

Common Access Point for the General State Authority Libraries

Project Report

Working Group on the Union Catalogue for the General State Authority Libraries

Coordinator: Domingo Arroyo Fernández

Sub-Directorate-General for Library Coordination, Ministry of Culture

Adviser: Alicia Sellés Carot. MASmedios

Date: 24/01/2011



BAGE UNION CATALOGUE WORKING GROUP

Coordinator:

Domingo Arroyo Fernández
Head of the Information Distribution and Communication Service
Sub-Directorate-General for Library Coordination
Ministry of Culture

Members:

National Library

José Ramón García Amo, Head of the Information Technology Coordination Unit
Marina Jiménez Piano, Head of the Automation and Organisation Area

Ministry of Foreign Affairs and Cooperation

María Begoña Ibáñez Ortega, Head of the Documentation and Publications Area
Ricardo Perera Manzanedo

Ministry of Culture

Elena García Bracamonte, Service Head of the Cultural Documentation Centre
Begoña Ricote Ricote, Sub-Directorate-General for Library Coordination

Ministry of Defence

Margarita García Moreno, Head of the Library Coordination Unit

Ministry of Economy and Finance

María Luisa Martínez Soro, Service Head of the Central Library
Agustín Torreblanca López, Technical Director responsible for Archives and Libraries

Ministry of Education

Ernesto Calbet Roselló, Head of the Documentation and Library Area

Ministry of Infrastructure

María Concepción Sanz Bombín, Director of the Transport Documentation Centre
Jaime Peón Pérez, Assistant Sub-Director of the Sub-Directorate-General for Cooperation and Interdepartmental Relations

Ministry of the Interior

Juan Antonio Narváez Gómez, Service Head of the Central Library
María Rosa Martín de Vega, Head of the Studies, Documentation and Publications Area

Ministry of Justice

Carmen Sañudo Sánchez-Garnica, Sub-Director-General for Documentation and Publications

Ministry of the Environment and Rural and Marine Affairs

Carmen Casal Fornós, Head of the Library and Documentation Area

Ministry of Regional Policy and Public Administration

Carlos Ibáñez Montoya, Head of the Publications, Libraries and Documentation Area

Ministry of the Presidency

Juan José Álvarez Galán, Area Head
Silvia Fuentes López, Head of the Libraries Service

Ministry of Health, Social Policy and Equality

Ramón Barga Hernández, Technical Director for Documentation and Libraries
María Teresa Escalada Buitrón, Technical Director of the Sub-Directorate-General for Cooperation and Institutional Relations

Ministry of Employment and Immigration

Felisa Suárez Riestra, Technical Director of Libraries
María José Moreno de Frutos, Service Head of the Sub-Directorate-General for the Data Process

EXECUTIVE SUMMARY

This report, drawn up by the General State Authority Libraries (BAGE) Union Catalogue Working Group, advised by the company MasMedios, presents a detailed solution to the mandate set out in Article 1.2 b) of *Royal Decree 1572/2007, governing the coordination bodies of the General State Authority Libraries and their public agencies*, which proposes the creation of a Common Access Point (the 'PCU') for the collections of the General State Authority Library and public agencies. This solution is based on the construction of a Union Catalogue formed by the harvesting of metadata from the catalogues of individual libraries.

It draws on the work performed by the BAGE Union Catalogue Working Group and the current state of development of Union Catalogues internationally.

The basis for development of the project is eXtensible Catalog, a suite of open-source tools which cover most functionalities required for a Union Catalogue, among others the possibility of importing metadata from different ILMS through OAI-PMH harvesting, an OAI layer for those ILMS which do not support this, the normalisation of metadata from different sources (MARC21 or Dublin Core), transformation into a schema which employs the most advanced bibliographical standards: FRBR and RDA; the possibility of building advanced consultation, discovery and data presentation interfaces: web, mobile, linked open data etc. It furthermore provides for real-time connection with individual ILMSs to obtain copy availability information via NCIP, allowing inter-library loan services to be set up.

A figure presents the functional structure of the Common Access Point, divided into three sections: the individual catalogues of libraries or groups of libraries; the Union Catalogue which is the basis for the common data which normalises, transforms and aggregates the records; and, lastly, the different user interfaces.

