FRBR: A Generalized Approach to Dublin Core Application Profiles

Maja Žumer
University of Ljubljana, Slovenia
Maja.Zumer@ff.uni-lj.si

Marcia Lei Zeng
Kent State University, USA
mzeng@kent.edu

Athena Salaba
Kent State University, USA
asalaba@kent.edu

Abstract

According to the Singapore Framework, any development of a Dublin Core Application Profile (DCAP) has to include the creation of a domain model. DC Scholarly Works Application Profile (SWAP) was the first one explicitly using the Functional Requirements for Bibliographic Records (FRBR) model in creating its domain model. FRBR has recently been extended with Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD) thus forming the so-called FRBR family. This paper first further develops the SWAP domain model to incorporate the FRBR family models. Then a generalized FRBR-family-based DCAP domain model is presented to be used as the basis for specific domain application profiles.

Keywords: application profiles; Dublin Core Application Profiles; DCAP; FRBR; SWAP; scholarly works; domain model

1. Introduction and Background

Functional Requirements for Bibliographic Records (FRBR) has already been used as an approach in domain model development, a step towards building a Dublin Core Application Profile (DCAP) (Coyle and Baker, 2009). A domain model is a description of the *things* the metadata will describe, and the relationships between those things. It is the basic blueprint for the construction of the application profile (Coyle and Baker, 2009). The Dublin Core (DC) Scholarly Works Application Profile (SWAP) was the first explicitly using FRBR as the base for developing its domain model (Johnston and Powell, 2009). SWAP was developed between 2005 and 2008, which was also a critical time period for the development of two complementary models of FRBR: Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD). The two models were released in 2009 and 2010 respectively, thus forming the so-called "FRBR family." This paper presents a generalized approach for domain models beyond scholarly works by modifying and extending SWAP for general applicability while taking into account these more recent FRBR family developments.

1.1. The FRBR Family: FRBR, FRAD and FRSAD

The IFLA Functional Requirements for Bibliographic Records (FRBR) conceptual model, published in 1998, focuses on the representation of the bibliographic universe, using an entity-relationship model (FRBR, 1998). FRBR defines three groups of entities. Group 1 comprises of the entities *Work* (distinct intellectual creation), *Expression* (realization of a work), *Manifestation* (embodiments of an expression) and *Item* (exemplar of a manifestation). Group 1 can be defined as all products of creative or artistic endeavor. Group 2 includes the entities *Person* (an individual) and *Corporate body* (organization or group of individuals). This group can be defined as all agents who are linked through a relationship with Group 1 entities, such as those responsible for the creation, production, distribution and ownership. The third group, Group 3, includes the entities *Concept*, *Object*, *Place* and *Event.* This Group can be defined as anything that can be the subject of a *work* (in addition to all Group 1 and 2 entities). Even though FRBR defined all entities and identified their attributes and relationships among entities, the main focus

of this model was on Group 1 entities and the remainder of entities from Groups 2 and 3 only as they relate to bibliographic information.

A second conceptual model, Functional Requirements for Authority Data (FRAD), was developed to cover mainly Group 2 entities and any other entity as it relates to authority (i.e., controlled access) information (FRANAR, 2009). FRAD added a new Group 2 entity, *Family*, in order to respond to the needs of archival collections. In addition, FRAD defined a number of new entities, attributes and relationships that correspond to the needs of authority data. New authority data entities in FRAD include *Name* (a character or group of words and/or characters by which an entity is known in the real world), *Identifier* (a number, code, phrase, etc. uniquely associated with an entity), *Controlled access point* (the term by which a record can be found), *Rules* (instructions governing the formulation of a controlled access point), and *Agency* (an organization that creates and maintains controlled access points).

The third conceptual model in the FRBR family is the Functional Requirements for Subject Authority Data (FRSAD). It should be noted that it is an extension of FRBR and has been an independent development parallel to FRAD. The focus of this model is to identify entities, attributes and relationships as they relate to subject authority data (FRSAR, 2010). Two main entities have been defined in FRSAD, *Thema* (any entity used as a subject of a *Work*) and *Nomen* (any sign or sequence of signs that a *Thema* is known by, referred to or addressed as).

