Formal Models for Digital Archives: NESTOR and the 5S

Nicola Ferro and Gianmaria Silvello

Department of Information Engineering, University of Padua, Italy
{ferro,silvello}@dei.unipd.it

Abstract. Archives are a valuable part of our cultural heritage but despite their importance, the models and technologies that have been developed over the past two decades in the Digital Library (DL) field have not been specifically tailored to them. This is especially true when it comes to formal and foundational frameworks, as the Streams, Structures, Spaces, Scenarios, Societies (5S) model is.

Therefore, we propose an innovative formal model, called NEsted SeTs for Object hieRarchies (NESTOR), for archives, explicitly built around the concepts of context and hierarchy which play a central role in the archival realm. We then use NESTOR to extend the 5S model offering the possibility of opening up the full wealth of DL methods to archives. We provide account for this by presenting two concrete applications.

1 Introduction

Nowadays, the scope of Digital Libraries (DLs) goes far beyond the realm of traditional libraries and also encompasses other kinds of cultural heritage institutions, such as archives and museums. Nevertheless, these institutions are quite different from several points-of-view: they have different internal organizations and traditions; their resources are different in nature, structure, and descriptions; and their users have different information needs which call for different access methods to resources [18].

Archives are not simply constituted by a series of objects that have been accumulated and filed with the passing of time but they represent the trace of the activities of a physical or juridical person in the course of their business which is preserved because of their continued value over time [2].

To this end, archives keep the *context* in which their records have been created and the network of relationships between them in order to preserve their informative content and provide understandable and useful information over time [8]. The fundamental characteristic of archives resides in their *hierarchical organization*. This expresses the *context* – i.e. the relationships and dependencies between the records of the archive – by using what is called the *archival bond* [4] and it distinguishes archives from other objects in the realm of cultural heritage – e.g. books – which in general are perceived as individual, repeatable and unrelated entities. Archives are in fact made up of series which, in turn, can be organized in sub-series formed of archival units, such as files, registers and so on [17].

T. Aalberg et al. (Eds.): TPDL 2013, LNCS 8092, pp. 192–203, 2013.

In this article we highlight the central role of formal models for the DL, because integration and cooperation between these models can turn into a real case of interoperability between the different facets of DL, including their community, methodology and technology. In this context a model for archives is sorely needed to formally define their characteristics and to prove that general DL methods and technologies can be embodied in this field and respect archival practice.

Therefore, we propose an innovative formal model for archives built around the notion of *archival bond* and *hierarchy*. The proposed model, called NEsted SeTs for Object hieRarchies (NESTOR), is based on the idea of expressing the hierarchical relationships between objects through the inclusion property between sets, in place of the binary relation between nodes exploited by the tree.

The set data models composing NESTOR are well-suited for archival practice; indeed, the idea of "set" shapes the concept of archival division which is a "container" comprising distinct elements that have some properties in common. An archive from the physical point-of-view resembles a Chinese boxes structure as there are boxes, folders, sheets, etc. contained one inside the other. Nested sets are closer to this view of reality than trees are. Indeed, although archival practice commonly considers archives as trees, a tree is actually a higher level abstraction than the nested sets as it only focuses on structural relationships; conversely, NESTOR comprises both the structure and the content of the archive.

DLs benefit from the existence of sophisticated formal models, such as the Streams, Structures, Spaces, Scenarios, Societies (5S) model [7,9], which allow us to formally describe them and to prove their properties and features. Despite the importance of archives, so far there has been no attempt to develop a dedicated formal model, built around their peculiar constituents. Nor can we exploit the 5S model as it is for archives because, as we will discuss later on, it needs some kind of extension and tailoring.

To this end, we exploit NESTOR to formally extend the 5S model to define a *digital archive* as a specific case of digital library. This defines an actual bridge between these two formal models which: (i) allows archives to exist and interact with other realities (i.e. libraries and museums); (ii) provides archives the possibility of exploiting the full wealth of DL technologies and methods; and, (iii) enables integrated access to heterogeneous contents.

As concrete accounts of this, we briefly describe how these formal models can be applied to overcome well-known issues in the archival field. We present two applications, the former regards the interoperability between digital archives and we formally exploit the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[1] to give a concrete account of how DL technologies can be adopted with archives. The latter one shows how the archives modeled with NESTOR can form compound digital objects made available as Linked Open Data (LOD) [10] on the Web adopting the Open Archives Initiative Object Reuse and Exchange (OAI-ORE)[2] as a working framework.