The design of the PCU accounted for all associated legal regulations: the Electronic Access Act 11/2007, the National Interoperability and Security Schemas, among others; standard metadata formats; and the most successful communications protocols for this type of application. Consideration has also been given to the methodological requirements which the General Coordination Committee established as a guideline for the project: open standards, open source in order to foster interoperability, and control over software development.

The project planning chapter aims to specify all the tasks, grouped into three major areas: project management and monitoring, the platform for the Common Access Point and, lastly, the section covering those adaptations which libraries will need to implement in order to take part.

Three project development phases are defined: initial phase, with the key aim of providing the general public with a BAGE Union Catalogue for consultation, with bibliographical records which have been subjected to basic aggregation and normalisation; in the second consolidation phase the proposal is for a higher level of service and integration than at the previous one, in terms both of the number of libraries and the quality of user services and records; the third and final phase is intended to develop all aspects viewed as value-added services.

We lastly focus on the development of the project at the initial phase, expected to last a period of eighteen months, providing details both of the tasks to be performed and the human resources and job profiles required, along with the technological resources and infrastructure.

Abbreviations employed in this report

AGE – *Administración General del Estado* [General State Authority]

API - Application Programming Interface

BAGEs – *Bibliotecas de la Administración General del Estado* [General State Authority Libraries]

DC – Dublin Core

EUPL – European Union Public Licence

FRBR – Functional Requirements for Bibliographic Records

LOD – Linked Open Date

MARC – MACHine-Readable Cataloging

MST – Metadata Service Toolkit (XC component)

NCIP - NISO Circulation Interchange Protocol

OAI-PMH - Open Archives Initiative - Protocol for Metadata Harvesting

PCU – *Punto de Consulta Único* [Common Access Point]

RDA – Resource Description and Access

ILMS – Integrated Library Management System

XC – eXtensible Catalog

SUMMARY

EXECUTIVE SUMMARY	3
ABBREVIATIONS EMPLOYED IN THIS REPORT	4
SUMMARY	5
1. INTRODUCTION	7
1.1 THE GENERAL STATE AUTHORITY LIBRARIES AND THEIR COORDINATION STRUCTURES.....	7
1.2 STATE PUBLIC LIBRARIES.....	7
1.3 UNION CATALOGUES AND THE CURRENT DIGITAL ENVIRONMENT	8
1.4 STARTING POINT	8
2. THE BAGE COMMON ACCESS POINT	11
3. EXTENSIBLE CATALOG	13
3.1 DESCRIPTION AND GENERAL CHARACTERISTICS	13
3.2 COMPONENTS.....	15
3.2.1 Metadata Services Toolkit	16
3.2.1.1 Objective and definition	16
3.2.1.2 Functional characteristics	16
3.2.1.3 XC schema	18
3.2.1.4 Application in the BAGE Common Access Point	20
3.2.2 Drupal Toolkit	20
3.2.2.1 Objective and definition	20
3.2.2.2 Functional characteristics	21
3.2.2.3 Application in the BAGE Common Access Point	22
3.2.3 OAI Toolkit.....	22
3.2.3.1 Objective and definition	22
3.2.3.2 Functional characteristics	22
3.2.3.3 Application in the BAGE Common Access Point	24
3.2.4 NCIP Toolkit	25
3.2.4.1 Objective and definition	25
3.2.4.2 Functional characteristics	25
3.2.4.3 Application in the BAGE Common Access Point	26
4. DESIGN OF THE COMMON ACCESS POINT	27
4.1 SCHEMA OF THE COMMON ACCESS POINT	27
4.1.1 The individual library catalogues	28
4.1.2 The Union Catalogue	28
4.1.3 User and systems interfaces	28
4.2 METHODOLOGICAL REQUIREMENTS	29
4.3 DOCUMENTARY ASPECTS	29
4.3.1 Legacy formats: MARC, DC	30
4.3.2 Metadata input control.....	30
4.3.3 Domains employing the XC schema	30
4.4 TECHNOLOGICAL ASPECTS.....	30
4.5 LEGAL ASPECTS.....	31
4.5.1 Royal Decree 1572/2007, on the regulation of General State Authority and public agencies library coordination bodies.....	31
4.5.2 Electronic Access of Citizens to Public Services Act 11/2007	31
4.5.3 Royal Decree 4/2010, the National Interoperability Schema	32
4.5.4 Royal Decree 3/2010, National Security Schema	32
4.6 KEY ASPECTS.....	33