The following figure is an overview of the FRBR family, representing the main entities and primary relationships. All relationships are declared in both directions, left-to-right first. The arrows are used to indicate the cardinality of the relations.

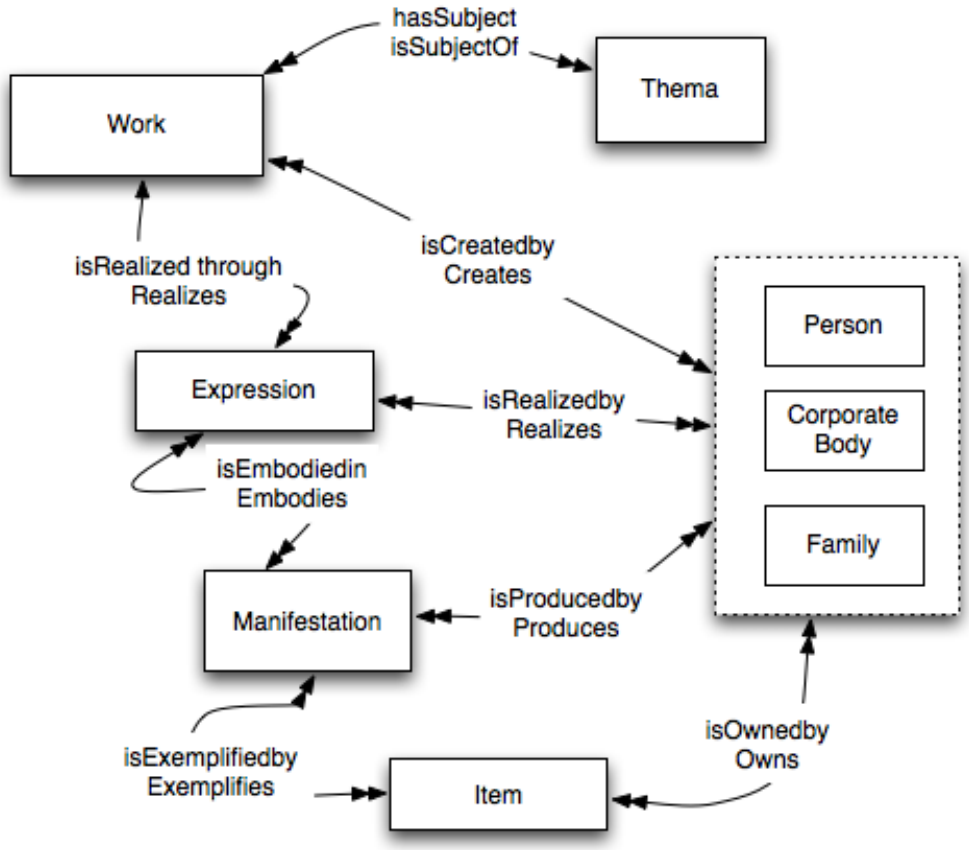

FIG. 1. "FRBR Family" Overview

After the latest model of the FRBR family, FRSAD, is published, a working group will need to be established within IFLA to develop a comprehensive harmonized model, covering all aspects of the bibliographic universe.

1.2. Domain Model Requirements in DCAP Related Documents

The Dublin Core Metadata Initiative (DCMI) groups have provided a series of recommended resources to guide the development of application profiles:

- The *Singapore Framework for Dublin Core Application Profiles* defines a set of descriptive components that are necessary or useful for documenting an application profile and describes how they relate to standard domain models as well as the Semantic Web fundamental standards. According to the Framework, a domain model is mandatory. The purpose of the domain model is to define a basic scope for the application profile, including the basic entities described by the application profile and their fundamental relationships (Nilsson, et al., 2008).
- The *Guidelines for Dublin Core Application Profiles* document provides a framework for the content and structure of any Dublin Core Application Profile (DCAP). The document explains the key components of a Dublin Core Application Profile and walks through the process of developing a profile. According to these guidelines, "[a] DCAP is a document (or set of documents) that specifies and describes the metadata used in a particular application. To accomplish this, a profile:
    - describes what a community wants to accomplish with its application (Functional Requirements);
    - characterizes the types of things described by the metadata and their relationships (Domain Model);
    - enumerates the metadata terms to be used and the rules for their use (Description Set Profile and Usage Guidelines); and
    - defines the machine syntax that will be used to encode the data (Syntax Guidelines and Data Formats)" (Coyle and Baker, 2009).