[1] http://www.openarchives.org/pmh/
[2] http://www.openarchives.org/ore/

The paper is organized as follow: Section 2 provides some background on digital archives. Section 3 describes the NESTOR Model and the 5S Model. Section 4 shows how the 5S model can be extended via NESTOR. Section 5 describes how the extended 5S addresses concrete issues in the DL realm. Finally, 6 draws some final remarks.

2 Digital Archives: Background

Archival descriptions have to reflect the peculiarities of the archive, retain all the evidential value of a record, and keep trace of the provenance and original order in which resources have been collected and filed by archival institutions [2,8]. They have to be organized in a hierarchical way to express the relationships and dependency links between the records of the archive in order to retain the archival bond [4]. Therefore, archival descriptions produced according to the International Standard for Archival Description (General) (ISAD(G)) [11] take the form of a tree. The principles of ISAD(G) are put into action by the Encoded Archival Description (EAD) standard [19] for encoding archival descriptions.

EAD reflects the archival structure and holds relationships between entities in an archive. In addition, EAD encourages archivists to use collective and multilevel description, and because of its flexible structure and broad applicability, it has been embraced by many repositories. On the other hand, EAD allows for several degrees of freedom in tagging practice, which may turn out to be problematic in the automatic processing of EAD files, since it is difficult to know in advance how an institution will use the hierarchical elements.

EAD represents an archive as a monolith and every description is embedded in the archival structure. This means that *content and structure are interlinked in the same XML file* and they cannot be handled separately. A direct consequence is that in a distributed environment where it is necessary to exchange data between repositories we are forced to exchange the archive as a whole. Indeed, we cannot share a specific piece of information – e.g. the descriptions of the documents belonging to a specific "series" – without extracting it from the EAD file and losing in this way the structural information retained thanks to the nested tags in the EAD itself [21]. This leads to difficulties in fully exploiting the OAI-PMH within the archives [16].

Furthermore, EAD presents some difficulties both for the expert user (i.e. the archivists who find the "complexity of EAD itself to be a deterrent to implementation" [21]) and the general user who has to consult and interpret the archival data without specific knowledge of archival theory and practice. The main difficulty is related to the reconstruction of the archival context starting from an element buried in the hierarchy; this difficulty related to the data model on which EAD is based may be reflected in a similar difficulty and disorientation for the user in the perception of the context which supply the information needed to satisfy the her/his information requirements.

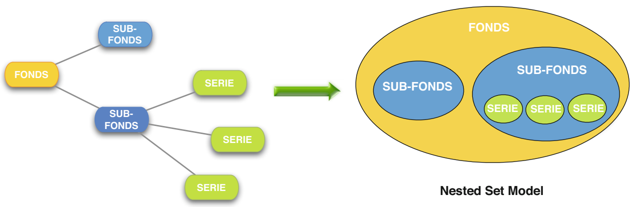

Fig. 1. An archive modeled by means of the NS-M

Lastly, when we need to relate one or more digital objects to their archival descriptions represented as metadata, EAD introduces some more limitations. Indeed, EAD allows for the linkage of only one digital object with an archival description (i.e. through the `<dao>` tag); if we need to link a bunch of digital objects with a description, we need to use the so-called digital wrappers which complicate the structure of EAD and limit its share-ability, interoperability [15] and the possibility of expose archives as LOD on the Web [5].

3 Formal Models for Digital Libraries and Archives

3.1 Overview of the NESTOR Model

NESTOR [6] is defined by two set data models – i.e. the Nested Sets Model (NS-M) and the Inverse Nested Sets Model (INS-M) – which are independent but complementary one to the other. For space reasons, in this context we present only the NS-M and its applications. INS-M provides different modeling features w.r.t. the NS-M, but its use as well as its integration with the 5S model can be derived by following the methodologies we present for the NS-M.

To illustrate the basic ideas behind NESTOR, let us consider an archive composed of six divisions: a fonds, two sub-fonds, and three series. As shown in Figure 1, the NS-M adopts a bottom-up approach: (i) each set corresponds to an archival division; (ii) the innermost sets are the leaves of the hierarchy, e.g. the series; (iii) you create supersets as you climb up the hierarchy, e.g. the sub-fonds and fonds. We can see that the NS-M is well-suited for bottom-up descriptive activities where the archivist does not know in advance the archive s/he is going to describe. In this case, s/he starts to study the documents and decides how to put them together in order to form an archival division, thus the archival hierarchy is built from the bottom. We call it bottom-up approach because in this case dividing the documents into archival divisions is an iterative process: the archivist starts from the whole set of documents (i.e. the fonds) and s/he defines the subsets (i.e. subfonds, series, etc.) by construction, analysing the documents one by one.

