JLIS.it

Metadata issues in Digital Libraries: key concepts and perspectives

Iryna Solodovnik

Brief overview of information management in Digital Library

Digital Library (DL) is a type of information retrieval system accessible by computers in which collections are stored in digital formats (as opposed to print, microform, or other media) (Greenstein and Thorin). In order to ensure effective search, access and retrieval of information resources in the DL environment, beside different metadata schemas for digital records (e.g. Dublin Core, DC) there can be used various metadata schemas based on classical methods, approaches and tools, such as bibliographic classification schemas. These last are used to provide for each document an appropriate classification system index reflecting content or information purpose of the document, and to place together (both physically and virtually) documents by subject and thematically. Cataloging practices are expressed through strictly defined rules on which collection of bibliographic records are created and are used to describe the document as a unique object that must be stored and retrieved, disclosing its individual preserved features. Today almost every library has a directory of electronic catalog with entries on the documents stored

JLIS.it. Vol. 2, n. 2 (Dicembre/December 2011).
DOI: 10.4403/jlis.it-4663

in its funds thus helping users to find and retrieve more efficiently the searched material. The task of entering information in catalog records requires considerable time and expertise, that's why in order to reduce costs and to avoid duplicated records, libraries (such as Library of Congress, national and major university libraries in different countries) often share their catalog records with each other for free.

Treating the issue of information management in DL, it is also beneficial make a reference to Functional Requirements for Bibliographic Records (FRBR), published by the International Federation of Library Association (IFLA) (IFLA Study Group on the Functional Requirements for Bibliographic Records; Gemberling; Ghilli and Guerrini). The novelty and originality of FRBR in the library practice is to try to develop a conceptual descriptive model for bibliographical records (Tillett, *FRBR: A Conceptual Model for the Bibliographic Universe*; Shotton and Peroni) through identifying optimal requirements of user's needs in searching and retrieving a bibliographical record from a catalog. Access and retrieval supported by FRBR are based on connections created between entities identifying context and content in a bibliographic record. In digital environment these connections are defined by links permitting to navigate through the hierarchy of different relations between entities. The originality of the FRBR is also reflected in its independency from specific cataloguing standards such as AACR2 or ISBD.

Apart from the three groups of entities:

1. Work, Expression, Manifestation, Item (WEMI reflecting the products of intellectual or artistic creation);

2. Person, Corporate body;

3. Concept, Object, Event, Place (intellectual endeavor of the subjects of first or second groups) and their possible connec-

> tions, FRBR defines four user tasks: to Find, to Identify, to Select, to Obtain. Accordingly, the user should be able to utilize the available data in order to: find the material that meets the search criteria set; identify an entity; select an entity that matches his/her needs; gain access to the described entity.

A fifth, informal task is to Navigate or Relate.

In figure 1 there is schematized the FRBR model (groups 1 and 2) built upon relationships between and among entities. These "relationships serve as the vehicle for depicting links between one entity and another, thus providing assistance for the user to 'navigate' the universe that is represented in a bibliography, catalogue or bibliographic database".[1]

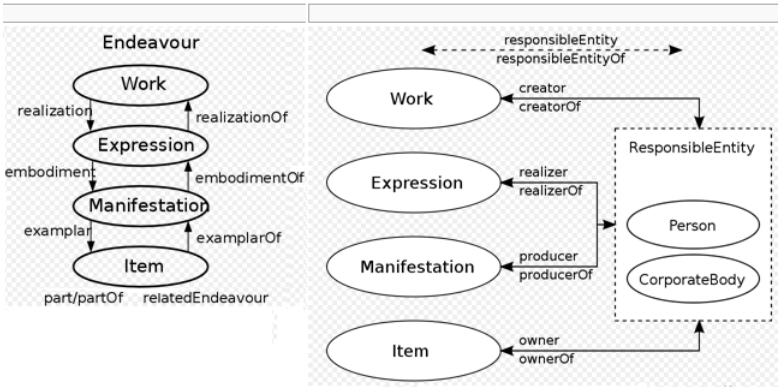

Figure 1: Entities and basic relations defined by FRBR (groups 1 and 2)

Each entity of the FRBR model has an associated set of attributes (directly related and external to the entity) by which user can formulate his/her information requests and interpret the results of the

[1] http://archive.ifla.org/VII/s13/frbr/frbr_current5.htm.

bibliographical search querying the relational entities (Christopher). FRBR should not only be seen as a conceptual model, but as an "element vocabulary" when used in Linked Data (Baker, *Designing Interoperable Metadata on Linked Data Principles. Tutorial: Dublin Core -Building blocks for interoperability*).

No matter what information management practice can be chosen to process bibliographic data, the quality of bibliographic records depends on their adequacy to the requirements of bibliographic standards, as well as on their conformation to cataloguing and classification rules, to thesauri and controlled vocabularies, called not only to define and describe records formally but also to normalize their data attributes and values and thereby to create controlled access enhancing uniform retrieval, interoperability and reuse of bibliographic data. In this perspective, Universal Bibliographic Control (IFLA UBC) "has traditionally imposed order upon the seeming chaos of the bibliographic universe, by subjecting individual documents to a rigorously-structured set of metonymic descriptions that enable these documents to be found under various useful circumstances" (Campbell). Information management in Digital Library requires the participation of suitably qualified cataloguers, an enormous and time-consuming indexing and classification work, mastery of a complex set of standards and appropriate descriptive schemas. When it comes to the rapidly growing digital resources and their sustainability by metadata schemas and tradition cataloguing rules, "there has been a yearning among knowledge organization professionals to find more efficient and accurate means for providing resource description" (Smiraglia).

