NESTOR: A formal model for digital archives

Nicola Ferro, Gianmaria Silvello *

Department of Information Engineering, University of Padua, Via Gradenigo 6/B, 35131 Padua, Italy

ARTICLE INFO

ABSTRACT

Archives are an extremely valuable part of our cultural heritage since they represent the trace of the activities of a physical or juridical person in the course of their business. 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 archives. 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. NESTOR is composed of two set-based data models: the Nested Sets Model (NS-M) and the Inverse Nested Sets Model (INS-M) that express the hierarchical relationships between objects through the inclusion property between sets. We formally study the properties of these models and prove their equivalence with the notion of hierarchy entailed by archives.

We then use NESTOR to extend the 5S model in order to take into account the specific features of archives and to tailor the notion of digital library accordingly. This offers the possibility of opening up the full wealth of DL methods and technologies to archives. We demonstrate the impact of NESTOR on this problem through three example use cases.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past two decades, digital libraries have been steadily evolving and shaping the way people and institutions access and interact with our cultural heritage, study and learn (Borgman, 1999, 2003; Fox et al., 1993; Fox et al., 1995; Fox et al., 2012; Lesk, 1997; Marchionini and Maurer, 1995; Witten and Bainbridge, 2003). Nowadays, the reach of digital libraries 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 methods for accessing resources.

Archives are not simply constituted by a series of objects that have been accumulated and filed with the passing of time – as usually happens with libraries that collect, for example, individual published books, journals, and serials. Instead, archives 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.

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

---

* Corresponding author. Tel: +39 049 827 79 29; fax: +39 049 827 77 99.
  E-mail addresses: silvello@dei.unipd.it, gianmaria.silvello@gmail.com (G. Silvello).

(Gilliland-Swetland, 2000). 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* 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 (Vitali, 2010). 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. These archival units have a homogeneous nature and can, in turn, be divided into subunits containing items such as letters, reports, contracts, testaments, photographs, and drawings (International Council on Archives, 1999).

Digital libraries benefit from the existence of sophisticated formal models, such as the Structures, Spaces, Scenarios, Societies (5S) model (Fox et al., 2012; Gonçalves, Fox, et al., 2004; Gonçalves et al., 2008), 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, such as the notion of *archival bond*. Nor can we exploit the 5S model as it is for archives because, as we will discuss later on, it needs to be extended and adapted to archives.

In this article we highlight the central role of formal models for the digital library, because integration and cooperation between these models can enhance the 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 digital library 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 notions 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 contrast to the binary relation between nodes exploited by the tree (Aho and Ullman, 1992).

Then we exploit NESTOR to formally extend the 5S model to define a *digital archive* as a specific case of digital library able to take into consideration the peculiar features of archives. 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 digital library technologies and methods; and (iii) enables integrated access to heterogeneous contents.

As concrete accounts of this and as substantial examples of their application, we apply NESTOR and the extended 5S model to three typical scenarios for digital archives and overcome well-known issues in the field. The first is called "detaching the archives" which is the case of interoperability between digital archives where we formally exploit the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to demonstrate how digital library technologies can be adopted with archives. The second scenario is called "unchaining the archives" which shows how archives modeled with NESTOR can form compound digital objects made available as Linked Open Data (LOD) (Heath and Bizer, 2011) on the Web adopting Open Archives Initiative Object Reuse and Exchange (OAI-ORE) as a working framework. Finally, the third scenario is called "socializing the archives" which describes how NESTOR together with the Flexible Annotation Semantic Tool (FAST) (Agosti and Ferro, 2008) can enhance the role of annotations in archives by helping both archivists and end-users in the description and interpretation of archival resources.

The paper is organized as follows. Section 2 provides an intuitive overview of the principles underlying the two set data models composing NESTOR (i.e. the NS-M and INS-M) and a presentation of the main contributions of this work. In Section 3 we provide some background on archives, formal models for digital libraries and discuss the related work about nested sets methodologies. In Section 4 we formally present NESTOR along with its properties. Section 5 shows the equivalence between NESTOR and the archival trees. In Section 6 we introduce our extension to the 5S model via NESTOR and in Sections 7–9 we apply NESTOR and this extension to three case studies. We draw conclusions and point to future work in Section 10. In Appendix A we report all the proofs of the properties and theorems presented in Sections 4 and 5.

2. NESTOR: overview and contributions

2.1. Intuitive overview of the model

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. If we consider the Chinese boxes metaphor, a hierarchy is composed of a sequence of boxes contained one inside the other; if we look at an archive from the physical point-of-view, we can see that it resembles the 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. Indeed, NESTOR comprises both the structure and the content of the archive, where the inclusion relationships represent the structure and the elements belonging to the sets represent the content.

To illustrate the basic ideas behind NESTOR, let us consider an archive composed of six divisions: a fonds, two sub-fonds, and three series.