Implementation of Metadata for OmniPaper RDF Prototype

Tomoko Yaginuma, Teresa Susana Mendes Pereira, César E. Ariza Ávila, Ana Alice Baptista
Department of Information Systems, School of Engineering, University of Minho, Portugal
{tomoko, tpereira, cariza, analice}@dsi.uminho.pt

Abstract

Information Society Technologies (IST) funded OmniPaper project investigates efficient ways for access to distributed and heterogeneous digital news archives using state-of-the-art technologies such as RDF, XTM and SOAP. An approach taken is to create small prototypes based on each of them. This paper presents the first stage of the prototype development, particularly of RDF approach, including analysis on existing news text format standards and metadata vocabularies, definition of metadata elements for OmniPaper, implementation of application profile and RDF schema and development of the RDF prototype in a web-based RDF specific application. The elaborated analysis shows that Dublin Core Metadata Element Set has to be a principal vocabulary to implement the OmniPaper application profile as it provides greater interoperability. The RDF prototype provides RDF "metadatabase" with searchable interface for simple and advance search on the defined metadata elements.
Keywords: Metadata, Dublin Core, Application Profile, Schema, RDF, OmniPaper.

1. Introduction

OmniPaper (Smart Access to European Newspapers, IST-2001-32174) [1] is a project from the European Commission IST program (Information Society Technologies), that investigates and proposes ways for access to different types of distributed information sources.

We intend to achieve the implementation of a prototype that enables users (professional journalists and occasional users) to have simultaneous and structured access to the articles of a large number of digital European news providers.

One of the principal aspects of the project is the whole metadata layer of the system. Conceptually, there are two metadata layers: (1) the first layer (Local Knowledge Layer) that is added to the local archives where the main purpose is to provide a standard semantic description of all the existent articles in order to enable a structured and uniform access to the available distributed archives; and (2) a second layer (Overall Knowledge Layer), an higher abstraction level that will provide a common access user interface by integrating and relating the metadata information coming from the local knowledge layer. Furthermore, this interface is meant to be personalized and multilingual. Figure 1 shows the conceptual view of OmniPaper architecture.

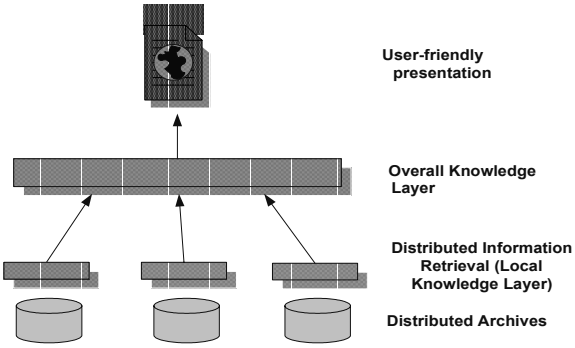

Figure 1. Concept of OmniPaper Architecture

In the OmniPaper project we are developing simultaneously two different metadata approaches to both layers: the Resource Description Framework (RDF) [2] approach, and the Topic Maps (TM) [3] approach.

The implemented work related to the RDF approach in the OmniPaper project has already taken the following steps:

(1) Elaborated analysis of several XML applications and vocabularies, such as XML News Industry Text Format (NITF) [4], NewsML (News Agency Implementation Guidelines) [5] developed by International Press Telecommunications Council (IPTC) [6], XMLNews [7], RDF Site Summary (RSS) 1.0 [8], Dublin Core Metadata Element Set (DCMES) [9], and Dublin Core Qualifiers (DCQ) [10];

(2) Selection or filter of the metadata elements that best define the features of the newspaper articles. These elements will be included in the application profile;

(3) Definition of other elements, that are necessary, and are not found in any widely used vocabulary. These will also be included in the application profile;

(4) Definition of the vocabulary mentioned in (3) in an RDF schema;

(5) Definition of the rules for the RDF/XML documents' description;

(6) Template of the documents' description in RDF/XML;

(7) Development of "Metadatabase" prototype using a native RDF Database Management System (DBMS) - RDF Gateway [11];

(8) Transformation of Article XML files and keyword XML files into RDF/XML files and

(9) Implementation of full prototype.