Metadata system and metadata schema

Just as traditional cataloging organizes the entire set of data items (e.g. by author, title, date, subject, coverage, number of location of the item on the shelf etc.) stored in a system of library collections, metadata system is considered to be central logical component of any DL (Gartner; Greenstein and Thorin) which makes use of metadata in online library catalogues (Birrell, Dunsire, and Menzies) belonging to Integrated Library Management System (ILMS). Any metadata system for the online library catalog may contain information on both library's existing holdings (stored in ILMS, using MARC metadata standard) and complex born-digital works requiring substantially more effort in their representation, management and preservation. With "the rise of new metadata systems, which infiltrate the documents themselves and affect the way they are used, cataloguing has moved into a new daunting territory, in which our traditional role as information intermediaries has become less rigidly defined and more subtly powerful" (Campbell). Anyway, as considered by traditional cataloguers, in order to maintain the requirements of quality in description of bibliographic resources, metadata systems are called to be continuously confronted with and adopted to requirements of authority data inherent in commonly shared descriptive cataloging rules, classification schemas, controlled vocabularies etc. (Gorman, "Metadata or cataloguing? A false choices"; Kurth, Ruddy, and Rupp) thus, through the means of authority control, to better identify, select, locate and permit a qualitative resource discovery, manipulability and portability (Hillmann, Dushay, and Phipps; El-Sherbini and Klim). In the library environment, the term schema – generally understood as a "structured framework" – can be applied to classificatory (DDC etc.) and terminological systems (Thesauri, LCSH etc.) as well as to the content standards (e.g. AACR) and to other rich semantic container-like

schemas (e.g. ontologies) which normally require the use, within a pre-determined number of content elements or metadata (the extent of which is depending on granularity or refinement of description), of the acceptable data with normalized values.[2] According to Greenberg ("Understanding Metadata and Metadata Schemas"), metadata schema is considered to be:

> A collection of metadata elements, forming a structured container, to which data values are added. Data values can be uncontrolled or controlled (e.g., taken from a source such as LCSH or a standardized list of values).

The complex metadata schemas with a higher number of structured elements related through determined properties and value types can be also intended as annotation ontologies (Valkeapää, Alm, and Hyvönen) combining "equivalent or similarly functioning metadata elements from two or more metadata schemas" (Greenberg, "Understanding Metadata and Metadata Schemas"). Let us now shortly return to the already cited metadata schema DC (Baker, *Basics of Dublin Core Metadata*) supported by the Dublin Core Metadata Initiative (DCMI) which aims at developing interoperable standards for online data as well as providing reference vocabularies and tools

[2]Reporting some of widely used metadata schemas supported by appropriate standards here we cite: Text Encoding Initiative (TEI), http://www.tei-c.org/index.xml; Encoded Archival Description (EAD), http://www.loc.gov/ead/; Dublin Core Metadata Element Set (DCMES), http://dublincore.org/documents/dces/; Visual Recourse Association's Core Categories (VRA), http://www.vraweb.org/projects/vracore4/; Gateway to Educational Materials (GEM), http://www.vraweb.org/projects/vracore4/; Learning Object Metadata (LOM), http://www.vraweb.org/projects/vracore4/; Moving Picture Experts Group (MPEG-21), http://searchsoa.techtarget.com/definition/MPEG-standards; Music Encoding Initiative (MEI), http://music-encoding.org/; Preservation Metadata (PREMIS), http://www.loc.gov/standards/premis/ (for links to preservation metadata see also DCMI Preservation Community, http://www.dublincore.org/groups/preservation); Metadata Encoding and Transmition Standard (METS), http://www.loc.gov/standards/mets/ etc.

to manage and encourage metadata mapping.

> DCMI has a potential role in promoting co-operation among vocabulary managers and in providing Best Practices for vocabulary alignment and interoperability. The inevitable and useful proliferation of vocabularies emerging in the Linked Data space demonstrates a need for increased vocabulary reuse and tools to facilitate this reuse, as well as central reference vocabularies and tools to manage and encourage vocabulary mapping (Metadatalibrarians post, 23 August 2011).

Dublin Core Metadata Element Set (DCMES),formalized by the standard ISO 15836, consists of 15 not qualified elements (Hillmann). These elements are considered to be essential to describe main characteristics of different bibliographical resources by representing a low-level general denominators for catalographic descriptions: DCMES can be seen as basic semantic layer for different bibliographic descriptions in digital environment that require the support of more complex metadata schemas. In other words, DC items can be seen as basic semantic elements for the different information resources on the web as well as for semantic descriptions supporting distributed bibliographical resources in DL which primarily were exclusively covered by MARC, ISBD and other catalographic descriptions. Being flexible enough (as well as optional and repeatable) to include any indication of semantic content and being officially translated in different languages, DC elements form a highly user-oriented informational model accepted worldwide for developing DL catalogs (Birrell, Dunsire, and Menzies).

In the digital world of scholarly communication supported by Open Access (OA) movement, the DC is selected as 'core metadata format' for Open Archives Initiative (OAI) architecture promoting OAI-PMH protocol important for the establishment of OA Repositories (discipline, institutional) (De Robbio; Guerrini) which can be a

constituent part of DL. The mandatory DC format within OAI-PHM architecture is called to provide basic information interoperability among different digital archives, whose contents might gain added semantic value in service providers able to create connections among various metadata schemas and specific contents.

> These principles consist of statements that highlight the need to establish a set of guidelines for data exchange, define a meta-language, recommend encoding principles, support compatibility with existing standards, and permit the conversion of resources to newer communication format (Greenberg, "Understanding Metadata and Metadata Schemas").

"Different ways may be needed to describe one object: MARC records for library catalogs, Dublin Core for simpler descriptions, specialized metadata for terms and conditions of use" (Baker, "Designing Interoperable Metadata on Linked Data Principles"). Anyway, it is important not only to select appropriate metadata schema/s but also to determine how metadata elements will depend on information content managed within DL, digital repository or archive, to document functional requirements for each metadata element, to qualitatively normalize metadata through data of appropriate authority lists, to make them interoperable through cross-walkers and harvesting mechanisms (Cole and Foulonneau) and, where it is possible, to enrich semantically metadata through mechanisms of ontologies, linked (open) data to further strengthen the quality of metadata environment.

Towards Semantic Interoperability

Just as internationally determined codes and standards (ISBD, AACR, authority control, subject access, classification systems,

> etc.) have fostered the goal of universal bibliographic control, so, too, have syntactic structures, semantic elements sets, transmission protocols, cross-schema mappings, and metadata harvesting tools been instrumental to realizing the concept of interoperability. In their respective roles as "change-agents" or catalysts to universal information access and exchange, bibliographic control and metadata seem less like "two solitudes", and more akin to soul-mates (Howarth).