  A section on "Selecting or Developing a Domain model" presents two domain models as examples: a simple model and a FRBR-based model (Coyle and Baker, 2009).
- The *DCMI Criteria for the Review of Application Profiles* document states that "[a]n application profile MUST provide a data model, if only a simple one, which describes the entities and relationships among the entities. … An application profile can be based on an externally defined data model. With regard to the data model the following questions have to be answered:
    - Does the model depict the set of entities to be described and the relationships among those entities?
    - If an application profile uses an externally defined data model:
        - Is the external data model identified?
        - Are deviations from the externally defined data model documented?" (DCMI Usage Board, 2009).

1.3. DCAP Development and the Domain Models

Several DCAPs have been developed throughout the years. (Note: all DCAPs referred to in this section are listed at the end of the paper.) The main points of reference for the review criteria applied to application profiles are the *Singapore Framework for Dublin Core Application Profiles*, the *DC Abstract Model* and a draft *Description Set Profile* specification (Nilsson, 2008).

Best-practice examples of application profiles include *Dublin Core Collections Application Profile* (DC-CAP) and *Scholarly Works Application Profile* (SWAP), both of which were reviewed by the DCMI Usage Board and became DC Application Profiles.

- *Dublin Core Collections Application Profile* (DC-CAP), published in March 2007, specifies how to construct a DC metadata description set that provides a collection level description. It provides a means for creating simple descriptions of collections (i.e., aggregations of physical or digital resources), as well as simple descriptions of catalogues and indexes (i.e., aggregations of the metadata that describe the collections). It employs a simple entity-relationship model for collections and their relationships to other entities. The entities are:
    - *Collection* -- An aggregation of *Items*.
    - *Item* -- A physical or digital resource.
    - *Location* -- A place where a *Collection* is held.
    - *Service* -- A system that provides access to the *Items* within the *Collection.*
    - *Catalogue or Index* -- An aggregation of *Items*, which describes a *Collection.*

  The DC Collections AP describes the use of properties to represent attributes of the *Collection* and of the *Catalogue or Index* as well as relationships between or among the entities *Collection* and *Catalogue or Index*. Another entity, *Agent*, is also presented in the context of the model but is not further defined (Dublin Core Collection Description Task Group, 2007).
- The Scholarly Works Application Profile (SWAP) was developed in 2006 and reviewed by the DCMI Usage Board in 2009 in order to provide a method for describing scholarly works. A *Scholarly Work* is defined as a distinct intellectual or artistic scholarly creation. The model comprises of the following five entities:
    - *Scholarly Work*
    - *Expression*
    - *Manifestation*
    - *Copy*
    - *Agent* (Allison et al., 2007).

  This application profile provides a way of describing these entities as part of a description set (a set of related DC descriptions). Details of this domain model will be discussed in Section 2 below.
- The *Images Application Profile* (IAP), a DCAP, came out from the work undertaken within the JISC Digital Repositories programme (UK) for describing images held in institutional repositories. The research concluded that FRBR could be used quite successfully to model some image types, particularly those that are the product of an artistic or intellectual process. However, the group found that the *Expression* entity is not applicable to all types of images and therefore a decision was made to exclude *Expression* from IAP domain model. According to IAP, any change in an image produces a new *Manifestation*. The group considered the possible consequences as IAP merges with other related FRBR-based repository profiles (Images Application Profile, 2008).

Other DCAP specifications that are in draft status include the following:

- The *DC Library Application Profile* (DC-Lib) intended to clarify the use of the Dublin Core Metadata Element Set in libraries and library-related applications and projects (DCMI-Libraries Working Group, 2004).
- The *DC-Education Application Profile* (DC-Ed AP) intended to define metadata elements for use in describing properties of resources related to their use in teaching and learning (DC-Education Application Profile Task Group, 2007).

- The *DC-Government Application Profile* (DC-GAP) task group was initiated to define how to describe metadata for governmental resources using the Dublin Core Metadata Element Set (DCMI Government Community Webpage, 2006). There is no final AP as of the time this paper is written.