This paper focuses on the RDF approach to the Local knowledge Layer, and deepens each step shown above.

2. Analysis of the Existing News Text Format Standards and Metadata Vocabularies

For development of application profile for OmniPaper, major news text formats and metadata vocabularies were thoroughly studied and analysed. Those examined formats and vocabularies are as follows.

- NITF
- NewsML
- XMLNews
- RSS
- DCMES and DCQ

In this section, each standard will be briefly introduced, and then advantages and disadvantages for OmniPaper application use will be discussed. Lastly, a decision for the selection will be made.

2.1 NITF

NITF is a text format standard which defines the content and structure of news articles. It was first created based on SGML and then after 1998 up-dated to be compliant with XML. NITF supports the identification and description of a number of news characteristics with rich in-line markups. It identifies, for instance, structural pieces of news article such as headings, bylines, paragraphs, tables, columns and footnotes. Metadata tags are also applied for describing the news content.

2.2 NewsML

NewsML is a one level higher standard for news representation and management. It is also based on XML. The main function of NewsML is the management of news item throughout its lifelong, including production, interchange and consumer use. It can contain many different types of objects including multimedia items such as images, videos, etc. The text written in NITF can be accommodated within NewsML in self-contained manner. Likewise, other formats like SportML and JPEG can be wrapped by NewsML. While NITF is a standardised format for description and structuring of news text, NewsML is used as an information wrapper and management tool.

2.3 XMLNews

XMLNews is a pair of specifications (XMLNews-Story and XMLNews-Meta) maintained by Megginson Technologies Ltd. XMLNews-Story defines the content of textual news stories while XMLNews-Meta defines a set of metadata information about news objects. Both specifications are developed based on XML in 1999. XMLNews-Story is a fully-compatible subset of the NITF. XMLNews-Meta is a news industry metadata format conforming to RDF. It is used as a single, common format to provide metadata about any kind of news object, whether textual or non-textual, separated from the news objects themselves.

2.4 RSS

RSS is an XML-based multipurpose extensible metadata description and syndication format. The latest version 1.0 is fully conforming to RDF. A typical RSS document describes a "channel" consisting of URL-retrievable items. Each item consists of a title, link and brief description. While items have traditionally been news headlines, RSS has been evolved to be used for multiple purposes such as aggregation, discussion threads, job listings, document cataloging, etc.

2.5 DCMES and DCQ

DCMES, which is now in line with the ISO standards, is a well-known metadata standard for cross-domain information resource description, developed and maintained by Dublin Core Metadata Initiatives (DCMI) [12]. DCMES is relatively simple 15 metadata elements set, easy to understand and extend to richer semantic description standards. DCQ was developed to qualify DCMES in ways that these qualifiers either make the meaning of an element narrower or more specific, or identify schemes that aid in the interpretation of an element value. DCQ has been incorporated with the DCMI type vocabulary and other legacy documents related to element refinements, and superseded by the DCMI Metadata Terms document [13].

2.6 Analysis

A close examination shows different aims of those existing standards: One is for mark-up of news article and the other is for metadata description of news article or any kind of resources. NITF, NewsML and XMLNews-Story belong to the former type of purpose while RSS, XMLNews-Meta, DCMES and DCQ belong to the latter.

A document written according to the former standards contains a full text (content) of a news article. Although NITF and NewsML-based documents can also include metadata elements, the main purpose of the standards is to provide a well defined and structured in-line mark-up for news text itself. This makes it easy to handle appearance and transformation of news article throughout its lifelong,

that seems to be one of the main advantages. The other advantages are explained in details by Yaginuma, et al [14].

On the other hand, the documents based on RSS, XMLNews-Meta or DCMES include only metadata description about a particular resource. Obviously, the metadata document is independent and separated from the original resource. This way, metadata extracted or transformed from an original resource is easily handled for further development. The size of document is much less than the one with news content, which is also preferable.