In order to enhance interoperability in digital information environment, different metadata representations need to move towards integrated approaches and this should be based on the mechanisms of standardization, normalization and enrichment of metadata schemas through linguistic-formal and conceptual entities such as terminologies, controlled vocabularies, classifications, ontologies, linked data etc., whose formulation and implementation is an extremely important part of planning and design of any interoperable information environment that tends to correspond to requirements of quality.

> Local metadata applications are more and more likely to desire interoperability with other applications, or at least the ability to link outside themselves to other stores of information such as information about a personal name contained in an Authority file, or information about a subject term contained in an external subject Thesaurus (Chopey).

In this vision it is important to develop additional semantic layers on top of metadata descriptions, investing in correct semantic design of their models with explanatory power rather than prescriptive power. Such an approach involves ontological reasoning and development of more detailed information by data linking. The interoperability contributes to economies of scale that can be achieved by adopting the strategies aiming to harmonize access methodology across col-

lections around the world. It is important to move towards interoperability among different description communities (digital archives and libraries, visual art museums, geospatial data providers, government information bodies, education field etc.) (Dunsire et al. 2010) enhancing, first of all, the interoperability among metadata schemas. These last created within relatively independent metadata communities are invited to move towards "interoperability of combinations of metadata specifications" (Nilsson) through harmonization of metadata standards. Conceivably, the most fundamental notion underlying the will for creation of a successful digital metadata environment is a semantic interoperability and ontology metadata standardization (Palma, Hartmann, and Gomez-Perez). Ontologies have been more and more identified themselves as the most appropriated enabling technologies to support metadata creation in developing Semantic Web. There must be more widely discussed the relation between ontologies as conceptual models and ontologies as codification artifacts (Delteil, Faron-Zucker, and Dieng) in complex domains supported by different metadata schemas. It is also important to extend the implementation of semantic annotation systems supporting the creation of flexible indexes to different metadata schemas in distributed applications. Even though semantic annotation systems offering ontological reasoning on the web aim at enhancing the potentiality of information retrieval systems as well as semantic auto-completion of information and concept extraction, a major source of interoperability problems on the Semantic Web is still due to the use of different value vocabularies supporting metadata descriptions in different linguistic communities. Hence, these last must aim at developing joint projects in view of the creation and broad implementation of domain-general frameworks harmonizing distribution and documentation of linguistic diversity of different metadata schemas. This task can be achieved by the development

of organic terminological platforms accurately qualified through values of widely-accepted terminologies, thesauri and taxonomies representing types of knowledge structures used to provide semantic information about data, to generate controlled lists of values for various data elements as well as to uniform semantic annotation and categorization (Gambarara). Such a practical approach to Information Management aims to maintain content coherence and to reduce translation costs, as well as to query term expansion, autocompletion, and concept mapping based on predefined reference vocabularies of categories and terms for faceted browsing (Bradley; Broughton). For the purpose of normalization of metadata values by linguistic resources, coordinated and collaborative development of a standard common set of open source "value vocabularies" based on semantic metadata should be widely supported and promoted (Valkeapää, Alm, and Hyvönen).

Flexible ontology knowledge models should be widely implemented in digital information environment thus helping to interchange different conceptual values between the properties of (meta)data.

Reflecting on possible transitions of metadata to ontologies, it is of benefit to refer back to the pioneer ontology-based metadata tool for ontological conceptualization of bibliographic items and relationships developed within the University of Michigan Digital Library (UMDL) Beethoven Project. The created ontology-based system is aimed at facilitating cataloguing of bibliographical resources supported by MARC data which could be simply added to the suitable "bibliographical family" at the appropriate hierarchical level, thus automatically inheriting and sharing data with common properties among different overlapping bibliographic databases. The aforementioned tool is mainly points to

> three important purposes: to describe collection content, enabling computational inference to support powerful user queries;

> to organize the space of available agent services, in a dynamic way that fosters the evolution of the society of agents; to deduce the licenses required to provide (given) services for work.

Metadata repositories enriched by requirements of widely used semantic standards, such as, for example, KIF, OWL, RDF, SKOS etc.,[3] would be definitely more effective to classify, associate and interlink heterogeneous information resources in a consistent and unified fashion. Currently RDF is considered to be the most common tool used to encode, exchange and reuse different metadata schemas on the web. The RDF model is based on triple: Subject (resource) – Predicate (property; relation) – Object (value).The goal of RDF is to encode metadata syntax as well as to contribute to design interoperable (meta)data on the principles of Linked Data (Baker, *Basics of Dublin Core Metadata*) (see figure 2 on the next page) based on the web standardized technologies such as HTTP, Uniform Resource Identifiers (URIs) and RDF data model. Being related to RDF, OWL enhances the semantic extension of RDF data assigning an additional meaning to certain RDF triples, and specifying exactly which triples have a specific meaning and what this meaning is. OWL creates data interferences generating new knowledge on related information objects. It is also “possible to extend and reuse metadata specifications and vocabularies distributed in the web using the OWL, by utilizing the language’s flexibility to create restrictions on inherit properties and to make interferences on web distributed resources" (Bermudez and Piasecki).

> In the ‘graph’ paradigm, it becomes easier to envision how Library metadata interacts with other metadata on the Open

[3]Knowledge Interchange Format, http://logic.stanford.edu/kif/dpans.html; Web Ontology Language, http://www.w3.org/TR/owl-features; Resource Description Framework, http://www.w3.org/RDF; Simple Knowledge Organization System, http://www.w3.org/2001/sw/wiki/SKOS.