These AP working drafts were initiated before the publication of the Singapore Framework and therefore conform to the previous version of the *DCAP Guidelines*. No domain model was required at that time. The generalized model this paper proposes should also apply to these APs if/when their domain models are developed in the future.

2. Extension and Generalization of the DC-SWAP Domain Model

With the extension of FRBR and the resulting FRBR family, there is a possibility to extend and generalize the existing DC-SWAP model.

2.1 Domain Model of the Scholarly Works Application Profile

DC-SWAP focuses on e-prints (including scholarly works, research papers or scholarly research texts), particularly for the purpose of repositories and aggregator services. In the initial stages, the DC-SWAP working group specified the goals of the application profile. General goals include consistency, richness (compared to simple DC), compatibility with library data, extensibility, and support for future added-value services. More specific goals include: data for preservation purposes, version control, identification of the most appropriate copy, access and rights management, identification of provenance and funding, support for browsing/filtering, use of controlled vocabularies and authority control (Eprints Application Profile Functional Requirements Specification, 2006).

As seen in Figure 2, the DC-SWAP model is based on FRBR with some modifications, particularly in labels for entities and relationships (Allison, et al., 2007).

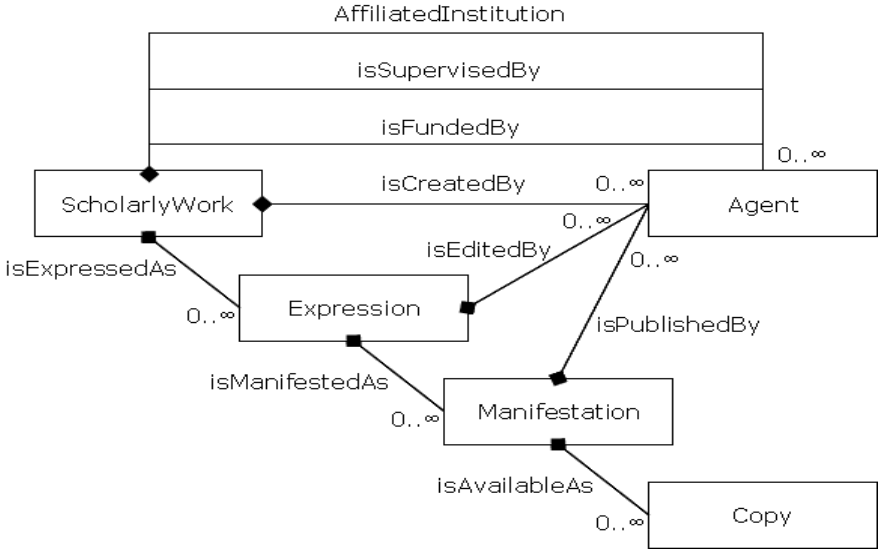

FIG. 2. SWAP Application Profile Domain Model,
Source: Allison, Johnston, and Powell. (2007). A Dublin Core Application Profile for Scholarly Works.

According to the SWAP Application Model presented by Allison, et al. (2007), it can be summarized that SWAP has introduced the following changes to FRBR:

- a *ScholarlyWork* is a FRBR *Work*
- a *Copy* is a FRBR *Item*
- an *Agent* is a FRBR *Person* or a FRBR *Corporate Body*

- the 'isExpressedAs' relationship is known as 'is realized through' in FRBR
- the 'isManifestedAs' relationship is known as 'is embodied in' in FRBR
- the 'isAvailableAs' relationship is known as 'is exemplified by' in FRBR
- the 'isCreatedBy' relationship is known as 'is created by' in FRBR
- the 'isPublishedBy' relationship is modelled as the "publisher" attribute of a *Manifestation* in FRBR.

2.2 Generalization and Extension of SWAP Domain Model

Keeping the majority of the chosen labels in SWAP, with minimum changes, while also taking into account the "FRBR Family," particularly FRSAD, the model could be generalized as shown in Figure 3.