For OmniPaper application, as explained in the previous section, only metadata of news articles are important in the development of Local Knowledge Layer prototypes. Therefore, the formats such as RSS and XMLNews-Meta or the vocabulary like DCMES may be preferable for the selection of metadata elements for OmniPaper. Nevertheless, elements from NITF and XMLNews are worth analysing for the possible use in RDF.

Further analysis indicated that RSS, as its name "Site Summary" implies, is more suitable for describing news "site" rather than details of each news article. Probably, it might be better used for syndication of completed OmniPaper site or personalisation of user interface.

In terms of XMLNews-Meta, it appears to be suitable for the OmniPaper purpose in a way that it can provide metadata separately from news content. Conformance to RDF seems to be beneficial for implementation of the RDF prototype. Specification introduces more than 40 metadata elements, which is rich enough to describe news objects.

However, our first priority for metadata description was given to DCMES and DCQ because of their simplicity and interoperability. They are rather general and abstract standards but have been widely used across the boundaries of disciplines or application domains. This means news articles described with DCMES have higher possibility to be discovered by external applications in different area, not only in news industry, which is very attractive to achieve Tim Berners Lee's Semantic Web concept [15]. Main reason why we did not make a full use of XMLNews-Meta results largely in that point.

To sum up, DCMES and DCQ are taken as OmniPaper's principal metadata vocabularies. Furthermore, when no appropriate element was found in DCMES or DCQ, we would refer to as many standards commonly used as possible. The last option has to be a creation of OmniPaper-original metadata elements. This must be avoided unless there is strong necessity.

2.7 Related Work

Along with the study of existing standards, we have examined the metadata elements that our news providing partners currently use for their systems. It is highly important to know what metadata elements are in use in the real world. Close examination showed several characteristics of the elements for describing news articles.

(1) **Article identification**: There are some common and essential metadata that all the companies use. They refer to the title of an article, author, issue date and the unique ID for an article.

(2) **Article classification and linking**: There are some kind of classification elements which sort the articles into categories, or reference elements which link the articles related to each other or in the same series. Each company has its own way of classifying and linking articles, using an existing or original thesaurus or other criteria.

(3) **Other media**: One company has several elements that describe the photos used for an article in details. Another company has an element which only describes the caption of the photo. These elements brought out the issue of how to deal with other media included in the news articles.

(4) **Edition and section**: Some elements are highly based on the paper version of news articles. For instance, they describe from which edition(s) the article came from, or in which section of the newspaper the article belongs to.

These factors were greatly useful for the selection of metadata elements. According to the analysis described in this section, our initial metadata elements were chosen.

2.8 Namespaces

Figure 2 summarises metadata sources for OmniPaper.

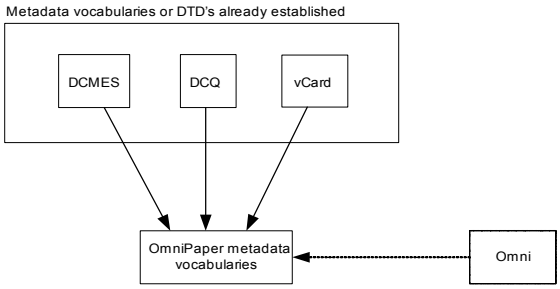

Figure 2. Metadata Sources for OmniPaper

The namespaces for these four sources are listed below:

- Dublin Core Metadata Element Set (DCMES) - http://purl.org/dc/elements/1.1/
- Dublin Core Qualifiers (DCQ) - http://purl.org/dc/terms
- OmniPaper RDF Schema (omni) - http://www.dsi.uminho.pt/omn/schemas/omn-schema#
- vCard - http://www.w3.org/2001/vcard-rdf/3.0#

The following section introduces the final set of metadata elements used for OmniPaper.

3. Metadata Elements Used in OmniPaper

We have defined 27 elements for OmniPaper. The tables below show all the elements sorted into 6 categories:

- Article Identification;
- Article Ownership;
- Article Location (Storage);
- Article Relevance/Audience;
- Article Classification;
- Link Information.