Formally, we define the NS-M as a collection of subsets where specific conditions must hold.

Definition 1. *Let A be a set and let $\mathcal{C}$ be a collection of subsets of A. Then $\mathcal{C}$ is a **Nested Sets Collection** (NS-C) if:*

$$A \in \mathcal{C}, \tag{3.1}$$

$$\forall H, K \in \mathcal{C} \mid H \cap K \neq \emptyset \Rightarrow H \subseteq K \vee K \subseteq H. \tag{3.2}$$

Therefore, we define a NS-C as a collection of subsets where two conditions must hold. The first condition (3.1) states that set A which contains all the subsets of the collection must belong to the NS-C itself. The second condition states the intersection of every couple of sets in the NS-C is not the empty-set only if one set is a proper subset of the other one. This definition formally defines how an archive can be modeled by means of the NS-M as shown in Figure 1. The collection of subsets $\mathcal{C}$ is the considered archive; the first condition says that there is a set – i.e. the "fonds" – which contains all the subsets – i.e. "subfonds", "series", etc. – of the archive. The second condition says that two subsets such as two "series" cannot have common elements, thus their intersection is always empty.

3.2 Overview of the 5S Model

The Streams, Structures, Spaces, Scenarios, Societies (5S) [7,9] is a formal model and draws upon the broad DL literature in order to have a comprehensive base of support. It was developed largely bottom up, starting with key definitions and with elucidation of the DL concepts from a minimalist approach. It is built around five main concepts: (i) *streams* are sequences of elements of an arbitrary type, e.g. bits, character, images, and so on; (ii) *structures* specify the way in which parts of a whole are arranged or organized, e.g. hypertexts, taxonomies, and so on; (iii) *spaces* are sets of objects together with operations on those objects that obey certain constraints, e.g. vector spaces, probabilistic spaces, and so on; (iv) *scenarios* are sequences of related transition events, for instance, a story that describes possible ways to use a system to accomplish some functions that user desires; and, (v) *societies* are sets of entities and relationships between them, e.g. humans, hardware and software components, and so on.

Starting from these five main concepts, it provides a definition for a minimal DL which is constituted by: (i) a repository of digital objects; (ii) a set of metadata catalogs containing metadata specifications for those digital objects; (iii) a set of services containing at least services for indexing, searching, and browsing; and, (iv) a society.

While these broad concepts can be also in common with archives, when you look at the specific way in which they are formally defined, you realize that the definitions cannot be straightforwardly applied to the archives case without at least some extension as we discuss in the next section.

4 Extending the 5S Model via NESTOR

The 5S model needs some kind of extension to be tailored to the specific case of archives. The notion of *descriptive metadata specification*[3] (definition 12 [9, p. 292]) is suitable either for representing, for each archival division, a descriptive metadata – e.g. a metadata describing a series, a sub-fonds, or an archival unit – or for representing the archive as a whole, as it happens in the case of EAD.

When it comes to the definition of *metadata catalog* (definition 18 [9, p. 295]), there is no means to impose a structure over the descriptive metadata in the catalog. Therefore, if you use separate *descriptive metadata specifications* for each archival division, as in the former case, this would prevent the possibility of expressing the relationships between these archival divisions, i.e. you would lose the possibility of retaining the archival bond. This means that an archive cannot be properly modeled throughout the 5S model without losing one of its main properties.

Moreover, in a *metadata catalog*, there is no means to associate (sub–)parts of the *descriptive metadata specifications* to the *digital objects* (definition 16 [9, p. 294]) that they describe, but you can only associate a whole descriptive metadata to a whole digital object.

Therefore, if you represent an archive as a whole with a single *descriptive metadata specification*, as in the latter case, it would not be possible to associate (sub-)parts of that descriptive metadata to the different digital objects corresponding to the various archival divisions; this does not allow the definition of compound digital objects and it is a barrier towards the adoption of the LOD paradigm in the archival context as discussed in Section 5.2. Furthermore, this strongly limits the interoperability between digital archives and the possibility of sharing archival metadata with variable granularity.