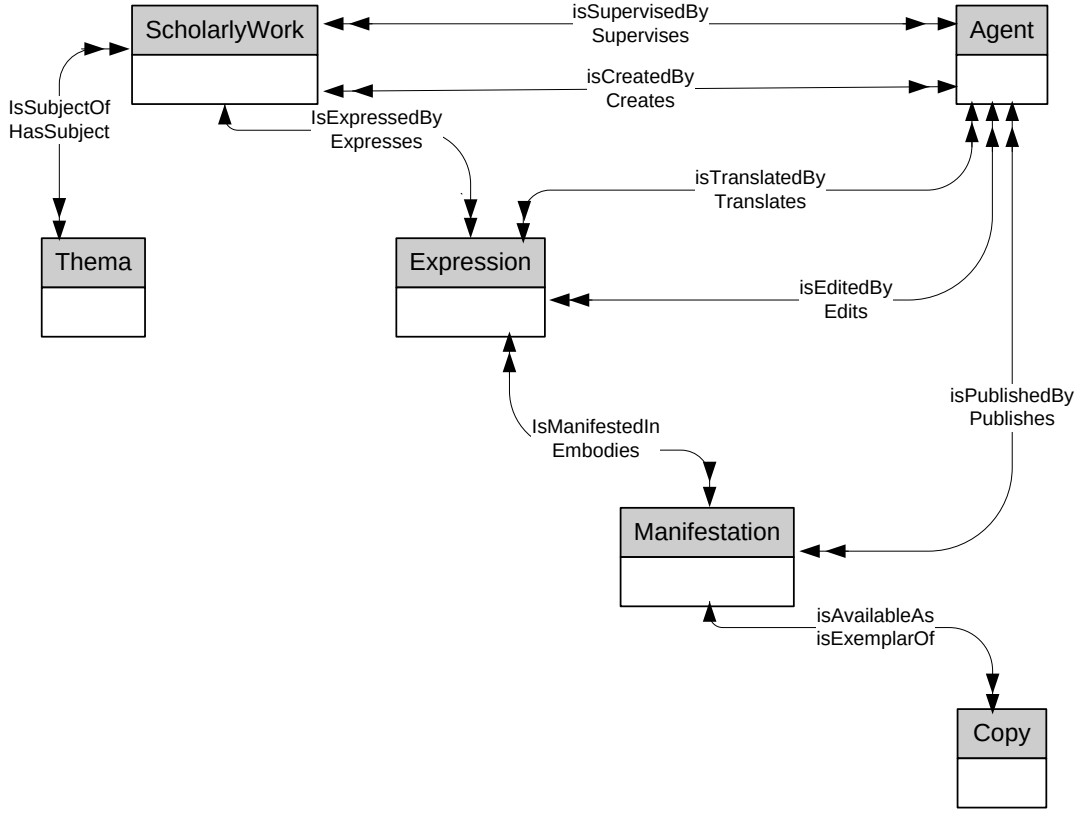

FIG. 3. Extended DC-SWAP Model

Some differences need to be pointed out:

1. All relationships are declared in both directions, left-to-right first. The arrows are indicating the cardinality of the relations, in line with FRBR conventions.
2. The most important change is the relationship between *Expression* and *Manifestation*: in addition to the obvious fact that any instance of *Expression* can be embodied in ('isManifestedIn') one or more *Manifestations*, we need to point out that any instance of *Manifestation* may embody one or more instances of *Expressions*. Examples of *Manifestations* embodying more *Expressions* include texts with illustrations, a journal issue, or any collection of articles. Here the relationship label has been changed from 'isManifestedAs' to 'isManifestedIn' to reflect the situation.
3. The 'hasSubject' relationship from *Work* to *Thema* is added. In SWAP "subject" is an attribute of a *ScholarlyWork*.

In practical terms, the *Copy* (FRBR *Item*) is not a particularly interesting or important part of the model in the e-print environment, because an instance of *Copy* is the actual document obtained by the user on his/her computer. In general, there would be no particular metadata describing it, since all important information guiding the access to and use of resources is tied to the *Manifestation*.

In the following discussion we are focusing on *ScholarlyWork*, *Expression* and *Manifestation* only. Built on what was introduced in the SWAP original model, Figure 4 presents a generalization of the 'hasPart' and 'isFundedBy' relationships since both can be applied to any Group 1 entity.