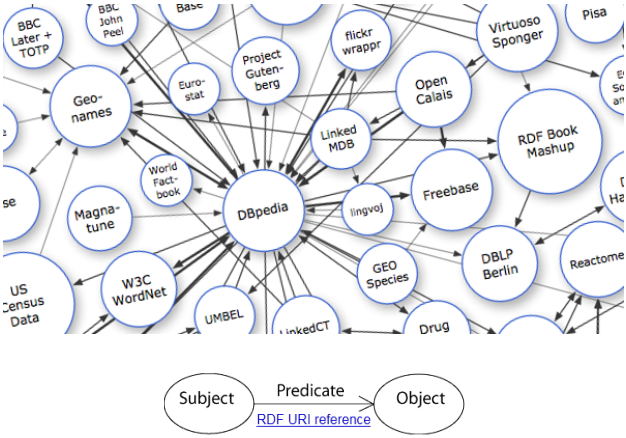

Figure 2: Linked Data. Part of the Linking Open Data (LOD) cloud diagram, http://lod-cloud.net/, http://linkeddata.org/

> web. As the graph grows, systems interested in metadata packages and records have a more diverse selection of descriptive information to utilize when building these structures [...] By 2006, Tim Berners-Lee had published a design note in which he reframed the Semantic Web discussion in much more useful terms by succinctly articulating both the simplicity and elegance of Linked Data (Harper).

How can library cataloging data be transformed in Linked Data within 'Web 3.0' and be understood by non-library web applications? To provide good answers to similar questions, different professionals from both library and semantic Web communities (e.g. W3C LLD XG Library Linked data incubator group[4]) are cooperating continuously exploring the situation in a technical and non-technical manner, making openly published their contributions about trans-

[4]http://www.w3.org/2005/Incubator/lld.

forming 'legacy library data' into Linked Data. The most important recent attempts of application of Linked Data in DL have been developed within the British National Library that provides its data as RDF download, in the German National Library that offers authority data linking to Wikipedia, DBpedia, and VIAF (Keßler); in the Hungarian National Library publishing its bibliographic and authority data using RDFDC for bibliographic data, Friend of a Friend (FOAF), and SKOS. To enhance semantic interoperability of bibliographical resources, DL community is looking to collaborate with an alpha project Linkypedia which is helpful enough in: exploring how the specific web content is used on Wikipedia (Linkypedia harvests all the links used in supporting reference in Wikipedia entries); extracting crowd-sourced information about certain web resources; connecting with Wikipedians who are using particular web bibliographical resources. Because many of links harvested from Wikipedia articles represent a lot of the citations point to Library/museum/archives information, Linkypedia's service can certainly be useful for innovative DL interoperability practices. Anyhow, Linked Data still miss resource aggregation mechanism based on univocal approach for publishing structured data (De Robbio and Giacomazzi). For this reason different descriptive models and metadata schemas have difficulty to be organized organically in a single space of Linked Data, even though different information communities are continuously promoting their new approaches for publication and use of shared data within Semantic Web. There is a need to individualize interoperable standards able to be integrated within LinkedOpenData (LOD); it is necessary to define the most proper formats and licenses for publishing metadata as LOD; it is important to get agreement on common minimum universally applicable unit of metadata for sharing information in the space of structured data ("minimal set of properties meaningful in data

sharing", LODe Recommendations).

LODe-BD data

Proceeding to overview the Linked Data application domains, it is also of great importance to cite a recent attempt to make bibliographic data LODe (Linked Open Data enabled) developed by AIMS team (Agricultural Information management Standards) (De Robbio and Giacomazzi). AIMS has developed and published on its site the Recommendations LODe-BD providing tools able to evaluate the choice of strategies and standards of metadata coding for the creation of bibliographical data LOD-enabled as well as Guidelines to code bibliographical data as LODe data in the space of Semantic Web.

The Recommendations provide a set of instructions to encode structures and properties of bibliographic (meta)data and are based on 5 fundamental principles:

1. to promote the use of well consolidated metadata standards and the emergent LOD-enabled vocabularies proposed by Linked Data community;

2. to encourage the use of Authority data, controlled vocabularies and the syntax of broadly shared coding metadata standards, thus improving the quality of semantic interoperability and effectiveness of information (structured data) interchange;

3. to encourage the use of URI as well as of literal and not-literal character strings to define most commonly values (seen as concepts/resources) of available bibliographic data. «When the thesaurus is published as Linked Data, the concept is considered as a resource and is given a unique URI, This means

that a URI reference is used to identify this concept as a resource» (LODE-BD Recommendations 1.1, 3. Explanation of Terminology);

4. to facilitate the decisional process [through the decisional trees] as it regards the individualization of the application of coding models and meta(data) profiles, so that they can be followed or reused by Data e Service Provider;

5. to provide a support to open references and suggestions regarding new properties and terms of metadata important to Linked Data community and producers of data.

The Recommendations inform how to publish and consume bibliographic Linked Data; where to retrieve vocabularies on their support; how to express metadata with different syntaxes: Text, HTML, XML, RDF, RDFa and why to publish bibliographical meta(data) as LOD. A conceptual model (Figure 4) introduced by the Recommendations provides a common sharable understanding on Entities and Relationships important for the creation of qualitive bibliographic relational data. This Model is developed on the base of FRBR (adjusted by AIMS group) allowing to extend and reconsider significantly the concept of bibliographic data LODe (Subirats, Nicolai, and Waltham).

The left part of LODe-BD conceptual model (see figure 3 on the facing page) provides a high-level abstract representation focusing on the Entity of bibliographic Resource. There can be identified principal relationships between an instance of a Resource, and of an Agent responsible for the content creation and dissemination of Resource, as well as between an instance of Resource and its Theme (Subject/Topic) that represents the content of Resource. Consequently, the model presents three basic entities: Resource, Agent, and Theme. In the right part of the Figure the implication of the gen-

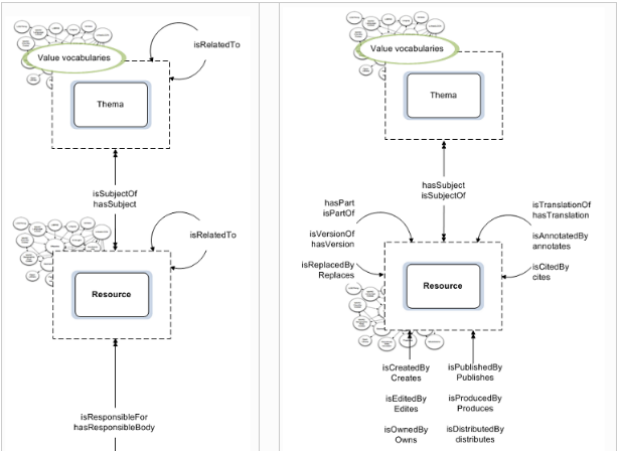

Figure 3: Conceptual LODe-BD model: definition of Subject, Theme, Agents and their connections: http://aims.fao.org/lode/bd/core-entities

eral conceptual model LODe-BD is shown and there are provided examples of possible relationships among the instances of different Entities:

1. The entity Resource is in the center of every description. The model doesn't exemplify the types of under-entities, which are different types of Resource. Resource is always the start point of each bibliographical description in LODe-BD decision trees.