Our extension to the 5S model is thus organized as follows:

- using the notion of *structure* (definition 2 [9, p. 288]), we introduce the notion of **NESTOR structure**, as a structure that complies with the constraints of NS-M;
- using the notion of *metadata catalog*, we introduce the notion of **NESTOR metadata catalog**, as a metadata catalog that exploits a NESTOR structure to retain the archival bonds;
- using the notion of *digital library* (definition 24 [9, p. 299]), we introduce the notion of **digital archive**, as a digital library where at least one of the *metadata catalogs* is a NESTOR metadata catalog.

Definition 2. *Let* $\mathcal{C}$ *be a Nested Set Collection (NS-C) on a set* A. *A* ***NESTOR structure***(A) *is a structure* $(NS\text{-}G, L, \mathcal{F})$, *where* L *is a set of label values,* $\mathcal{F}$ *is a labeling function, and* $NS\text{-}G = (V, E)$ *is a directed graph where* $\forall v_j \in V, \exists!\ J \in \mathcal{C} \wedge \forall e_{j,k} \in E, \exists!\ J, K \in \mathcal{C} \mid K \subseteq J$.

[3] In this section, we use italics to highlight definitions taken from the 5S model.

Definition 2 applies to the definition of NS-M (i.e. Definition 1) ensuring that the resulting structure complies with the NS-M[4]. Note that the set of label values L and the labeling function $\mathcal{F}$ are not strictly needed for the NS-M, but they can be useful in the context of the 5S and this feature, in turn, may extend the NS-M with semantic possibilities.

The definition of *metadata catalog* in the 5S model can be expressed as follows. Let H be a set of handles to *digital objects* and M a set of *descriptive metadata specifications*, then a *metadata catalog* is a function $DM : H \times 2^M$.

Definition 3. *Let H be a set of handles to digital objects and M a set of descriptive metadata specifications, a metadata catalog DM is a* ***NESTOR metadata catalog*** *if:*

$$\forall h_i \in H \mid \exists M_i \in 2^M \ \wedge \ DM(h_i) = M_i \Rightarrow |M_i| = 1 \tag{4.1}$$

$$\exists \text{ NESTOR structure}(M) \tag{4.2}$$

Condition 4.1 imposes that, if exists, there is only one *descriptive metadata specification* for a given *digital object* because, in archival practice, every single metadata describes a unique archival division, being it a level in the archive or a digital object [11]. Condition 4.2 ensures that the relationships between the different archival divisions are compliant with the *descriptive metadata specifications* in M.

Definition 4. *A* ***digital archive*** $(\mathcal{R}, DM, \text{Serv}, Soc)$ *is a digital library where*

- $\mathcal{R}$ *is a repository;*
- *at least one of the metadata catalogs in the set of metadata catalogs DM is a NESTOR metadata catalog;*
- *Serv is a set of services containing at least services for indexing, searching, and browsing;*
- *Soc is a society.*

Definition 4 extends the definition of *digital library* in the 5S model requiring that at least one of the *metadata catalog*s is a NESTOR one, i.e. there exists at least one *metadata catalog* capable of retaining the archival bonds. This definition has several consequences. Firstly, more than one NESTOR metadata catalogs can be present in the same digital archive, thus making it possible to express different archival descriptions over the same set of *digital objects*. This extends the current practice in which a system for managing an archive is usually capable of managing only one description of the archive, thus giving only one point-of-view on the material held [3,12]. Secondly, you can mix NESTOR and not-NESTOR metadata catalogs which allows for the seamless integration of different visions of the managed *digital objects* within the same digital archive. This opens up the possibility of exploiting the whole breadth of methodologies and tools available in the DL field with the archives.

[4] This definition can be extended to enclose also the INS-M if needed.

5 Applications of the 5S Model Extended via NESTOR

The extension of the 5S model via NESTOR represents an actual bridge between these two formal models which allows the archives to live and cooperate with other methodologies initially not built for archives. A first application shows how OAI-PMH can now be employed by the archives without changing its internal functioning and broadening its functionalities. A second example shows how this theoretical framework can be employed for exposing archives through the LOD paradigm realized by means of OAI-ORE.