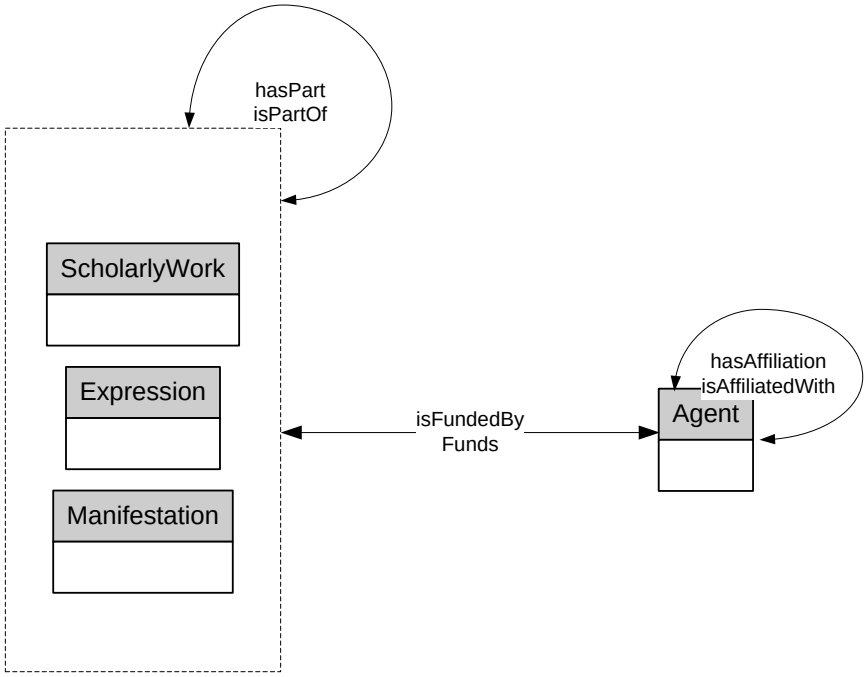

FIG. 4. 'HasPart' and 'isFundedBy' Relationships Declared on all Group 1 Entities and 'hasAffiliation' Relationship Between Instances of *Agent* Entity

Moreover, the relationships between Group1 and *Agent* can be further generalized, as shown in Figure 5. The 'isSupportedBy' relationship may include any funding or non-material support; 'isSupervisedBy' relationship may include a formal (such as thesis advising) or informal supervision; 'rightsControlledBy' relationship serves as a mechanism for attributing rights. In particular implementations the general relationships, shown here, may be replaced by more specific ones.

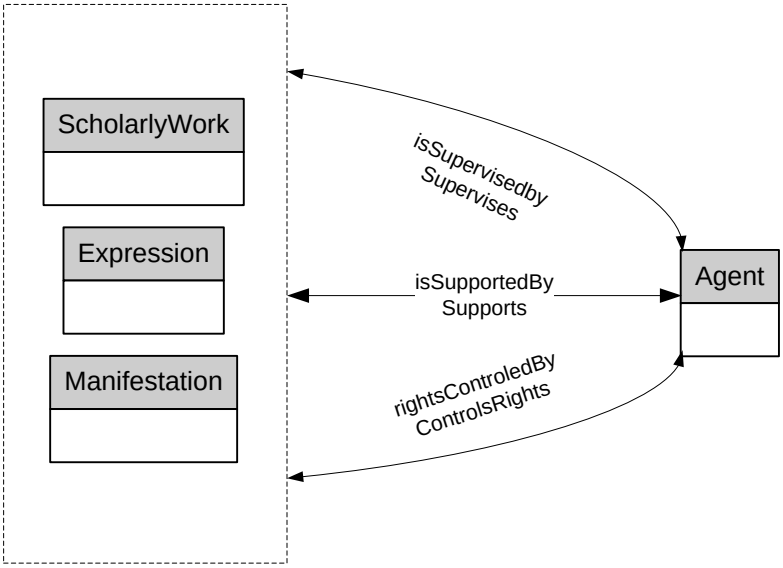

FIG. 5. Relationships 'isSupportedBy', 'isSupervisedBy' and 'rightsControlledBy' between Group 1 entities and *Agent*

3. Towards a General DCAP Domain Model

As suggested in the *Guidelines for Dublin Core Application Profiles* (2009), the functional requirements are the first step: we have to define the activities that are to be supported by metadata. FRBR user tasks are general and applicable to all information resources:

- to **find** entities that correspond to the user's search criteria
- to **identify** an entity
- to **select** an entity that is appropriate to the user's needs
- to acquire or **obtain** access to the entity described (FRBR, 1998).

