

Proceedings of the
**SPDECE-2012. Ninth multidisciplinary
symposium on the design and evaluation of
digital content for education**

13–15 June 2011
Universidad de Alicante
Alicante, Spain

Edited by
Manuel Marco Such
Pedro Pernías Peco



Copyright 2011 The authors

Printed in Alicante, Spain



Available online at www.sciencedirect.com



**Procedia
Social and
Behavioral
Sciences**

Procedia - Social and Behavioral Sciences 00 (2011) 000–000

www.elsevier.com/locate/procedia

SPDECE 2012

BRINGING LIBRARIES TO WEB 3.0

G.Candela^a, MP.Escobar^a, MJ. Serna^a, M.Marco^a, P.Pernías^a

Abstract

In the last decade the evolution on virtual libraries forced an extraordinary interest in new methods for delivering bibliographic content to users. Generally the actual libraries provide information from several databases and repositories that sometimes are not really integrated.

Our approach implements a conceptual model to persist and serve the information. This model follows the guidelines set by The Functional Requirements for Bibliographic Records (FRBR) that consists of an entity-relationship model developed by International Federation of Library Associations and Institutions (IFLA) providing concepts to describe the generalized view of the bibliographic universe. In this paper we propose the steps to transform the information of the actual Miguel Cervantes Virtual BVMC library to the FRBR model. The information has been extracted and transformed using business intelligence techniques.

The system is build according to SOA (Service Oriented Architecture), coupled with AJAX-based user interface. All the services build for BVMC are SOAP (Simple Object Access Protocol) allowing applications to exchange information.

We investigate the semantic approach enforcing the transition from a static information to a dynamic (collaborative) knowledge space where users are contributors as well. Semantic technologies can offer more efficient solutions for accessing content and metadata. Based on RDF, the different metadata sources like bibliographic descriptions, user profiles or bookmarks, are integrated providing communication interoperability.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [name organizer]

Keywords: digital libraries, SOA, web service, business intelligence, web semantic, social.;

1. Introduction

Information retrieval (IR) has changed considerably in the last years with the expansion of the Web (World Wide Web) and the advent of modern and inexpensive graphical user interfaces and mass storage devices. However these interfaces usually do not focus on the final user's perspective. Our approach relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user's perspective.

In this paper we propose a whole cycle of transformation and technology renewal that has been implemented in the BVMC to support the latest web trends like WEB 3.0 and remain a world leader in digital libraries.

The first problem that we faced was the different data source repositories that were not really integrated. The best solution in this case is a general and standard format that provides a model to enclose all the instances of the library. In particular, we show the migration process based on business intelligence techniques that allows the definition of often complex processes to extract, transform, and load heterogeneous data from one or more sources into a data warehouse.

In our solution we have followed and customized the IFLA study group on the Functional Requirements for Bibliographic Records (FRBR). The aim of this study is to produce a framework that provides a clear, precisely stated, and commonly shared understanding of what it is that the bibliographic record aims to provide information about.

Based on FRBR we have created a set of web services and interfaces to achieve the user's needs. Moreover we present a first version of the semantic BVMC where we are working at this moment.

2. FRBR

There are many approaches to the development of library description formats like FRBR (Functional Requirements for Bibliographic Records), MARC21 or Dublin Core. The growing importance of FRBR is due in part in a response to a growing ability of library systems to represent FRBR relationships in search systems. FRBR offers a fresh perspective on the structure and relationships of bibliographic and authority records, and also a more precise vocabulary to help future cataloguing rule makers and system designers in meeting user needs.

FRBR is a recommendation that needs to be customized to accomplish the requisites of the BVMC. The component item has been eliminated on our approach. Bárbara Tillet describes the item concept in her article as “For example, when we say book to describe a physical object that has paper pages and a binding and can sometimes be used to prop open a door or hold up a table leg, FRBR calls this an item.” [1] BVMC is a digital library and therefore there are no physical objects in its domain as it is shown in Fig. 1.



Fig. 1. FRBR structure

Based on FRBR object-oriented (FRBRoo [2]) definition and mapping to FRBRER and oriented object programming concepts, our approach includes an abstract class from where Work, Expression and Manifestation extends. This customization facilitates the search operations because all the items are treated on the same level. Moreover each FRBR concept has been extended with more concrete data. For instance, image, sound and serial expression extend from the original FRBR expression concept.