5.1 Employ OAI-PMH within Digital Archives

OAI-PMH is the *de-facto* standard for metadata exchange in DLs [1,14] and it has also been modeled by means of the 5S model allowing for "the specification and automatic generation of DL applications" [9]. In the 5S model, Data and Service providers are represented as (electronic) *Societies*; the communications between these providers are *Streams*; the sets, metadata, and schemas are *Structures*; and, each request-response pair (e.g. harvesting the records belonging to specific sets) is associated with a *Scenario* [9, p. 283].

The 5S model formally specifies OAI-PMH by defining its components and services allowing us to know in advance how the harvesting service works in the library context. NESTOR does not provide a formalization of OAI-PMH, thus there is no formal way of defining the harvesting procedure for digital archives. A formal definition of OAI-PMH in the archival context is needed to know if archival properties and constraints are respected when the data are shared via harvesting in distributed environments.

As we have previously seen, OAI-PMH cannot be properly employed with digital archives modeled and realized with EAD files; to this end, we exploit the extended 5S to propose a general solution for modeling the archives, thus overcoming the limitations of EAD and enabling a full exploitation of OAI-PMH. To do this we exploit the set organization of OAI-PMH; indeed, OAI-PMH can organize records into OAI-sets, each one identified by a `setSpec` which is unique handle for a set within the repository. OAI-set organization may be hierarchical expressed thanks to the `setSpec` field using a colon [:] separated list indicating the path from the root of the set hierarchy to the respective node. This feature is used to map an archive modeled by the NS-M into an OAI-PMH set organization respecting all the archival constraints; for a detailed description of the mapping procedure refer to [6].

As an example, let us consider the archive represented by the NS-C in Figure 1. As we can see in Figure 2, each set composing this nested set structure is mapped into an OAI-Set with a proper `setSpec`; the set called "fonds" is mapped into an OAI-set with < setspec > 0001 < /setspec >. This set has two subsets that are mapped into two OAI-sets: < setspec > 0001 : 0002 < /setspec > and < setspec > 0001 : 0003 < /setspec > and so on for the other sets.

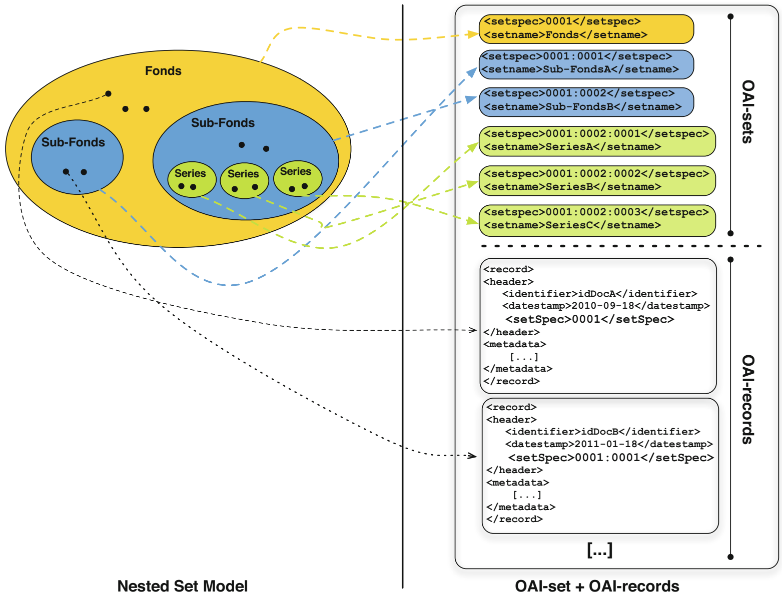

Fig. 2. An archive represented as a NS-C mapped into OAI-PMH

We can see that the hierarchical relationships and thus the inclusion order between the sets is maintained by the OAI-sets and each single archival description is mapped into a metadata belonging to an OAI-set. With this model we can exchange metadata and sets with variable granularity being able to reconstruct their original context. The extended 5S model allows us to formally employ OAI-PMH within the archives where the 5S is used to model the protocol and its features and NESTOR is used to model the archives within the protocol which allows us to respect the archival practice also when there is the necessity of sharing data in a distributed environment.

5.2 Expose Archives as LOD on the Web

Currently, archival practice is moving towards the definition of complex relationships between the resources of interest as well as the constitution of compound digital objects. To this end archives can take advantage of using the LOD paradigm which eases the access to the resources, enhances the interoperability by moving the focus from the systems managing the data to the data themselves, and provides additional and flexible representations of archival resources. In the context of the DL, the LOD paradigm can be instantiated by means of Open Archives Initiative Object Reuse and Exchange (OAI-ORE) which has a precise focus in the representation and management of compound digital objects.