"Explore" was added in FRSAD and comprises of the need to investigate the bibliographic world and/or domain:

- to **explore** relationships between entities (FRSAR, 2010).

3.1 General Domain Model

A general DCAP domain model is needed for ensuring the interoperability of all application profiles and metadata created using these APs. A general AP domain model will not cover all specific tasks or specific features of the different forms/genre/types of resources. These special needs are, and should be, taken care of in specific APs.

The General DCAP Domain Model (Figure 6) is proposed based on previous developments. In The General DCAP Domain Model, the relationships are declared on the most general level; any particular AP will then define more specific relationships, for example 'isFundedBy' could replace 'isSupportedBy.' Similarly the general 'isCreatedBy' can be refined as 'isTranslatedBy' (in the case of *Expression*) or 'isPublishedBy' for *Manifestation.*

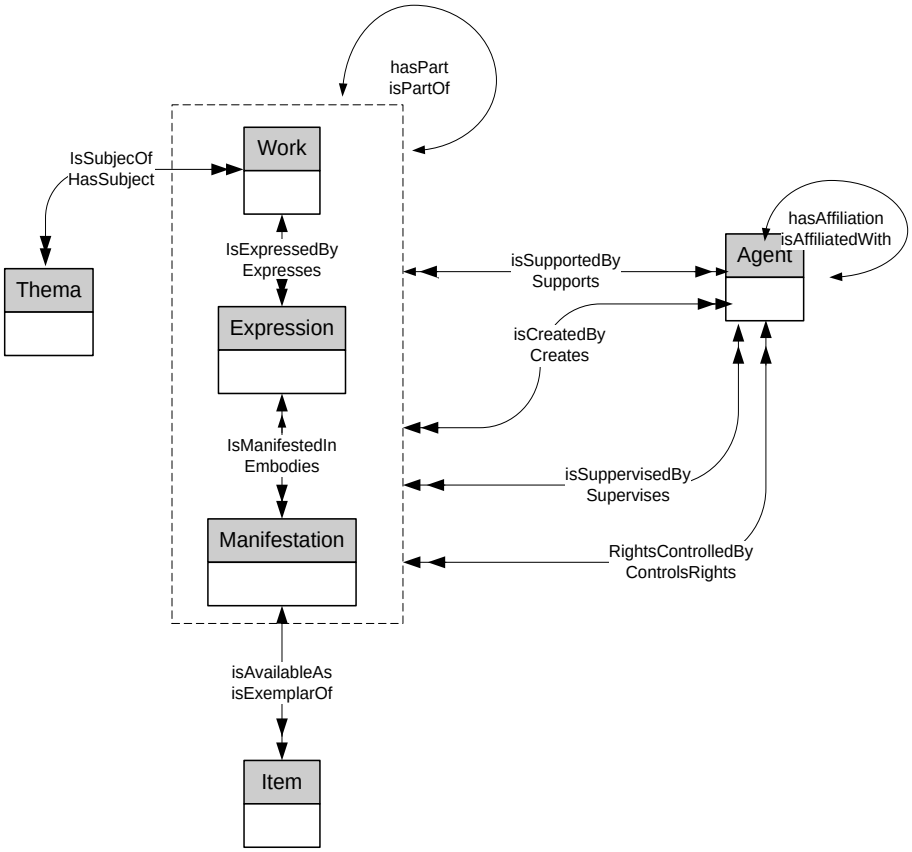

FIG. 6. General DCAP Domain Model

3.2 Basis for Authority Control

Figure 7 presents the general relationship between a "thing" and its appellation as modeled by FRSAD. Including an entity for the appellation is necessary to model the authority control process, an important mechanism of controlling variant appellations. If authority control is not implemented, the appellation (title, name, label etc.) of any entity may be modeled as an attribute of that entity instead of being an independent entity itself. Such a practice not only limits the opportunity of defining appellation attributes and the relationships between different appellations, but also mixes inter-concept semantic relationships with the inter-appellation lexical relationships.