4.6.1	Identification of sources and types of metadata.....	33
4.6.2	Metadata harvesting and extraction	34
4.6.3	Processing and storage of metadata	36
4.6.4	Creation of the Common Access Point: Service	37
5.	PLANNING.....	39
5.1	PCU WORKING LINES.....	39
5.1.1	AREA 1: Platform for the PCU (Common Access Point)	40
5.1.2	AREA 2: Participating libraries and specific needs	42
5.1.3	AREA 3: Project management and coordination.....	43
5.2	DEVELOPMENT PHASES OF THE COMMON ACCESS POINT PROJECT	44
5.2.1	INITIAL PHASE: Definition and objectives.....	44
5.2.1.1	COMMON ACCESS POINT:	44
5.2.1.2	BAGEs	45
5.2.1.3	Project management and coordination	46
5.2.2	INTERMEDIATE PHASE	46
5.2.2.1	COMMON ACCESS POINT:	46
5.2.2.2	BAGEs	47
5.2.2.3	Project management and coordination	48
5.2.3	ADVANCED PHASE	48
5.2.3.1	COMMON ACCESS POINT:	49
5.2.3.2	BAGEs	50
5.2.3.3	Project management and coordination	50
6.	DEVELOPMENT OF THE INITIAL PHASE.....	53
6.1	DETAIL OF TASKS.....	53
6.1.1	Detail	53
6.1.2	Timescale	55
6.2	HUMAN RESOURCES AND WORKING TEAM	57
6.3	TECHNOLOGICAL RESOURCES	58

1. Introduction

1.1 The General State Authority Libraries and their coordination structures

According to data from the Spanish library directory¹, there are currently 982 General State Authority libraries, most of them covered by the terms of Royal Decree 1572/2007 governing the coordination bodies of libraries of the General State Authority and of its public agencies.

The BAGEs overall present a highly varied landscape, in terms of both library and service development and automation processes and systems. The results of the survey² presented in October 2009, and drawn up by the Permanent Secretariat of the General Coordination Committee for the BAGEs, reveals this diversity in terms of systems and formats, along with serious shortcomings: a substantial proportion of libraries are not automated (40%) and standards are not sufficiently employed.

The BAGE coordination bodies, the Ministerial Committees and General Committee, developed in the aforementioned Royal Decree 1572/2007, represent the organisational framework³ for libraries and those responsible for coordination to take the action required in order to allow a project such as the Common Access Point to prosper. In items [3](#) and [4](#) we detail these actions.

1.2 State Public Libraries.

Under the terms of Article 2.2 of Royal Decree 1572/2007, State Public Libraries, the administration of which is transferred to the corresponding Autonomous Region, are excluded from the scope of application of the Royal Decree. However, to the extent that these libraries are state-owned it has been deemed desirable for them to be involved in projects set up within the context of the coordination structures governed by Royal Decree 1572/2007, albeit at all times within the terms and scope laid down in the State Public Libraries and Spanish Library System Regulation (Royal Decree 582/1989, of 19 May 1989, in the corresponding management agreements for those libraries and in Article 14.3(d)) of Reading, Books and Libraries Act 10/2007, of 22 June 2007.

¹ Ministry of Culture *Spanish libraries directory*, available at <http://directoriobibliotecas.mcu.es/>

² Ministry of Culture *Results of the survey into the state of automation at General State Authority libraries*, 2009; available at <http://www.mcu.es/bibliotecas/MC/BAGE/SecretariaPer.html>

³ Further information on the BAGE coordination bodies at <http://www.mcu.es/bibliotecas/MC/BAGE/index.html>

Introduction

1.3 Union Catalogues and the current digital environment

Libraries are currently experiencing a debate as to the purpose of catalogues within the digital environment, with user expectations having changed radically over recent years. There are certain basic agreements which we must consider within our project:

- Users seeking information expect access which is as immediate as possible to the referenced resource, either by accessing the complete text at that time or having some simple way of obtaining it. They furthermore demand a series of functionalities to which they are now accustomed in their daily access to the digital world: simple searches, integration with social networks, intuitive navigation, access from any device.
- Search and loan systems need to be reconsidered in accordance with general working frameworks and standards, allowing libraries to interoperate with other participants as an additional player in the digital world.
- Traditional Union Catalogues, which display only references, are of very limited use. Their future is far from certain unless they can be integrated with other agents in the digital world and provide value-added services.