In order to exploit OAI-ORE within the archives there is the need to model the archival structure – which is the mean to retain all the archival characteristics such as the archival bond – into OAI-ORE. The 5S models OAI-ORE as

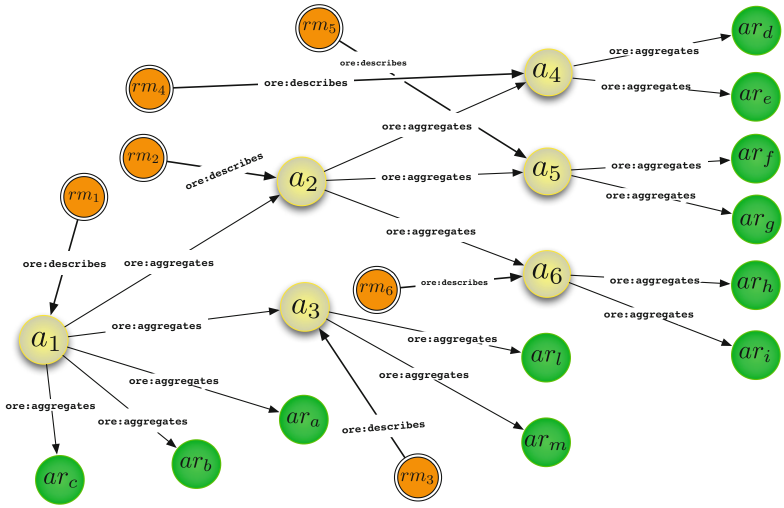

Fig. 3. The archive shown in Figure 1 mapped into an OAI-ORE instance

a *Structure* [13]; this very model extended via NESTOR allows us to impose conditions on the 5S *Structure* – i.e. by defining it as a *NESTOR Structure* (Definition 2) – thus creating OAI-ORE instances accordingly to the archival practice.

To model an archive throughout OAI-ORE we have to consider the features of this framework (refer to [20] for a detailed description) which are: Aggregations (i.e. A) composed of Aggregated Resources (i.e. AR), Resource Maps (i.e. RM) which are materializations of the Aggregations, and Proxies (i.e. P) which allows the definition of relationships between Aggregated Resources. Another important feature is the possibility of defining *Nested Aggregations* which enables the definition of Aggregations of Aggregations. Thanks to this feature, an order (i.e. $\prec_a$) exists between Aggregations: for all $a_i, a_j \in A$ we say that $a_i \prec_a a_j$ if and only if the Aggregation a_i is aggregated by a_j.

We can represent an archive by means of OAI-ORE by modeling it as a NS-C and then defining a mapping between this NS-C and an OAI-ORE instance exploiting the conditions defined by the *NESTOR structure*. So, starting from a the NS-M representation of an archive, the intuitive mapping idea is that every set $H \in \mathcal{C}$ becomes an Aggregation $a_h \in A$ and consequently, every resource $r_t \in R$ belonging to H becomes an aggregated resource $ar_t \in AR$ aggregated by a_h as shown in Figure 3 where the archive of Figure 1 is mapped into an OAI-ORE instance. Furthermore, for every pair of sets $\{H, K\} \in \mathcal{C}$ such that $H \subseteq K$, it is possible to create a pair of aggregations $\{a_h, a_k\} \in A$ such that $a_h \prec_a a_k$ where $\prec_a$ is the order relation defined above[5].

Every set in the NS-C is mapped into an Aggregation in the OAI-ORE model materialized by a Resource Map which in the 5S model is a modeled as a

[5] Further details about the formal mapping can be found in [5].

Structure and in the extended 5S model is modeled as a *NESTOR Structure*. The inclusion order between the sets is maintained by the relation defined between the Nested Aggregations of OAI-ORE. Every resource belonging to a given set H in the NS-C is mapped into Aggregated Resources belonging to the Aggregation mapped from H. Note that OAI-ORE proxies allows us to handle complex relationships between metadata and digital objects in an aggregation. Therefore, we can map a NS-C into a correspondent OAI-ORE instance ruled by the definition of NESTOR Structure in the extended 5S model.