Each element is described with name, definition of the element, the source of the element and recommended encoding scheme. The encoding schemes are the qualifiers used for expressing the element value. These schemes include controlled vocabulary and formal notation. For the metadata elements taken from DCMES and DCQ, we followed the encoding schemes recommended by DCMI.

As can be seen in the following tables, 18 elements are taken from Dublin Core, either DCMES or DCQ, and the rest 9 elements are OmniPaper-original. Majority of elements are taken from Dublin Core, which is ideal for maximising interoperability.

Table 1. Article Identification

Source	Element	Definition	Encoding Scheme(s)
DCMES	Identifier	An unambiguous reference to a resource.	URI
OMNI	UniqueId	An OmniPaper specific ID for a resource.	
DCMES	Creator	Author(s) of an article.	
DCQ	Issued	Date of publication of an article.	W3C-DTF
DCMES	Title	Title of an article.	
DCQ	Alternative	Substitute or alternative to the formal title of article.	
DCMES	Publisher	The entity responsible for making the resource available.	
DCMES	Language	The language in which an article is written.	ISO 1766 & 639
OMNI	KindOfArticle	Nature or genre of an article.	
OMNI	Section	Named section of a publication where an article appears.	
OMNI	Edition	The name(s) of edition(s) in which an article is distributed.	
DCMES	Contributor	An entity responsible for making contributions to the content of the resource.	
DCQ	Extent	Size of an article.	

Table 2. Article Ownership

Source	Element	Description	Encoding Scheme(s)
OMNI	Copyright	Container for copyright information.	
OMNI	Supplier	The local archive that owns article.	

Table 3. Article Location (Storage)

Source	Element	Description	Encoding Scheme(s)
DCQ	Medium	The physical or digital carrier of an article.	IMT
DCMES	Source	A reference to an article from which the present article is derived.	URI

Table 4. Article Relevance/Audience

Source	Element	Description	Encoding Scheme(s)
OMNI	OfInterestTo	The target audience for an article, based on demographic, geographic or other groups.	
DCQ	Valid	Date (often a range) of validity of an article.	W3C-DTF
DCQ	Spatial	Geographical location that an article treats or is related to.	ISO 3166

Table 5. Article Classification

Source	Element	Description	Encoding Scheme(s)
DCQ	Abstract	A summary of the content of an article.	
OMNI	Key_list	A list of most relevant keywords extracted from an article document.	
DCMES	Subject	Topic of the content of an article, specified according to the IPTC Subject Code.	IPTC Subject Code System

Table 6. Link Information

Source	Element	Description	Encoding Scheme(s)
DCQ	HasPart	The described article includes the referenced resource physically, such as photo, table, diagram, etc.	URI
DCQ	IsVersionOf	The described article is a version, edition or adaptation of the referenced version.	URI

OMNI	Series	The name of series that an article belongs to.	
DCQ	References	The described article references, cites or points to the referenced article.	URI

All the elements are repeatable and optional, though, some basic elements such as Title, Issued and so on should preferably be provided. Additional explanation is given to the elements below.

UniqueID: UniqueID is an OmniPaper-original element. The purpose of this element is to identify a certain article with OmniPaper specific string-based ID for the process within the OmniPaper system. The main difference from the Identifier element is that UniqueId should not be public to the outside of the system.

KindOfArticle: The purpose of the element is to classify the form of the article as genre. This element is an OmniPaper original so that the vocabulary for the values of the property such as interview has to be also defined (see Figure 4).

Supplier: This element has to describe the organisation which supplies news articles to the OmniPaper system. It can be called "Local archive" as conventionally used in the OmniPaper architecture and consortium. It has to be clearly distinguished from the author of the article, news companies or distributors which provide news to the "local archives". NITF Distributor element was one of the candidates but we took the original "Supplier" to emphasise the specific meaning.