At the international level the Union Catalogues currently being designed aim to incorporate functionalities serving to meet the expectations of today's user, and through appropriate interoperability with other systems, to integrate their services into the current digital environment. One example would be LIBRIS (the Swedish Union Catalogue).

At the national level we have within the Library Cooperation Council, the Technological Foundations Working Group, which is working on a Union Catalogue for Public Libraries which will allow a nationwide inter-library loan service to be constructed.

1.4 Starting point

The AGE General Library Coordination Committee analysed at its meeting on 22 September 2009 the results of the survey into the state of automation of AGE libraries and debated the technical options for the implementation of the Common Access Point, based on the reports drawn up by the Sub-Directorate-General for Library Coordination.⁴

The same meeting agreed to set up its own Working Group, the BAGE Union Catalogue Working Group, with the following objectives:

⁴ See: *Establishment of a common access point: draft project; Development of a ILMS – draft project* ; both draft projects available at <http://www.mcu.es/bibliotecas/MC/BAGE/SecretariaPer.html>

Introduction

- **The development of an integrated library management system** [ILMS], based on some kind of open source solution to assist libraries in their automation process within the Common Access Point project.
- The creation of the **PCU** Common Access Point itself (the object of this report).

In order to fulfil the first objective the working group drew up an evaluation report and requirements for the initial phase of development⁵. Development work has already begun, and the new system, named **KOBLI**, based on KOHA, will be available in the first quarter of 2011.

Another result of the Working Group was the development of a format converter from IBERMARC to MARC21⁶, assisting towards both objectives. This converter has been released as open source code under an EUPL licence.

⁵ BAGE Union Catalogue Working Group, Evaluation report of ILMS KOHA for the General State Authority libraries, Sept. 2010. Available at <http://hdl.handle.net/10421/4711>

⁶ BAGE Union Catalogue Working Group, Transformation from IBERMARC to MARC21 with ibermarc2marc21.xsl, <http://hdl.handle.net/10421/4804>

2. The BAGE Common Access Point

The BAGE Common Access Point [*Punto de Consulta Único, or 'PCU'*] project comprises the generation of a Union Catalogue, in accordance with Royal Decree 1522/2007, through the harvesting of metadata from the various BAGE catalogues.

The PCU will constitute the Information System for the BAGEs and serve as the basis for the provision of a range of services: document access, bibliographical tools, social tools, etc., serving to integrate BAGE services within the current digital environment.

The mission of the PCU is, taking as its starting point the diversity of bibliographical records at the BAGEs, to create uniform metadata with a rich structure providing users with additional quality services.

The PCU aims to support and contribute to the development of a new Union Catalogue model by obtaining, through automated management, a complex metadata structure, while also making use of advanced discovery interfaces.

The PCU is, meanwhile, intended to increase the visibility of the BAGEs' valuable collections and promote the use among the BAGEs of metadata and interoperability standards.

The Working Group chose the tools developed by the eXtensible Catalog project as being appropriate in order to achieve these goals.

3. eXtensible Catalog

eXtensible Catalog is an open source user-centred software intended as the software for next generation libraries. It has 4 software components which can be used independently, to meet specific needs, or in combination in order to provide an end-to-end discovery system to connect library users to resources, and vice versa.

