



UNIMARC in The European Library and related projects

Nuno Freire

National Library of Portugal

Lisbon, Portugal

E-mail: nuno.freire@bnportugal.pt

Meeting:

148. UNIMARC

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY

10-15 August 2010, Gothenburg, Sweden

<http://www.ifla.org/en/ifla76>

Abstract:

This paper provides an overview of the use of UNIMARC in The European Library service as well as in related projects, highlighting the current and future prospects of UNIMARC data handling and exploitation in such contexts. It includes information on issues such as UNIMARC conversion and harvesting in specific profiles, distributed search and retrieval, FRBRization, rights management and data provision to Europeana.

INTRODUCTION

The European Library¹ was created in 2005 as a service of the Conference of the European National Libraries² (CENL), aiming to offer a common access point to the distributed collections of the 48 national libraries of Europe in 32 languages. The European Library currently gives access to 150 million bibliographic records across Europe, concerning resources that can be both digital and analogical, and is a free service.

The European Library

Die Europäische Bibliothek L
os The European Library De l
uropeos La Biblioteca Europe

➔ An online service for all European National Libraries which provides access to their catalogues & digital collections via one central, multi-lingual web-interface (portal)



The European Library of today is the result of a number of earlier successful projects:

- Project Gabriel - Gateway and Bridge to Europe's National Libraries: It ran during 1996-2004 and established a joint web portal of European national libraries;
- TEL Project (The European Library: Gateway to Europe's Knowledge): It ran during 2001-2004, and laid down the foundations of the current service. The project had the participation of 9 national libraries: Finland, Germany, Italy (Florence), Italy (Rome), Netherlands, Portugal, Slovenia, Switzerland and United Kingdom.

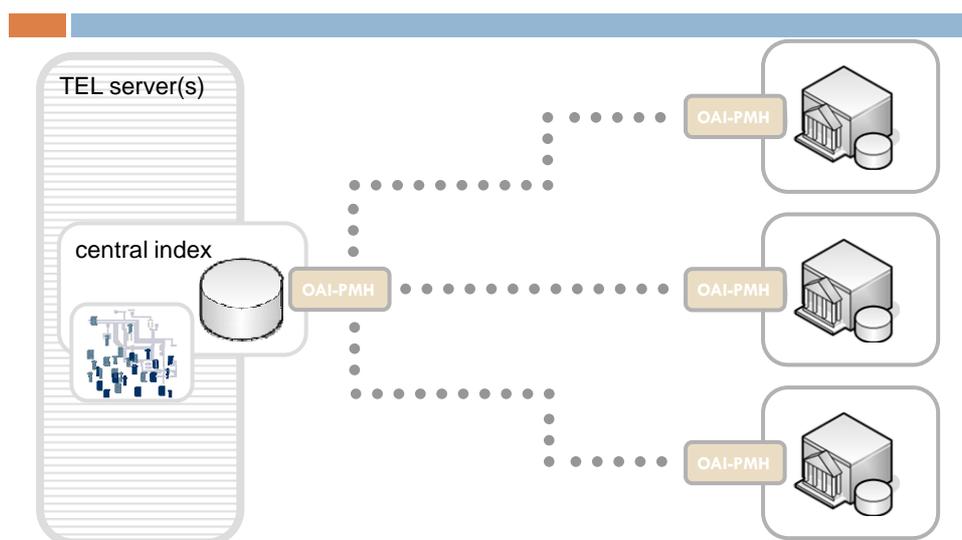
The success of The European Library has led to further developments and improvements. The number of national libraries members grew from the initial 9 to the current 48. Its growth was supported by European co-funded projects. TEL-ME-MOR (2005-2007), had as its main goal to support, stimulate and facilitate the participation of the 10 national libraries from the New Member States, which were partners in the project; the EDL project (2006-2007) worked towards the integration of additional national libraries and in enhancing the multilingual capabilities of The European Library portal. Both projects have contributed to the inclusion in the service of the collections of 19 national libraries.

TELplus was another project of The European Library context, undertaken by a consortium of 26 partners, including national libraries and research centres. TELplus provided value-adding services and products for The European Library. Its primary objectives concerned improvements to the service architectural performance, the enhancement of multilingual access functionalities, and providing access to full-text documents originating from OCR processes.