Spatial: Previously, there were several more elements which refer to the semantic content of the document, namely, Doc_scope (NITF), Location (NITF), Company and Organisation. Doc-scope indicates an area where the article may be of interest. Location defines a relationship between an article and a geographical location about which the article treats. Company and Organisation define the relationship between an article and a company/organisation about which the article treats. The first two are very similar so that we had to have only one element which describes the geographical space. Then, the Spatial element, the qualifier of Coverage (DCMES) was chosen instead of NITF based Doc_space and Location. The latter two elements are rather too specific and we decided to discard them.

It is also worth noting that some of the OmniPaper-original elements are inspired in other formats (NITF or NewsML). As these formats are not metadata vocabularies but DTD's or XML Schemas and have not yet got the respective schemas or namespace schemas, they have been included in our schema always with reference to the original context in which they were developed and are widely used. The elements referenced to NITF are **Section**, **Edition**, **Copyright**, **Key-list** and **Series**. The element referenced to NewsML is **OfInterestTo**. In the next section, we will discuss about application profile and RDF Schema.

4. OmniPaper Application Profile and RDF Schema

4.1. OmniPaper Application Profile

The concept of application profiles has emerged in discussions on metadata schemas, in relation to work that is being done on metadata registries, specifically in DCMI. The application profiles grew out of UKOLN's work [16] on the DESIRE project [17].

Heery and Patel define application profiles as schemas which consist of data elements drawn from one or more namespace schemas, combined together by implementers and optimised for a particular local application [18]. They distinguish namespace schema from application profile schema.

Namespace is defined within the W3C XML schema activity [19] and allows for unique identification of elements, that may be defined through an RDF Schema. Within the W3C XML and RDF Schema Specifications [20], namespaces are the domain names associated with elements which, along with the individual element name, produce a URL that uniquely identifies the element [18].

According to Heery and Patel [18], with 'namespace', it is possible to:

- identify the management authority for an element set;
- support definition of unique identifiers for elements;
- uniquely define particular data element sets or vocabularies.

In application profiles, it is possible to [18]:

- use elements of one or more namespaces;
- not introduce new elements;
- specify schemas and allowed values;
- refine normalized definitions.

In the context of the OmniPaper project, the definition of the application profile is intended to describe not only the elements from different metadata vocabularies but also to describe the elements defined in our own schema implemented according to the implementation of the RDF/XML metadata structure that is fully described by Yaginuma et al. [21].

After an elaborated analysis on the elements for OmniPaper, described in the previous section, an RDF/XML application profile [22] based on vocabulary SMES [23] was developed in the scope of OmniPaper project. The complete OmniPaper application profile can be seen in CORES registry [24, 25].

4.2. OmniPaper RDF Schema

The Resource Description Framework Schema Specification (RDFSS) [20] is a proposed recommendation of the World Wide Web Consortium (W3C) [26]. The schema specification language is a declarative representation language influenced by ideas from knowledge representation (e.g. semantic nets, frames, predicate logic) as well as database schema specification languages and graph data models [20].

The main goal of the RDFSS consists of the definition of a schema specification language that allows the definition of mechanisms needed to define elements from specific vocabularies, to define classes of resources which those elements may be used with, according to restriction mechanisms, based on the constraint rdfs:domain, and in the definition of values for the properties through the rdfs:range.

The RDFSS provides mechanisms of extensibility of RDF vocabularies, through the use of classes and specific properties of the RDF Schema. We can define our own vocabularies that are not more than specializations of vocabularies normalised as it is the case of DCMES.

In the OmniPaper project, some metadata elements that were not previously defined neither in standard vocabularies nor in widely-used vocabularies are used. For this purpose, a specific OmniPaper vocabulary containing these metadata elements was created using RDF Schema.

In Figure 3, illustrated below, we show the properties defined in the OmniPaper RDF Schema [27].