Another customization for the BVMC case is the group of items for thematic sites. Each site has assigned specific works. Therefore, a new relationship has been established between work and site.

The main goal of the application of this model is to facilitate international standardization and reduce costs for cataloging on a global scale as the BVMC does.

3. Migration Process

Generally the actual libraries provide information from different data repositories that sometimes are not really integrated. The best solution in this case is a general and standard format that provides a model to enclose all the instances of the library.

A common scenario in data migration is data warehousing, often applied in the areas of business intelligence, customer relationship management, data mining or master data management. In practice, developers often use data integration tools instead of hand-coded scripts for loading data warehouses. Extract-Transform-Load (ETL) tools allow the definition of often complex processes to extract, transform, and load heterogeneous data from one or more sources into a data warehouse. Our selected option was Pentaho Data Integration Community Edition also known as Kettle [3]. Pentaho Data Integration delivers powerful Extraction, Transformation and Loading capabilities using an intuitive, graphical, drag and drop design environment, and a proven, scalable, standards-based architecture as it is shown in Fig. 2.



Fig. 2. Pentaho ETL example

The source data in this research includes the records from a medium sized academic library and consist of 150000 records. Although the FRBR data model is well defined and mapping algorithms have been defined to migrate traditional MARC into FRBR modeled data, the application of these algorithms have typically been done in a ‘perfect’ world environment. For instance, some MARC elements do not map to anything in the FRBR model, such as MARC elements for record processing. Some FRBR attributes do not map unambiguously into MARC, or may be recorded in non-specific textual fields, such as general notes.

After analyzing the mapping between the different data sources and the FRBR model [4], we proposed a high level management of ETL sub-processes to enable their flexible re-use, optimization, and rapid development. This whole process creates a new MySQL database to gather the FRBR information as it is shown in Fig.3.

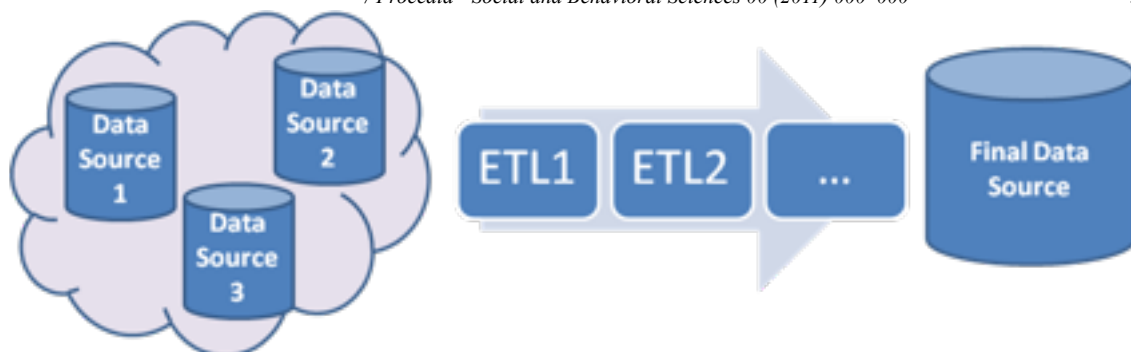


Fig. 3. Migration process

As a control system, we maintain old identifiers related with the new ones in a temporal table. Thus, database validations are possible once all data has been migrated to the new model.

In addition, our approach faced a challenge of ensuring that data being reported is current and up-to-date. Our process takes about two hours to migrate all the data to the new model. Cataloguers keep working on the old system till the moment our team decides to change to the new model.

4. Client Interface

Much of the success of Web Services and arguably SOA as a result is attributed to interoperability between platforms and programming languages. Web services are also versatile by design. They can be accessed by humans via a Web-based client interface, or they can be accessed by other applications and other Web services. Code re-use is another positive side-effect of Web services' interoperability and flexibility. One service might be utilized by several clients, all of which employ the operations provided to fulfill different business objectives. Instead of having to create a custom service for each unique requirement, portions of a service are simply re-used as necessary.

Web services are loosely coupled software components, published, located and invoked across the web. A web service comprises several operations. Each operation takes a SOAP package containing a list of input parameters, fulfills a certain task, and returns the result in an output SOAP package.