Metadata usage at The European Library

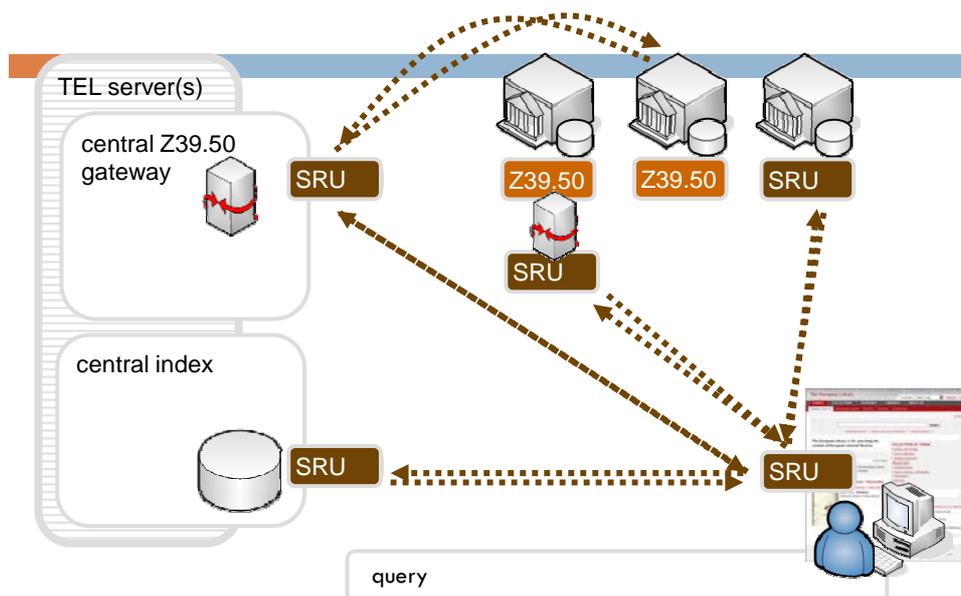
Collections from the national libraries can be made available in The European Library portal via three communication protocols: the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), Z39.50 or the Search/Retrieve via URL (SRU).

Harvesting via OAI-PMH

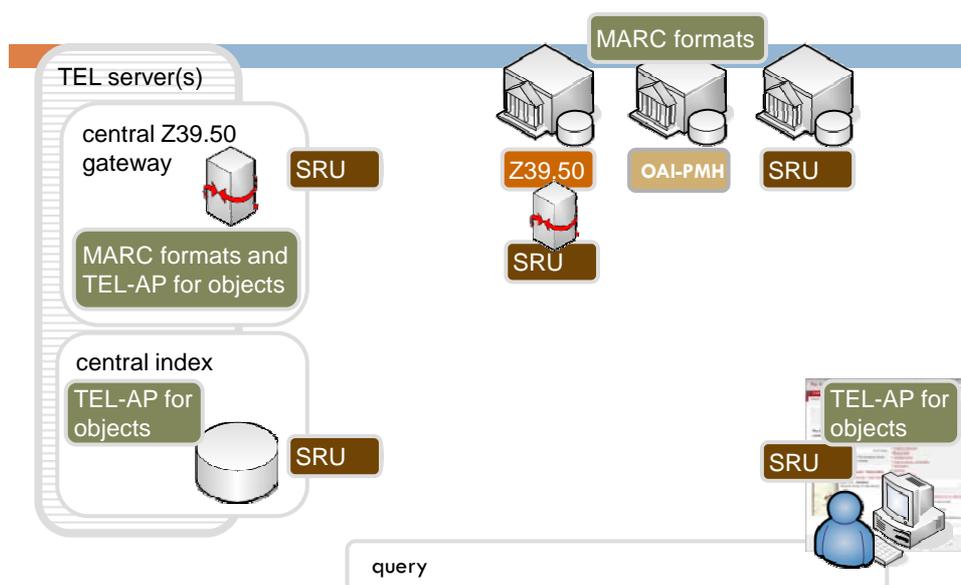


The choice of the communication protocol greatly influences the functionalities that the portal can provide to the end-user. Although all three protocols provide a standard for communication between the portal and the libraries' systems, the underlying communication paradigm is substantially different. While OAI-PMH's design allows the portal to harvest all metadata records from the libraries into a central repository, Z39.50 and SRU were designed for remote search and retrieval, therefore metadata records remain only at the data provider.

Search and retrieval via SRU



Metadata formats



The European Library portal relies on a metadata application profile based on Dublin Core for achieving interoperability between the diverse bibliographic metadata formats used by the European national libraries – The European Library Application Profile for Objects (TEL-AP). The portal and the central index are designed to work on this metadata format.