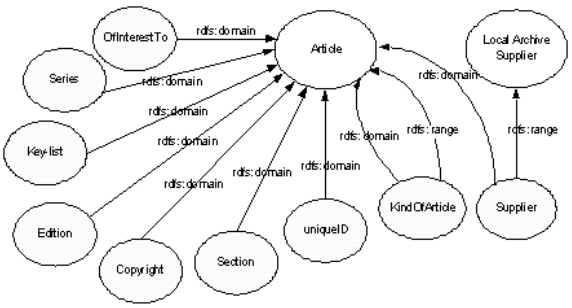

Figure 3. Schema Properties

UniqueID, Section, Copyright, Edition, Key-list, Series and OfInterestTo: Applies to all kinds of articles.

KindOfArticle: Applies to all kinds of articles, may have a value of any kind of article (any or its subclasses).

Supplier: Applies to all kind of local archive suppliers. May have a value of any kind of Local Archive Supplier (any of its subclasses).

Each property has a specific meaning, defines its permitted values, the types of resources it can describe, and its relationships with other properties [28].

In order to support this properties' definition, a set of classes was defined in the RDF Schema as explained below.

The constraint rdfs:range indicates the classes that the values of a property must be members of. Each property can only have at maximum one rdfs:range constraint. "Although it is possible to express two or more range constraints on a property, a similar outcome can achieve by defining a common super class for any classes that represent appropriate values for some property" [20]. For this purpose were defined the classes Article and LocalArchivesSupplier where the classes Interview, Review, OpinionLetter and News were defined as subclasses of the class Article, and the classes PTE, MyNews and Mediargus were defined as subclasses of the class LocalArchivesSupplier, as shown in the Figure 4 and 5.

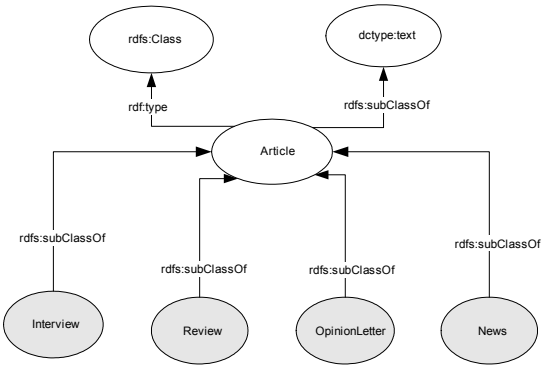

Figure 4. OmniPaper Schema – Relationships between classes

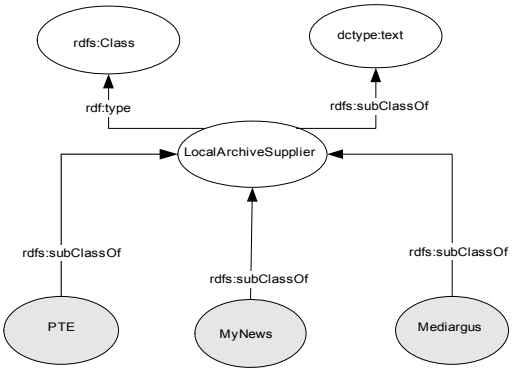

Figure 5. Omnipaper Schema – Relationship between LocalArchieveSupplier class

5. RDF Prototype

5.1. Functionalities

To recap, the main purpose of the Local Knowledge Layer is to provide a standard semantic description of all the existent articles in order to make a structured and uniform access to the available distributed archives. The RDF approach tries to achieve the best way to describe and store

metadata of provided news articles using RDF and related technologies.

The first phase of RDF prototype has the following functionalities:

(1) **Metadata transformation**: the prototype gets metadata from a XML file (news article document) and transforms into RDF file structured with the predefined RDF/XML template using XSLT.

(2) **Search**:
- Simple search (single query input field): full-text search in all metadata fields.
- Advanced search: full-text search in selected metadata fields. Fields are combined by default using the AND operator. Used fields are Title, Creator, Issued (Date), Key-list and Publisher.

(3) **Search results**:
- Display of article info (Title, Issued and Publisher)
- Display of search time
- Display of number of search results
- Ranking of search results
- Case sensitiveness
- Query keyword stemming

(4) **Article retrieval**: from XML file using XSLT.