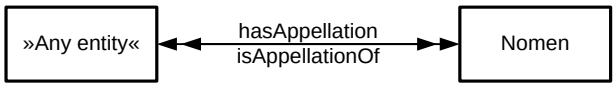

FIG. 7. The "has Appellation" Relationship as Basis for Authority Control

In the framework of FRBR family, "Any entity" refers to any of the entity types in the model. Defining *Nomen* as an entity (instead of an attribute) enables management of appellations. As with the general relationships discussed earlier, a specific AP may declare and define additional entities and attributes to better fit the stated goals of the application.

4. Conclusion

A number of Dublin Core Application Profile domain models have been developed and two of them (most significantly SWAP) are based on the Functional Requirements for Bibliographic Records (FRBR) entity-relationship model. FRBR has recently been extended with two models, FRAD and FRSAD, both focusing on authority control. The three models are now known as the FRBR-family. This paper first expanded DC-SWAP domain model, taking into account the FRBR family. This domain model is then further developed into a generalized FRBR-family-based DCAP domain model. Such a general model could support the development of compatible specific domain models, which in turn would result in interoperable DC application profiles.

References

Allison, Julie, Pete Johnston, and Andy Powell. (2007). A Dublin Core application profile for scholarly works. Retrieved March 25, 2010, from http://www.ariadne.ac.uk/issue50/allinson-et-al/

Coyle, Karen and Thomas Baker. (2009). Guidelines for Dublin Core application profiles. Retrieved March 25, 2010, from http://dublincore.org/documents/profile-guidelines/

DCMI Usage Board. (2009). Criteria for the review of application profiles. Retrieved March 25, 2010, from http://dublincore.org/documents/2009/03/02/profile-review-criteria/

Eprints application profile functional requirements specification. (2006). Retrieved July 12, 2010, from http://www.ukoln.ac.uk/repositories/digirep/index/Functional_Requirements

IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR). (1998). Functional requirements for bibliographic records - final report. Munich: K.G. Saur. Also retrieved March 25, 2010, from http://www.ifla.org/VII/s13/frbr/index.htm

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). (2009.) Functional requirements for authority data – A Conceptual model. Edited by Glenn E. Patton. München: K.G. Saur.

IFLA Working Group on the functional Requirements for Subject Authority Records (FRSAR). (2010). Functional requirements for subject authority data – A Conceptual model. Edited by Marcia L. Zeng, Maja Žumer, Athena Salaba. Retrieved July 12, 2010, from http://www.ifla.org/node/1297

Nilsson, Mikael, Thomas Baker, and Pete Johnston. (2008). The Singapore Framework for Dublin Core application profiles. Retrieved March 25, 2010, from http://dublincore.org/documents/2008/01/14/singapore-framework/

Nilsson, Mikael. (2008). Description set profiles: A constraint language for Dublin Core application profiles. Retrieved March 25, 2010, from http://dublincore.org/documents/2008/03/31/dc-dsp/

DC Application Profiles Referred to in the Paper

Dublin Core scholarly works application profile (SWAP). SWAP Working Group, JISC Digital Repositories programme. Coordinated by Julie Allinson and Andy Powell. Retrieved March 25, 2010, from http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile

Dublin Core Collection Description Task Group. Dublin Core collections application profile. Retrieved March 25, 2010, from http://dublincore.org/groups/collections/collection-application-profile/2007-03-09/

JISC Digital Repositories Programme. Images application profile. Coordinated by Polly Christie and Mick Eadi Retrieved March 25, 2010, from http://www.ukoln.ac.uk/repositories/digirep/index/Images_Application_Profile

DCMI-Libraries Working Group. Dublin Core library application profile (DC-Lib). Retrieved March 25, 2010, from http://www.dublincore.org/documents/2004/09/10/library-application-profile/

DC-Education Application Profile Task Group. Dublin Core education application profile (Working Draft of v0.4). Retrieved March 25, 2010, from http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile

DCMI Government Application Profile Task Force. DC-Government application profile (DC-GAP). Retrieved March 25, 2010, from http://www.dublincore.org/dcgapwiki. Refer to: DCMI Government Community Webpage. Retrieved March 25, 2010, from http://dublincore.org/groups/government/index.shtml