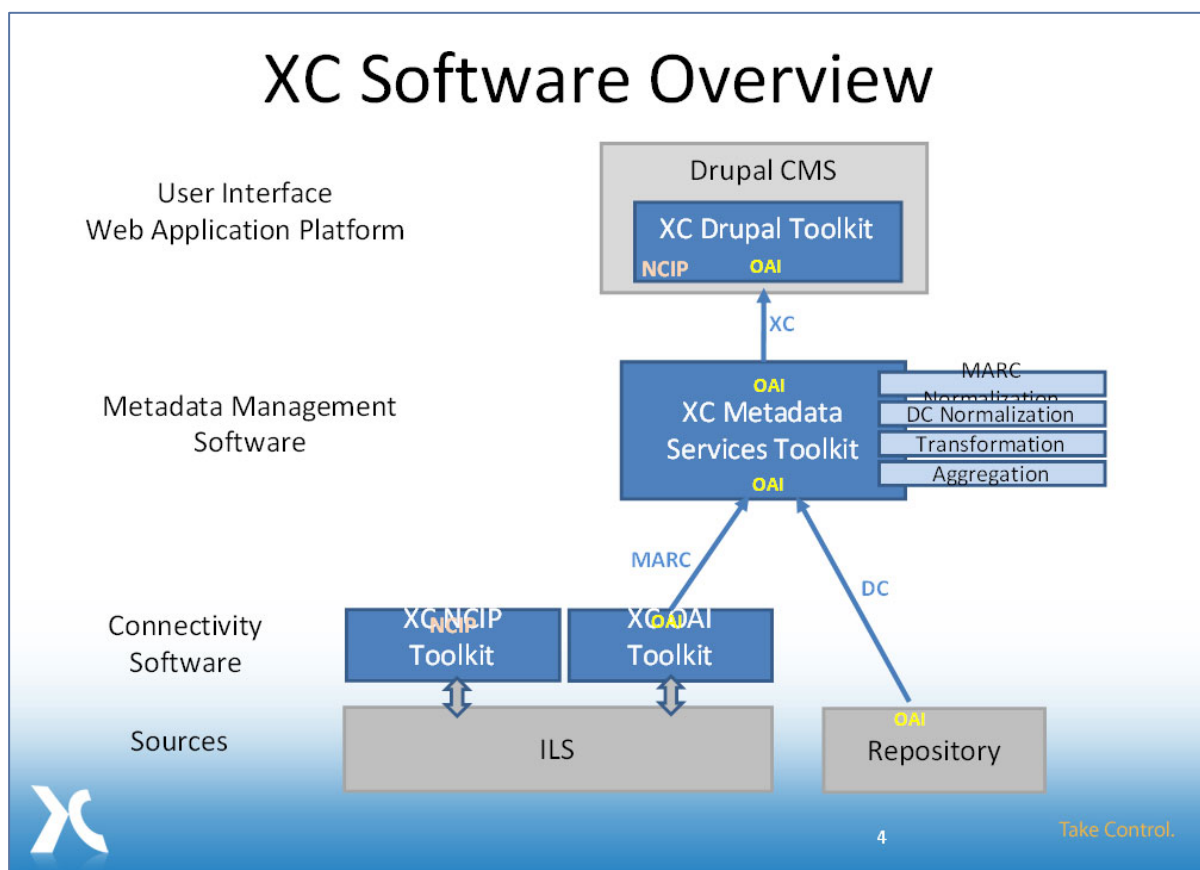
3.1 Description and general characteristics

The aim of the eXtensible Catalog (**XC**) project is to provide open source applications which simplify user access to all digital and non-digital library resources.

The XC project team have defined a series of objectives and characteristics:

- Provide a software which can harvest, store and serve up information about library resources within the applications which the libraries are already using (library management systems, learning tools, repositories and/or content management systems).
- User interface including a full set of Web 2.0 functionality (blogs, labelling, folksonomies (social labelling), RSS).
- Provide libraries with tools to create their own web service.
- Remove barriers to the exploitation and reuse of resources in terms of metadata and systems, providing an integration platform which normalises the existing metadata, generates relationships or facets for the metadata automatically, and is capable of integrating heterogeneous data and systems in one single location through metadata harvesting.
- Allow libraries to harvest, share and make available the metadata generated by users (such as resource recommendations).
- User analysis for the design and development of XC software functionality.
- Publish the XC code as open source code allowing it to be supported by a community.

The following figure sets out the functional system and communication philosophy established between the four XC components. A later section details the characteristics of each one.