5.2. System Architecture

The RDF prototype is a web-oriented application that the user will send and receive information through the HTTP protocol. The HTML pages are generated by a RDF Server pages (an ASP fashioned server).

The code inside RDF Server Pages (RSP) files interacts with users and the database engine, executing queries within the metadatabase and sending answers to the users. The system design is shown in Figure 6.

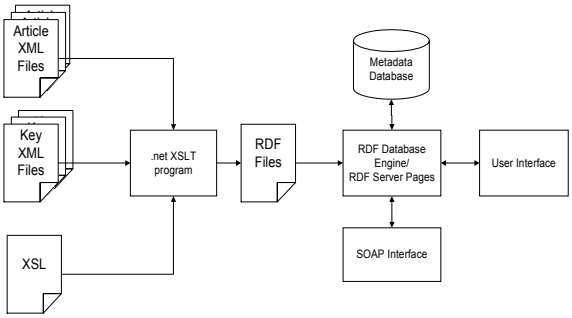

Figure 6. System Design

For this prototype's implementation, 1881 English news articles are used as a test sample. Keywords are extracted in advance by automatic keyword extraction module, and stored in XML formatted files. Metadata already exist in the news articles are mapped to OmniPaper metadata elements when transformed to RDF/XML files.

The prototype (metadatabase) stores these metadata extracted or transformed from news articles in a form of triples. Each RDF file holds one description about one article. An RDF/XML template has also been built. Figure 7 presents formats and relations of data in the RDF prototype.

In the final prototype, SOAP interface will allow the communication with all the distributed archives. Each archive has to map its own metadata to OmniPaper metadata and also implement SOAP interface to respond to the server's SOAP requests.

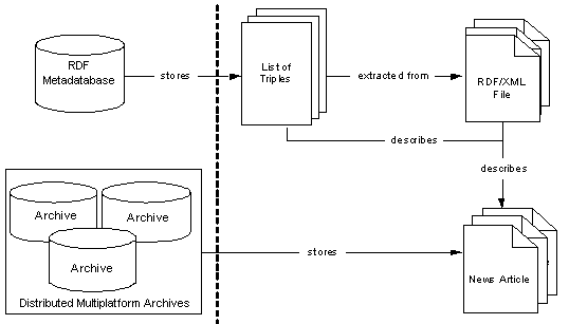

Figure 7. Format and Relation of Data

The RDF prototype was developed over RDF-Gateway, since RDF-Gateway can generate HTML pages to users by its Network IO Module, and also can create a Native Database and a Package that are by themselves the applications.

The application was developed in a scripting language called RDFQL. RDFQL is based on ECMA Script (Java Script) with query extensions to perform federated searches across multiple data sources. More details of the system architecture are described by Pereira, et al [29].

6. Future Work

The work for the Local Knowledge Layer, the construction of rich metadata layer to access the distributed heterogeneous digital news archives, is currently in the last stage. The RDF prototype for this layer will be cross-tested and compared with the Topic Maps prototype and Direct Retrieval approach.

An extension of the RDF prototype for the Overall Knowledge Layer has also been executed. In this stage, navigation of subject and concept of news articles are underlined.

All the metadata elements are common to all the distributed archives so that the common thesaurus is also used for cross-archive browsing and navigation. Particularly, IPTC Subject Code [30] is used for Subject View, and Euro Word Net [31] is used for Concept View of news articles. Currently, mapping of metadata to the common thesaurus is one of the major problems.

7. Conclusion

In the first year of the OmniPaper project, a number of significant works have been done for definition of metadata element set for the description of digital news articles and development of searchable uniformed metadatabase.

Having defined 27 metadata elements for the OmniPaper, these semantically rich elements are covering various aspects of news articles that might result in a valuable contribution for future applications related to digital libraries. In this sense, it was important to create OmniPaper Application Profile and Schema which are now public through CORES registry [25].