2. The relationships are established among the entity Resource and other two important entities: Agent and Theme.

3. There can be also relationships among instances of an Entity: Resource can be connected to another Resource, as well as Agent to another Agent.

4. The relationship between any pair of instances may vary at

different levels. For example, an Agent can fund the creation of an original work, the translation of a work, or the production of a new format of a translation.

5. Authority Control is an important element of the model: Agents, Concepts, Titles, Themes regardless of their relation with a Resource should be managed through authority files of names.

6. Authority data is getting to be widely provided by LOD. The model intentionally represents an extracted piece of the LOD cloud inheriting the entities of authorities files.

SWAP (Scholarly Works Application Profile): levels of metadata and connections

SWAP is the document designated by DCMI in 2009 as a practical application of Dublin Core Application Profile in the domain of scientific works (peer-reviewed articles, pre-prints, working papers, thesis, parts of the book, scientific reports etc.) recognizing them as a distinct academic intellectual or artistic creations. The principal purpose of the SWAP is to offer a solution to the problems of metadata management and interoperability identified in "Eprints UK", a project financed by JISC (Joint Information Systems Committee) in 2007. The indications of SWAP allow to describe a vast range of digital academic works and are aimed at:

- providing a richer and more consistent metadata profile for the description of e-prints;
- facilitating the search, navigation and selection of the contents;
- allowing the identification of the most recent versions of resources and facilitating the navigation among different versions;

- sustaining services with added value, particularly those based on the use of OpenURL ContextObjects (ANSI/NISO Z39.88);
- implementing univocal/not ambiguous method to identify full-text of digital resources;
- allowing the identification of the research financier and the code of the project;
- facilitating the identification of OA materials.

The basic model used by SWAP is FRBR model represented by 4 Entities: 'Work', 'Expression', 'Manifestation', 'Item', and also the fifth Entity that is 'Agent' supporting the Entities 'Person' and 'Corporate Body'. The SWAP documentation includes both a schematic description of the domain model and the expressions of the diagram as follows. A ScholarlyWork may be expressed as one or more Expressions. Each Expression may be manifested as one or more Manifestations. Each Manifestation may be available as one or more Copies. Each ScholarlyWork may have one or more Creators, Funders and Supervisors.Each Expression may have one or more Editors and each Manifestation - one or more Publishers. Hereinafter we provide the schematic representation of metadata describing Entities of scientific works within SWAP application Profile (figure 4 on the next page).

Dublin Core Abstract Model (DCAM) is important for the choice of descriptive terms for metadata, as well as of correlated descriptions allowing to application models to capture, codify and share a set of more complex entities. Making use of information offered by the DCAM, the SWAP profile is able to capture and to group the descriptions of multiple entities in a single set of description. SWAP takes in consideration functional requirements for descriptions of scientific works as well as user's needs inherent in the richest and the

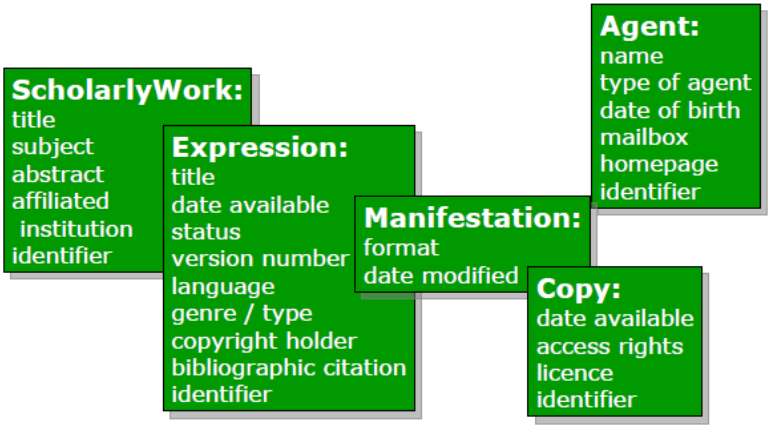

Figure 4: Examples of metadata/properties of qualification for SWAP Entities

most functional metadata. SWAP has rationalized the 'traditional' citations among Expressions and hyperlinks among Copies, as well as the navigation through different Versions and the individualization of OA full-text copies. The apparently complex SWAP model can be manifested in the relative simple navigation interfaces. The community supporting SWAP has developed Community Acceptance Plan in which has declared all the assumptions for common development and acceptance of the application profile to improve description and interoperability of the digital scientific works managed by OpenSource Software such as Eprints, Dspace and Fedora involved in the implementation of Open Access Repositories.

The strategic role of information management in a DL project, – equally if it is oriented to the production of electronic resources, to the digitization of analog documents, to retrieval of and access to

already available digital resources – should operate on the prudent choice of material to be treated, on discovery of information needs, and therefore of metadata to be implemented to support all necessary functions of acquisition, production, access and use of digital objects in long-term perspective. Metadata should be seen as an important attribute of any element of information that is in electronic format or is required to be catalogued electronically.Opportunities to share the digital bibliographic information have been increased with event of Open and Linked data. In this context, the SWAP model is aimed to help cataloguers to create organic and highly interoperable bibliographic data relating to a broad population of metadata schemas supporting the description of scientific works, thus also promoting the strategic role of FRBR and DCAM in a larger context of digital information communities.