The key piece of our approach is the search web service that provides simple search, synonym search, fuzzy search, pagination, most visited items, recent items...Search operation is performed based on different fields like author, role, title, subject or date. The most relevant operations in the new search service are listed below specifying their characteristics and functionality:

- Faceted search: by author, subject and date
 - Inputs:
 - Search string.
 - Number of items to return.
 - Starting number of pagination.
 - Output: JSON text with result list.

- Catalog and content search.
 - Inputs:
 - Search string.
 - Number of items to return.
 - Starting number of pagination.
 - Output: JSON text with result list.

The new architecture uses just one data repository providing the same results as the old one. The advantage of this unique data source resides on the scalability and code re-use. Moreover our approach provides a new whole module to create groups of items that may be used in the personalized sites. The FRBR has been customized to achieve this functionality.

For instance Fig. 4 depicts a situation where the user searches “el quijote” and uses some facets. The results are grouped in two tabs, one for metadata search and one for content search. When the user checks one item, a new page is opened with all the information about the authors, subjects, title, dates... The interface provides a fragment search when the user can check the coincidences as it is shown in Fig 5.



Fig. 4. Faceted Search



Fig. 5. Coincidences Search

5. Ongoing work and conclusions

Although the practical importance of library description formats is significant, only some work on FRBR has been performed in the library research community. Most related research results are about migrating to FRBR without taking in account customized characteristics. There are other approaches where the interface appears as in WorldCat [5]. Thus, none of the existing approaches work on the full set of FRBR mechanism or support fundamental FRBR aspects, such as client interface and an administration module.

In this paper we present a generic approach for FRBR migration, meaning that we have created a complete process that migrates the data to FRBR model. Once the data is migrated, we have developed a new whole application based on an AJAX client interface to add, update or delete entities. For searching and faceting results, we provide a list of web services based on SOAP architecture.

Typical digital libraries usually focus on cataloguing resources where information retrieval relies on text search engines. Simple searching returns too many results which have to be filtered somehow. This approach does not use the metadata information to link entities and leave the user in the background.

Recent research and development in digital libraries domain focuses on using the Semantic Web and social networking technologies. Social networking services can improve the information discovery and sharing; users become active producers of the metadata and libraries can provide more focused and accurate results using recommendations techniques. Based on SOA architecture, the semantic library helps to interconnect with other applications using web services.

Usually, when a user browses a digital library, some articles and materials seem to him more valuable than others. Common practice is to bookmark those resources. Users want to see the bookmarks of their friends, and use the knowledge collected by them. To facilitate this knowledge sharing, we are developing a semantic social networking based on FOAF, FRBR and our own ontology.

At this moment, all search and browsing results in BVMC are serialized in JSON format. The main goal is allowing other libraries to use the BVMC searching services without taking in account the architecture of the foreign systems. The new web semantic approach uses RDF format where the information is also linked.

REFERENCES

- Bárbara Tillet, What is FRBR? <http://www.loc.gov/cds/downloads/FRBR.PDF>
- FRBRoo: por una perspectiva común de la información de las instituciones de la memoria. Pat Riva, Martin Doerr, Maja Žumer. Traducción: Ricardo Santos Muñoz. Biblioteca Nacional de España.
- Pentaho Kettle tool. <http://kettle.pentaho.com/>
- Application of FRBR to existing MARC data: A comparison of outcomes. Erik Mitchell, Carolyn McCallum, Patricia Strickland
- The World's Catalog. <http://www.worldcat.org>
- JeromeDL – a Semantic Digital Library. Sebastian Ryszard Kruk, Tomasz Woroniecki, Adam Gzella, and Maciej. Dąbrowski
- Functional Analysis of the MARC 21. Bibliographic and Holdings Formats. Analysis with Tables. Library of Congress. April 6, 2006. <http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html#table>

Functional Requirement for Bibliographic Records: Final Report. (2008). International Federation of Library Associations and Institutions. Retrieved January 5, 2011 from http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

Conversion Services of documentary resources from the Biblioteca Virtual Miguel de Cervantes in multiformat educative resources. 2011.

EJB Design Patterns: Advanced Patterns, Processes, and Idioms. Floyd Marinescu.