Libraries' metadata is in generally in one of the formats of the MARC family, although other formats are also in use, particularly within digital and specialized collections. Conversion from these formats to the TEL-AP may be performed at the national

libraries or at The European Library. In the second case, conversion may be performed in the central Z39.50 gateway or during ingest in the central index.

UNIMARC at The European Library

A study performed in 2007 addressed the bibliographic data formats usage on national libraries³. The study included twenty nine national libraries and found that UNIMARC or a derivative was in use at ten of those libraries: Albania, Bulgaria, Cyprus, France, Greece, Lithuania, Portugal, Republic of Macedonia, Serbia, and Slovenia.

UNIMARC usage in The European Library

Die Europäische Bibliothek L
sos The European Library De l
uropeos La Biblioteca Europe

- At the National Libraries
 - ▣ From a study conducted in 2007 by Christoph Schmidt-Supprian:
 - The study included 29 National Libraries
 - *“... Concerning data formats and descriptive rules, international standards were clearly dominant. Among 29 libraries, ten used UNIMARC or a derivative...”*
 - Albania, Bulgaria, Cyprus, France, Greece, Lithuania, Portugal, Republic of Macedonia, Serbia, and Slovenia

Within The European Library, UNIMARC was supported since the original project, where three of the eight countries used UNIMARC. A conversion of UNIMARC to TEL-AP is supported by the Z39.50 gateway and by the OAI-PMH metadata harvesting system. This conversion is documented in The European Library Handbook, which is available in the members' area of the portal.

A particular XSLT implementation of the conversion between UNIMARC and TEL-AP is available, and supported by the OAI-PMH software made available for libraries (<http://repor.ist.utl.pt>).

UNIMARC usage in The European Library

- Transformation of UNIMARC to the TEL-AP for objects
 - Supported in the Central Z39.50 Gateway
 - Documentation for transformation of UNIMARC is available in The European Library Handbook (members area)
 - XSLT Stylesheets for transformation of UNIMARC are available in the OAI-PMH software made available for libraries
 - <http://repor.ist.utl.pt>

Other recent or ongoing developments in The European Library include support for UNIMARC. These are described in the following sections.

UNIMARC in Project TELplus

Project TELplus, which ended in December 2009, included UNIMARC support in two of its work packages. Work package 2 - “Improving the usability of The European Library through OAI-PMH compliancy” – aimed at increasing the adoption of OAI-PMH within the participating libraries of The European Library. It provided tools and guidelines for the implementation of OAI-PMH on libraries’ catalogues and collections. It also addressed the central usage of OAI-PMH at The European Library with the development of tools and their integration with the central index. It was in this work package that the XSLT implementation of the UNIMARC conversion to TEL-AP was developed and made available for use in the national libraries and in The European Library.

UNIMARC usage in Project TELplus WP2



- WP2 – Improving the usability of The European Library through OAI-PMH compliancy
 - ▣ Provide tools and guidelines for the implementation of OAI-PMH on libraries' catalogues and collections
 - ▣ Provide tools to expand The European Library central index
- Implemented transformation of UNIMARC to TEL-AP for objects:
 - ▣ Based on XSLT style sheets (for MARCXML and MarcXchange)
 - ▣ In the OAI-PMH software made available for libraries
 - Allowing libraries to transform their UNIMARC records to TEL-AP for objects
 - ▣ In the central OAI-PMH software of The European Library
 - Allowing UNIMARC metadata to be harvested and transformed centrally to TEL-AP for objects

Work package 3 of TELplus - “Improving accessibility” – was aimed at improving accessibility in three ways: by enhancing full-text searching; by aggregating related resources; and investigating automatic vocabulary mappings”. It included a task where UNIMARC was used: “Setting up a FRBR repository for the European Library”. This task had the support of several national libraries that contributed to this experiment with bibliographic and authority data, and as advisors.

UNIMARC usage in Project TELplus WP 3



- WP 3, task 3.3 – FRBR aggregation, search and browsing
 - ▣ The purpose of this task was to develop solutions to support alternative services of searching in The European Library according to the FRBR paradigm
- The work focused on a specific collection, especially to make it easier to assess the results
 - ▣ The works of the Nobel Prize winners in Literature
 - ▣ Proof of concept available at:
 - <http://digmap2.ist.utl.pt:8080/lgteFrbr/evaluation.html>

The main objective of this task was to investigate possible improvements for search and accessibility of the contents of The European Library through the implementation of the FRBR model in the search functionality of the portal.