Works Cited

Baker, Thomas. *Basics of Dublin Core Metadata*. Seoul, 2009. http://dublincore.org/resources/training/dc-2009/Tbaker.pdf. (Cit. on pp. 6, 12). Web.

——. "Designing Interoperable Metadata on Linked Data Principles". *Tutorial: Dublin Core -Building blocks for interoperability*. 2009. http://dublincore.org/resources/training/NISO_Webinar_20100825/dcmi-webinar-03.pdf. (Cit. on p. 8). Web.

——. *Designing Interoperable Metadata on Linked Data Principles. Tutorial: Dublin Core -Building blocks for interoperability*. 2009. http://dublincore.org/resources/training/NISO_Webinar_20100825/dcmi-webinar-03.pdf. (Cit. on p. 4). Web.

Baker, Thomas, et al. "What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives". *Journal of Digital Information* 2.2. (2001). Web.

Bearman, David, et al. "A common model to support interoperable metadata: progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities". *D-Lib magazine* 5.1. (1999). Web.

Bermudez, L. and M. Piasecki. "Metadata Community Profiles for the Semantic Web". *GeoInformatica* 10.2. DOI: 10.1007/s10707-006-7577-2. (2006): 159–176. (Cit. on p. 12). Web.

Birrell, Duncan, Gordon Dunsire, and Kathleen Menzies. *Web Catalogue and Repository Interoperability Study (OCRIS). Final Report, Centre for Digital Library Research, University of Strathclyde*. 2009. http://eprints.rclis.org/bitstream/10760/13187/1/strathprints014037.pdf. (Cit. on pp. 5, 7). Web.

Bradley, P. Allen. *Faceted Navigation of User-Generated Metadata, Metadata Tools for Digital Resource Repositories*. 2006. http://www.ils.unc.edu/mrc/jcdl2006/abstracts/allen.html. (Cit. on p. 11). Web.

Broughton, Vanda. "Concepts and terms in the faceted classification: the case of UDC. Paper presented at UDC International Seminar: Classification at the crossroads, The Hague, 29-30 October, 2009". (Cit. on p. 11). Print.

Campbell, Grant D. "Metadata, Metaphor, and Metonymy". *Metadata: A Cataloger's Primer*. Ed. Richard Smiraglia. New York: Haworth Information, 2006. (Cit. on pp. 4, 5). Print.

Caplan, Priscilla. *Metadata fundamentals for all librarians*. Chicago: American Library Association, 2003. Print.

Chan, Lois M and Zeng M. Lei. "Metadata Interoperability and Standardization – A Study of Methodology. Achieving Interoperability at the Schema Level". *D-Lib magazine* 12.6. (2006). http://www.dlib.org/dlib/june06/chan/06chan.html. Web.

Chopey, Michael A. "Planning and Implementing a Metadata-Driven Digital Repository". *Metadata: A Cataloger's Primer*. Ed. Richard Smiraglia. New York: Haworth Information, 2006. (Cit. on p. 9). Print.

Christopher, Lee. "A framework for contextual information in digital collections". *Journal of Documentation* 67.1. (2011). (Cit. on p. 4). Print.

Cole, Timothy W. and Muriel Foulonneau. *Using the Open Archive Initiative Protocol for Metadata Harvesting*. London: Westport Connecticut, 2007. (Cit. on p. 8). Print.

DCMI. *DCMI Specifications*. 2011. http://dublincore.org/specifications. Web.

——. *Dublin Core Metadata Element Set, version 1.0: Reference description*. 1998. http://www.dublincore.org/documents/1998/09/dces. Web.

De Sompel, Herbert Van, Thomas Krichel, and Michael L. Nelson. "The UPS Prototype. An Experimental End-User Service across E-Print Archives". *D-Lib Magazine* 6.2. (2000). Web.

De Robbio, Antonella. *Archivi aperti e comunicazione scientifica*. Napoli: ClioPress, 2007. http://www.storia.unina.it/cliopress/derobbio.htm. (Cit. on p. 7). Web.

De Robbio, Antonella and Silvia Giacomazzi. "Dati aperti con LODe". *Bibliotime* 14.2. (2011). http://didattica.spbo.unibo.it/bibliotime/num-xiv-2/derobbio.htm. (Cit. on pp. 14, 15). Web.

Delteil, Alexandre, Catherine Faron-Zucker, and Rose Dieng. "Learning ontologies from RDF annotations". *Proceedings of IJCAI-01 Workshop on Ontology Learning OL-2001*. Ed. Alexander Maedche, et al. Seattle, 2001. (Cit. on p. 10). Print.

Dunsire, Gordon. "Collecting metadata from institutional repositories". *OCLC Systems & Services* 24.1. (2008). Print.

Dunsire, Gordon and Wirna Miller. "Initiatives to make standard library metadata models and structures available to the Semantic Web. Meeting 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management". Proc. Of World Library and Information Congress: 76th IFLA General Conference And Assembly, Gothenburg. 2010. http://www.ifla.org/files/hq/papers/ifla76/149-dunsire-en.pdf. Web.

Duval, Erik, et al. "Metadata Principles and Practicalities". *D-Lib Magazine* 8.4. (2002). http://www.dlib.org/dlib/april02/weibel/04weibel.html. Web.

El-Sherbini, Magda. "Metadata and the future of cataloging". *Library Review* 50.1. (2001): 16–27. Print.

El-Sherbini, Magda and George Klim. "Metadata and cataloging practices". *Electronic Library* 22.3. (2004): 238–248. (Cit. on p. 5). Print.

Encoded Archival Description. 2002. http://www.loc.gov/ead. Web.

Functional Requirements of Authority Numbering and Records. 2010. http://www.ifla.org/en/taxonomy/term/474. Web.

Functional Requirements: the FRBR Family of Models. IFLA official website. 2011. http://www.ifla.org/node/2016. Web.

Gambarara, Daniele. "Thesauri, mappe semantiche, ontologie. Problemi semantici e costruzioni concettuali". *Dal documento all'informazione*. Ed. Roberto Guarasci. Milano: ITER, 2008. (Cit. on p. 11). Print.

