.fr/var/culture/storage/pub/feuille_de_route__metadonnees_culture_lles_et_transition_web_3_0_janvier_2014/index.htm

¹³ Currently, the HADOC model is only available on a textual form (<http://www.culturecommunication.gouv.fr/Ressources/HADOC/Modele-de-donnees2>). It is being converted into RDF.

¹⁴ Information about ISNI can be found at: <http://www.isni.org/> The ISNI database can be searched at: <http://www.isni.org/search>

¹⁵ See the mandate of the ICA Experts Group on Archival Description for the revision of the international descriptive standards and the development of an archival conceptual model: <http://www.ica.org/13845/egad-activities-and-projects/egad-strategic-work-plan.html>

¹⁶ You will also find a presentation of a graduated approach to applying linked data methods to archival description presented by the Linked Archival Metadata: A Guidebook", released by the Digital Collections and Archives at Tufts University. See: <http://sites.tufts.edu/liam/deliverables/prospectus-for-linked-archival-metadata-a-guidebook/>