This procedure makes it possible to model and describe the archives from scratch by means of OAI-ORE while allowing archivists to easily express relationships between archival metadata and digital objects. Archival principles are preserved and still have primary importance for understanding archival resources; at the same time, OAI-ORE offers the possibility of defining new relationships between the resources enabling the definition of new services over the archives. Furthermore, this methodology provides a means to define archival compound objects that can be shared with the systems which already employ OAI-ORE and that can be exposed as LOD on the Web.

6 Final Remarks

The 5S model has been built around the most general concepts but without specifically dealing with the peculiar features of the archives. This hampers the possibility of fully exploiting and applying them for defining a theory for digital archives, intended as digital libraries with specific characteristics that fit in the archival domain.

The archival domain deserves a formal theory as well and this theory has to be reconciled with the more general theories for DL in order to disclose to archives the full breadth of methodologies which have been developed in the DL field.

To this end, we have introduced a formal model, called NESTOR and we extended the 5S model through it in order to introduce the notion of digital archive as a specific case of digital library complying with archival constraints. Finally, we applied this extension to two concrete cases showing how we can overcome current issues with state-of-the-art digital archive solutions.

References

1. Bell, J., Lewis, S.: Using OAI-PMH and METS for exporting metadata and digital objects between repositories. Program: Electronic Library and Information Systems 40(3), 268–276 (2006)
2. Cook, T.: Archival Science and Postmodernism: New Formulations For Old Concepts. Archival Science 1(1), 3–24 (2001)
3. Discovery, E., Shaw, S., Reynolds, P.: Creating the Next Generation of Archival Finding Aids. D-Lib Magazine 13(5/6) (May/June 2007)
4. Duranti, L.: The Archival Bond. Archives and Museum Informatics 11(3–4), 213–218 (1997)

5. Ferro, N., Silvello, G.: Modeling Archives by means of OAI-ORE. In: Agosti, M., Esposito, F., Ferilli, S., Ferro, N. (eds.) IRCDL 2012. CCIS, vol. 354, pp. 216–227. Springer, Heidelberg (2013)
6. Ferro, N., Silvello, G.: NESTOR: A Formal Model for Digital Archives. Information Processing & Management (in print, 2013), http://dx.doi.org/10.1016/j.ipm.2013.05.001
7. Fox, E.A., Gonçalves, M.A., Shen, R.: Theoretical Foundations for Digital Libraries: The 5S (Societies, Scenarios, Spaces, Structures, Streams) Approach. Morgan & Claypool Publishers, USA (2012)
8. Gilliland-Swetland, A.J.: Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment. Council on Library and Information Resources (2000)
9. Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A.: Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. ACM Transactions on Information Systems (TOIS) 22(2), 270–312 (2004)
10. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. In: Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, USA (2011)
11. International Council on Archives. In: ISAD(G): General International Standard Archival Description, 2nd edn., International Council on Archives, Ottawa (1999)
12. Kaplan, D., Sauer, A., Wilczek, E.: Archival description in OAI-ORE. Journal of Digital Information 12(2) (2011)
13. Kozievitch, N.P., da S. Torres, R.: Describing OAI-ORE from the 5S Framework Perspective. In: Chowdhury, G., Koo, C., Hunter, J. (eds.) ICADL 2010. LNCS, vol. 6102, pp. 260–261. Springer, Heidelberg (2010)
14. Lagoze, C., Van De Sompel, H., Nelson, M., Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0 (December 2008), http://www.openarchives.org/OAI/openarchivesprotocol.html
15. University of California. California Digital Library. CDL Guidelines for Digital Objects. Version 2.0. Technical report, 34 pages (January 2011)
16. Prom, C.J., Habing, T.G.: Using the Open Archives Initiative Protocols with EAD. In: Hersh, W., Marchionini, G. (eds.) Proc. 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2002), pp. 171–180. ACM Press, New York (2002)
17. Ridener, J.: From Polders to Postmodernism: A Concise History of Archival Theory. Litwin Books (2009)
18. Ross, S.: Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries. New Rev. Inf. Netw. 17(1), 43–68 (2012)
19. Society of American Archivists. Encoded Archival Description: Tag Library, ver. 2002. Society of American Archivists (2003), http://www.loc.gov/ead/tglib/
20. Van de Sompel, H., Lagoze, C.: Interoperability for the Discovery, Use, and Re-Use of Units of Scholarly Communication. CTWatch Quarterly 3(3) (2007)
21. Yako, S.: It Complicated: Barriers to EAD Implementation. American Archivist 71(2), 456–475 (2008)