Gartner, R. *Metadata for Digital Libraries: state of the art and future directions*. Bristol: JISC, 2008. (Cit. on p. 5). Print.

Gemberling, T. "Thema and FRBR's Third Group". *Cataloging & Classification Quarterly* 48.5. DOI: 10.1080/01639371003745413. (2001): 445–449. (Cit. on p. 2). Print.

Ghilli, C. and M. Guerrini. *Introduzione a FRBR. Functional Requirements for Bibliographic Records - Requisiti funzionali per record bibliografici*. Milano: Editrice Bibliografica, 2001. (Cit. on p. 2). Print.

Gill, Tony. "Metadata and the Web". *Introduction to Metadata*. Ed. Murtha Baca. Los Angeles: Getty Research Institute, 2008. http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html. Web.

Gilliland, Anne J. "Metadata - Where Are We Going?" *International Yearbook of Library and Information Management: Metadata Applications and Management*. Ed. Gary E. Gorman. London: Facet Publishing, 2003. Print.

——. "Setting the Stage". *Introduction to Metadata*. Ed. Murtha Baca. Los Angeles: Getty Research Institute, 2008. http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html. Web.

Gorman, Michael. "Electronic Resources: which are worth preserving & what is their role in library collections?" Proc. Of International Conference Electronic

Resources: Definition, Selection and Cataloguing, Rome. 2001. http://eprints.rclis.org/bitstream/10760/4141/1/gorman_eng.pdf. Web.

——. "Metadata or cataloguing? A false choices". *Journal of Internet Cataloging* 2.1. (1999). http://www.informaworld.com/smpp/content~db=all~content=a909291879~frm=abslin. (Cit. on p. 5). Web.

Gradmann, Stefan. "Cataloguing vs. metadata: old wine in new bottles?" *International cataloguing and bibliographic control* 28.4. (1999). Print.

Graham, Rebecca A. "Metadata harvesting". *Library Hi Tech* 19.3. (2001). Print.

Greenberg, Jane. "Metadata Generation: Processes, People and Tools". *Bulletin of the American Society for Information Science and Technology* 29.2. (2003). http://www.asis.org/Bulletin/Dec-02/greenberg.html. Web.

——. "Understanding Metadata and Metadata Schemas". *Metadata: A Cataloger's Primer*. Ed. Richard Smiraglia. New York: The Haworth Information, 2006. http://www.ils.unc.edu/mrc/pdf/greenberg05understanding.pdf. (Cit. on pp. 6, 8). Web.

Greenstein, D. and S.E. Thorin. *The Digital Library: A Biography*. Washington: Digital Library Federation Council on Library and Information Resources, 2005. http://www.clir.org/PUBS/reports/pub109/pub109.pdf. (Cit. on pp. 1, 5). Web.

Groenewald, Ria and Amelia Breytenbach. "The use of metadata and preservation methods for continuous access to digital data". *Electronic Library* 29.2. (2011). Print.

Guerrini, Mauro. *Gli archivi istituzionali. Open access, valutazione della ricerca e diritto d'autore*. Milano: Editrice Bibliografica, 2010. http://eprints.rclis.org/bitstream/10760/15609/1/Gli_archivi_istituzionali.pdf. (Cit. on p. 7). Web.

Harper, Corey. "Dublin Core metadata initiative: Beyond the Element Set". *Information Standards Quarterly* 22.1. (2010). http://www.niso.org/publications/isq/free/FE_DCMI_Harper_isqv22no1.pdf. (Cit. on p. 13). Web.

Hillmann, Diane. *Using Dublin Core*. 2005. http://dublincore.org/documents/usageguide. (Cit. on p. 7). Web.

Hillmann, Diane, Naomi Dushay, and Jon Phipps. "Improving metadata quality: augmentation and recombination". *Proceedings of the 2004 international conference on Dublin Core and metadata applications: metadata across languages and cultures*. Shanghai, China: Dublin Core Metadata Initiative, 2004. http://dl.acm.org/citation.cfm?id=1383385.1383395. (Cit. on p. 5). Web.

Hodge, Gail. "Metadata for electronic information resources: From variety to interoperability". *Information Services and Use* 25.1. (2005). http://dl.acm.org/citation.cfm?id=1383385.1383395. Web.

Howarth, Lynne C. "Metadata and bibliographic control: Soul-mates or two solitudes?" *Cataloging & Classification Quarterly* 40.3-4. (2005). (Cit. on p. 9). Print.

IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: Final report*. München: Saur, 1998. http://www.ifla.org/publications/functional-requirements-for-bibliographic-records. (Cit. on p. 2). Print.

Information Technology- Metadata registries (MDR). 2004. http://metadata-stds.org/11179. Web.

Information Technology Identification and Mapping of Various Categories of Jurisdictional Domains. 2002. http://metadata-stds.org/18038/index.html. Web.

Information technology Metadata Registry Interoperability and Bindings (MDRIB). 2008. http://metadata-stds.org/20944/index.html. Web.

Information Technology Procedures for Achieving Metadata Registry Content Consistency. 2003. http://metadata-stds.org/20943/index.html. Web.

International Standard Bibliographic Description for Electronic Resources. 1997. http://archive.ifla.org/VII/s13/pubs/isbd.htm. Web.

Jones, Wayne. "Cataloging the Web: Metadata, AACR, and MARC 21". *ALCTS Paper Series on Library Technical Services and Collection* 10. (2002). Print.

Jorgensen, Poul H. "Practical application of FRBR and RDF". Proc. Of International Conference Electronic Resources: Definition, Selection and Cataloguing, Rome. 2001. http://eprints.rclis.org/bitstream/10760/4114/1/jorgensen_eng.pdf. Web.

Keßler, M. *Linked Open Data of the German National Library. ECO4r Workshop "LOD of DNB"*. Berlin: Deutsche National Bibliotek, 2010. http://www.eco4r.org/workshop2010/eco4r_workshop2010_mirjam_kessler.pdf. (Cit. on p. 14). Web.