This work was focused on a specific collection rather than on the complete catalogues, since a controlled collection would allow a better evaluation of the results. The chosen collection was the works of the Nobel Prize winners in Literature.

UNIMARC usage in Project TELplus WP3



- Both bibliographic and authority records from library members of The European Library were gathered
- 5 out of 12 libraries that contributed to the collection sent UNIMARC metadata records
- FRBRization of UNIMARC data was conducted for the following National Libraries:
 - ▣ French National Library
 - ▣ Martynas Mazvydas National Library of Lithuania
 - ▣ National Library of Portugal
 - ▣ National Library of Serbia
 - ▣ Royal Library of Belgium

Both bibliographic and authority records from library members of The European Library were gathered and a FRBRization process was applied to them. Five out of twelve libraries that contributed to the collection sent UNIMARC metadata records.

The particular FRBRization of UNIMARC data was addressed and applied to the collections of: French National Library, Martynas Mazvydas National Library of Lithuania, National Library of Portugal, National Library of Serbia, and Royal Library of Belgium.

This task resulted in a proof of concept which is still available at: <http://digmap2.ist.utl.pt:8080/lgteFrbr/evaluation.html>

UNIMARC in Project ARROW

Project ARROW⁴ - Accessible Registries of Rights Information and Orphan Works - addresses the interoperability of rights information along the process of rights clearance for digitization purposes. It covers the identification of the work, the

clarification of its intellectual property rights status and the identification of the rights holders.

UNIMARC usage in Project ARROW



- **ARROW** - Accessible Registries of Rights Information and Orphan Works towards Europeana
- Consortium of European national libraries, publishers and collective management organisations, aiming to:
 - Deliver a rights information infrastructure operable within and across borders
 - Facilitate identification of right holders (authors/visual artists/publishers)
 - Facilitate the identification of the IPR status
 - Provide test beds for
 - Business models for digital libraries
 - Registries for Orphan works (OW) and Out of Print Works (OPW)
 - Rights clearance mechanisms
 - Model licenses developed by the EC High Level Expert Group (HLEG)
 - Interoperability and criteria for interoperability

ARROW depends on the availability of existing bibliographic data and rights information. There is already an established and generally well-regarded information infrastructure for print material, through national bibliographies, books in print and the databases of rights organisations. Currently, these sources are not inter-operable because of differences in data collection policies and data schemas. Bibliographic databases rarely include metadata about rights ownership and usage policies. Such information is usually held in a wide array of formats by publishers, collecting societies and authors.

Bibliographic data from the catalogues of Europe's national libraries are one of the key data sources in ARROW. For this reason, ARROW is building on the existing interoperability achieved through The European Library, using it as a core source of bibliographic data.

UNIMARC usage in Project ARROW



- The European Library, as an aggregator of bibliographic metadata from the National Libraries
 - ▣ Will be the main source of information to identify all manifestations related with a work
- Early results from the project identified that TEL-AP for objects would not fulfil the requirements of the project
 - ▣ As a result a new central index based on MARC formats is being prepared
 - Support for MARC21 already implemented
 - Support for UNIMARC is under development

Early results from the project identified that TEL-AP would not fulfil the requirements of the project. As a result a new central index based on MARC formats is being prepared at The European Library. The first library to have its catalogue in ARROW, by means of the UNIMARC format, is the Bibliothèque nationale de France.

UNIMARC in data provision to Europeana

The European Library is also an aggregator for Europeana. Although so far most of the provision of metadata from The European Library to Europeana has been based on the TEL-AP, in the future it will be based on the MARC formats. The release of the new Europeana Data Model, planned for Spring 2011, will allow the semantic quality of the metadata from the MARC formats to be better preserved when it reaches Europeana. Conversion of UNIMARC to the Europeana Data Model will be a priority for The European Library.

Future work

- In the context of The European Library as an aggregator for Europeana
 - ▣ New Europeana Data Model (EDM) planned for the Danube release (Spring 2011)
 - ▣ The European Library will support the transformation of UNIMARC to EDM



References

¹ The European Library portal is available at: <http://www.theeuropeanlibrary.org>

² Members of CENL are the national librarians of the Council of Europe member states. Further information about CENL is available at: <http://www.cenl.org/>

³ Schmidt-Supprian, Christop. 2007. Controlled Vocabularies in a Multilingual Federated Search Environment: The Example of The European Library

⁴ <http://www.arrow-net.eu>