Choice of DCMES as a principal vocabulary for the OmniPaper metadata set is ideal as it has already widely been used across the different application domains. This must lead to greater interoperability particularly on Internet.

The RDF prototype brings all works done together into one application. It will be one of the first studies on how metadata and RDF contribute to information retrieval for heterogeneous digital resources when compared with the other approaches in the OmniPaper project. Furthermore, integration of ontologies will have to make the prototype much more powerful in navigation of information resources.

References

[1] OmniPaper – Smart Access to European Newspapers, EU project IST 2001-32174. http://www.omnipaper.org/, 2002.

[2] Resource Description Framework. http://www.w3.org/RDF/

[3] Pepper, S. and Moore, G. (2001) *XML Topic Maps (XTM) 1.0*. from http://www.topicmaps.org/xtm/1.0/

[4] News Industry Text Format. http://www.nitf.org/

[5] News ML. http://www.newsml.org/

[6] International Press Telecommunications Council. http://www.iptc.org/

[7] XMLNews. http://www.xmlnews.org/

[8] RDF Site Summary (RSS) 1.0. from http://www.purl.org/rss/1.0/spec

[9] *Dublin Core Metadata Elements Set, Version 1.1: Reference Description*. from http://dublincore.org/documents/dces/

[10] *Dublin Core Qualifiers*. from http://dublincore.org/documents/dcmes-qualifiers/

[11] RDFGateway. from http://www.intellidimension.com/

[12] Dublin Core Metadata Initiative. http://dublincore.org/

[13] DCMI Metadata Terms. From http://purl.org/dc/terms

[14] Yaginuma, T., Pereira, T. and Baptista, A. A. (2003) *Design of Metadata Elements for Digital News Articles in the OmniPaper Project*. From the proceedings of the 7th International Conference on Electronic Publishing, Portugal.

[15] Berners-Lee, T., Hendler, J. and Lassila, O. (2001). *The Semantic Web*. from http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2

[16] UKOLN. http://www.ukoln.ac.uk/

[17] DESIRE. http://www.lub.lu.se/desire/

[18] Heery, R. and Patel, M. (2002), Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne, Issue 25*. Retrieved May 12, 2003, from http://www.ariadne.ac.uk/issue25/

[19] Extensive Markup Language (XML) Activity Statement. from *http://www.w3.org/XML/Activity*

[20] Brickley, D. and Guha, R. V. (2003) *RDF Schema*. from http://www.w3.org/TR/rdf-schema/

[21] Yaginuma, T., Pereira, T. S. M. and Baptista, A. A. (2003) *Metadata Elements for Digital News Resource Description*. From the proceedings of the III Congresso Luso-Moçambicano de Engenharia, Maputo, Mozambique.

[22] Pereira, T. and Baptista, A. A. (2002) *OmniPaper Application Profile*. http://www.dsi.uminho.pt/omn/schemas/Omnipaper-ap.rdf

[23] SCHEMAS: Forum for Metadata Schema Implementors. From http://www.ukoln.ac.uk/metadata/schemas/

[24] CORES. http://www.cores-eu.net/

[25] OmniPaper Application Profile registered in CORES. (2003) Available at http://cores.dsd.sztaki.hu/?class=http%3A%2F%2Fwww.cores-eu.net%2Fregistry%2Freg%2FAppProfile

[26] World Wide Web Consortium. http://www.w3.org/

[27] Pereira, T. and Baptista, A. A. *OmniPaper Schema Namespace*. http://www.dsi.uminho.pt/omn/schemas/omn-schema

[28] Lassila, O. and Swick, R. R. (1999) *RDF Model and Syntax Specification*. http://www.w3.org/TR/REC-rdf-syntax/

[29] Pereira, T. and Baptista, A. A. (2003) *The OmniPaper Metadata RDF/XML Prototype Implementation*. From the proceedings of the 7th International Conference on Electronic Publishing, Portugal.

[30] IPTC Subject Codes. Available at http://www.iptc.org/metadata/index.php

[31] EuroWordNet. http://www.illc.uva.nl/EuroWordNet/