Krichel, Thomas. *ReDIF v.1*. 2000. http://openlib.org/acmes/root/docu/redif_1.html. Web.

Kurth, Martin, David Ruddy, and Nathan Rupp. "Repurposing MARC metadata: using digital project experience to develop a metadata management design". *Library Hi Tech* 22.2. (2004). (Cit. on p. 5). Print.

Lagoze, Carl. "The Warwick Framework: A Container Architecture for Diverse Sets of Metadata". *D-Lib Magazine*. (1996). Print.

Matei, Dan and Peter Noerr. "User benefits for a new bibliographic model: follow-up the IFLA functional requirements study". Proc. Of 64th IFLA General Conference, Amsterdam. 1998. http://www.ifla.org/IV/ifla64/084-126e.htm. Web.

Metadata Encoding and Transmission Standard. 2011. http://www.loc.gov/standards/mets. Web.

Myers, Jack. *Metadata Ontology Language*. 2008. http://www.metadata.com/products/metadata_language.pdf. Web.

Nilsson, Mikael. *From Interoperability to Harmonization in Metadata. Standardization Designing an Evolvable Framework for Metadata Harmonization*. 2010. http://kmr.

nada.kth.se/papers/SemanticWeb/FromInteropToHarm-MikaelsThesis.pdf. (Cit. on p. 10). Web.

Palma, Raul, Jens Hartmann, and A. Gomez-Perez. *Towards an Ontology Metadata Standard*. 2007. http://oa.upm.es/6259/1/Towards_an_Ontology_Metadata_Standard.pdf. (Cit. on p. 10). Web.

Powell, A., et al. *DCMI Abstract Model*. 2007. http://dublincore.org/documents/abstract-model. Web.

Shotton, David and Silvio Peroni. *FaBiO, the FRBR-aligned Bibliographic Ontology*. 2011. http://speroni.web.cs.unibo.it/cgi-bin/lode/req.py?req=http:/purl.org/spar/fabio. (Cit. on p. 2). Web.

Smiraglia, Richard. "Introducing Metadata". *Metadata: A Cataloger's Primer*. Ed. Richard Smiraglia. New York: Haworth Information, 2006. (Cit. on p. 4). Print.

Subirats, I., C. Nicolai, and N. Waltham. *FRBR/FAO Model Implementation and Benefits at ASIST Annual Meeting 2010: Information for Submission*. 2010. http://www.slideshare.net/faoaims/frbrfao-model-implementation-and-benefits. (Cit. on p. 16). Web.

Text Encoding Initiative. 2011. http://www.tei-c.org/index.xml. Web.

Tillett, Barbara. "AACR2 and metadata: Library opportunities in the global semantic web". *Cataloguing & classification quarterly* 36.3-4. (2003). Print.

——. *FRBR: A Conceptual Model for the Bibliographic Universe*. 2004. http://www.loc.gov/cds/downloads/FRBR.pdf. (Cit. on p. 2). Web.

UNIMARC to MARC 21 Conversion Specifications Prepared for the Network Development and MARC Standards Office Library of Congress. 2001. http://www.loc.gov/marc/unimarctomarc21.html. Web.

Universal Bibliographic Control and International MARC Core Activity. 2009. http://archive.ifla.org/VI/3/nd1/publist.htm. Web.

Valkeapää, Onni, Olli Alm, and Eero Hyvönen. "Efficient Content Creation on the Semantic Web Using Metadata Schemas with Domain Ontology Services (System Description)". *Proceedings of the 4th European conference on The Semantic Web: Research and Applications*. Heidelberg: Springer, 2007. (Cit. on pp. 6, 11). Print.

VRA, v.3.0. Visual Recourse Association's Core Categories. 2000. http://www.vraweb.org/projects/vracore3/index.html. Web.

Weibel, Stuart. *Approval of initial Dublin Core interoperability qualifiers*. 2002. http://www.mailbase.ac.uk/lists/dc-general/2000-04/0010.html. Web.

Weibel, Stuart and Traugott Koch. "The Dublin Core Metadata Initiative: mission, current activities, and future directions". *D-Lib Magazine* 6.12. (2000). http://www.dlib.org/dlib/december00/weibel/12weibel.html. Web.

Weinstein, Peter C. "Ontology-Based Metadata:Transforming the MARC Legacy". *Proceedings of the Third ACM Digital Library conference*. Pittsburg, 1998. http://www-personal.umich.edu/~peterw/Ontology/beethoven.paper.rtfg. Web.

Woodley, Mary. *DCMI Glossary. Dublin Core Metadata Initiative*. 2003. http://dublincore.org/documents/usageguide/glossary.shtml#admin. Web.

Zhuhadar, Leyla, et al. "Metadata as seeds for building an ontology driven information retrieval systems". *International Journal of Hybrid Intelligent Systems* 6.3. (2009). http://iospress.metapress.com/content/mt01x14600l52p77/?p=47b8253923e14c09b9954dfbe0c8f069&pi=8. Web.

IRYNA SOLODOVNIK, University of Calabria
email iryna.solodovnik@unical.it

Solodovnik, I."Metadata issues in Digital Libraries: key concepts and perspectives". *JLIS.it* Vol. 2, n. 2 (Dicembre/December 2011): 4663-1–4663-27. DOI: 10.4403/jlis.it-4663. Web.

ABSTRACT: The article sets out to investigate the meaning, role and implications of some information management approaches used in Digital Library practice. A greater focus on innovation in managing online resources and on improving their interoperability can be achieved by normalizing metadata schemas through interoperable standards, world-wide accepted controlled vocabularies as well as by their enrichment through qualitatively constructed ontologies and linked data, which are key to the expansion of the semantic reasoning on the web through building and connection of additional semantic layers on top of metadata descriptions. Reviewing some innovative methods of information representation (LODe-BD, SWAP), the paper tries to lead the reader to discover some new ways of knowledge creation in digital information environment, in particular what concerns digital bibliographic records.

KEYWORDS: Digital library; FRBR; Linked data; LODe-BD; Metadata; SWAP.

Submitted: 2011-06-02 / Accepted: 2011-09-22 / Published: 2011-06-15.