

# The UNIMARC Metadata Registry

José Borbinha, Hugo Manguinhas



• UNIMARC is a family of metadata schemas with formats for descriptive information, classification, authorities and holdings. The UNIMARC activities are an IFLA Core Activity, the ICABS: IFLA-CDNL Alliance for Bibliographic Standards.

• In 2004 the National Library of Portugal (BN) started an activity to develop a UNIMARC Registry. The purpose is to support the processes around the evolution of the UNIMARC formats and provide a reference point for the professionals and organizations using it.

• The first target was on the bibliographic format, especially the development of a formal description of its structure and grammar in XML. The application of those results was tested and validated in two practical cases:

- The on-line textual publication of the format, in this moment in Portuguese and English languages
- The automatic validation of bibliographic records.

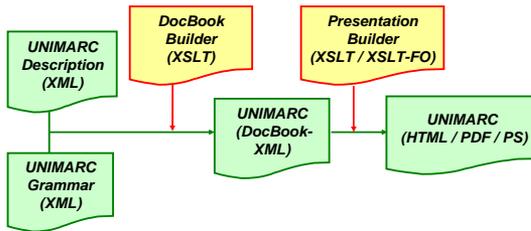
• The system for the on-line publication provides persistent URN identifiers for the most recent version of the elements of the format. In this moment those URN are resolved only for HTML pages, but in the future it'll be possible to return also the same information in a structured XML reply.

• A standalone tool, MANGAS, is available for the automatic validation of bibliographic records. This tool validates the integrity of a UNIMARC record and it's information, supporting also features to provide detailed reports and correct records. The engine is also used in an equivalent on-line service.

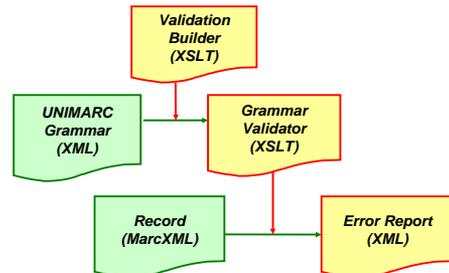
• The next steps of this activity will be:

- To revise and finish the formal UNIMARC Registry according to the ISO/IEC 11179 (formal description of terms, revision of the URN space, etc.)
- To develop a stable web-service for machine access
- Add more UNIMARC formats, versions and languages, and formalize the descriptions of changes between versions.

## UNIMARC Publishing Workflow



## UNIMARC Validation Workflow



## The UNIMARC Registry

## UNIMARC in XML

## Applications

### Purpose

- Publishing of Human Readable Versions of UNIMARC
- Multilingual
- Multi-version
- For all the formats
- Publishing of Machine Readable Versions of UNIMARC
- Grammar (structured rules)
- Descriptions (structured text)

### Main Concepts

- Each format has its own :
  - XML grammar file
  - URN space
  - Version control
- Multilingual, having for each language:
  - Its own XML description file
  - Its own URN sub-space

### Applications

- Online Publishing
  - For Humans (access to a readable description any version of the format in any language)
  - For Systems and Machines (access to structured descriptions of the format for building user interface applications)
- Record Validation
  - Quality Control Services (validation of the consistency of records according to any version of the format)
  - Editing and Cataloguing Services (automatic configuration of editing UNIMARC interfaces for professionals)

> UNIMARC Grammar

```
grammar format="bibliographic" version="1.0" locale="pt-PT"
<?xml version="1.0" encoding="UTF-8" standalone="true" repeatable="false"
  <!-- Control Fields -->
  <controlfield tag="001" mandatory="true" repeatable="false"
    <!-- Data Fields -->
    <datafield tag="010" mandatory="true" repeatable="true"
      <!-- Element -->
      <!-- Subfield -->
      <!-- Value-set -->
      <!-- Indicator -->
      <!-- Element Value -->
      <!-- Subfield Value -->
      <!-- Element Value Set -->
```

> UNIMARC Description

```
<description-list xmlns="urn:unimarc:grammar:bibliographic:1.0" locale="pt-PT">
  <description-list>
  <description-item>
    <description>
    <examples>
    <definition>
    <notes>
    <occurrence>
```

Service at <http://www.unimarc.info>

Block, Field, Formats and Versions, Format Index, Indicator, Subfield, Navigation, Record View, Batch progress information, Detailed Validation Report

### UNIMARC in DocBook (<http://www.docbook.org>)

### Mangas On-line (Diagnostic Service)

Record, Reports, Detected Errors, Diagnostic Report (HTML), Diagnostic Report (XML)