We highlight below the issues we consider fundamental in the design of the tool:

- **Toolkit:** The toolkit concept breaks with the traditional idea of integrated management in terms of:
 - Independence: respecting the systems which will communicate in XC and also moving away from the restrictions associated with each of them (metadata, standards, local policies...).
 - Modularity: There is no need to deploy all modules; each project will need to analyse its needs and objectives, and implement them on this basis.
- **Technology, protocols, standards:** XC employs a variety of technologies and protocols, mainly:
 - Java, PHP, SQL, XML
 - Apache SOLR and Lucene
 - Web services
 - CMS Drupal
 - MARC metadata and Dublin Core

eXtensible Catalog

- LDAP
- OAI – PMH
- **Version:** XC is still a tool undergoing development and evolution. The eXtensible Catalog Organization, LLC, is responsible for:
 - Continuous improvement and maintenance of the XC software
 - Guiding and directing the evolution of XC
 - Support for the creation of a community of users and developers
 - They have developed and released certain components and functionalities.
- **Uses and applications:** XC has considerable potential for application to the BAGE Common Access Point, as it has been developed in wide-ranging environments such as:
 - Using Drupal toolkit as a discovery tool, requiring the use of the AOI toolkit and MST (with metadata transformation). E.g. Consultation of a consortium catalogue.
 - Using the AOI toolkit and MST to harvest MARC metadata from the ILMS for reuse in an “*ad hoc*” web application (potentially including access to circulation modules using NCIP)
 - Automatic generation of a Union Catalogue using OAI Toolkit and MST (planning harvestings and normalisation services)
 - Complete use of the application employing OAI Toolkit to convert MARC data into MARCXML, and MST to normalise and correct them, etc, maintaining the original MARC data to be returned to the ILMS; it would otherwise prove necessary to redefine batch processes within the ILMS to restore the original data to it.

3.2 Components

Each of the XC components can operate independently allowing its operational method to be evaluated, with aggregation or communication ultimately with the Drupal CMS.

The eXtensible Catalog components refer to 3 fundamental aspects:

- User interface or aspect: Drupal Toolkit. Integrated with the Drupal CMS.
- Metadata management: Metadata Services Toolkit. Tools for the processing and aggregation of metadata.

eXtensible Catalog

- Connectivity with the Library Management System:
 - OAI Toolkit. Synchronisation of MARC metadata with XC.
 - NCIP Toolkit. Connects the ILMS circulation system to XC.

As already mentioned, the XC tool is still under development; some modules are fully functional, others are being developed but have active functionalities, while others are yet to be developed.

3.2.1 Metadata Services Toolkit

3.2.1.1 Objective and definition

Metadata Services Toolkit comprises a main web application (MST) and a suite of plug-ins and metadata services designed to process metadata and generate new records (transforming MARC and Dublin Core) to even provide control over authorities, FRBR metadata, and work with faceted navigation interfaces.

Metadata Services Toolkit includes a web-based user interface which administrators and cataloguers can use to harvest metadata from repositories via OAI, as well as control, cleanse and configure the data normalisation or transformation process.

The MST presents the results of each service as an OAI-PMH (set/subset) repository, which will make them available for harvesting by other XC software components (Drupal Toolkit), while also facilitating its output for other non-XC applications.

MST can be installed and configured in different ways. MST can be configured as a centre with services installed separately or one single installation. These options offer institutions the greatest possible number of alternatives in creating and housing their own individual services, employing the single installation to share them on a network.

In addition to the functionalities listed for the management of repositories, users, harvesting scheduling, the display of metadata and metadata transfer via OAI-PMH, the basic functionality of this module involves the handling and processing of metadata, including normalisation, transformation to a bespoke metadata system, the control of authorities and the aggregation of items/records.

This involves allowing users to manage services and the processing rules associated with them, as detailed in the following section.

3.2.1.2 Functional characteristics

The initial version of MST includes as its main functionality:

- Normalisation services (MARC and Dublin Core)
- Transformation services (MARC for XC and DC for XC)
- Authority control services

eXtensible Catalog

■ Aggregation

The figure below presents a clear figure of how MST operates in terms of the collection, processing and transformation of records:

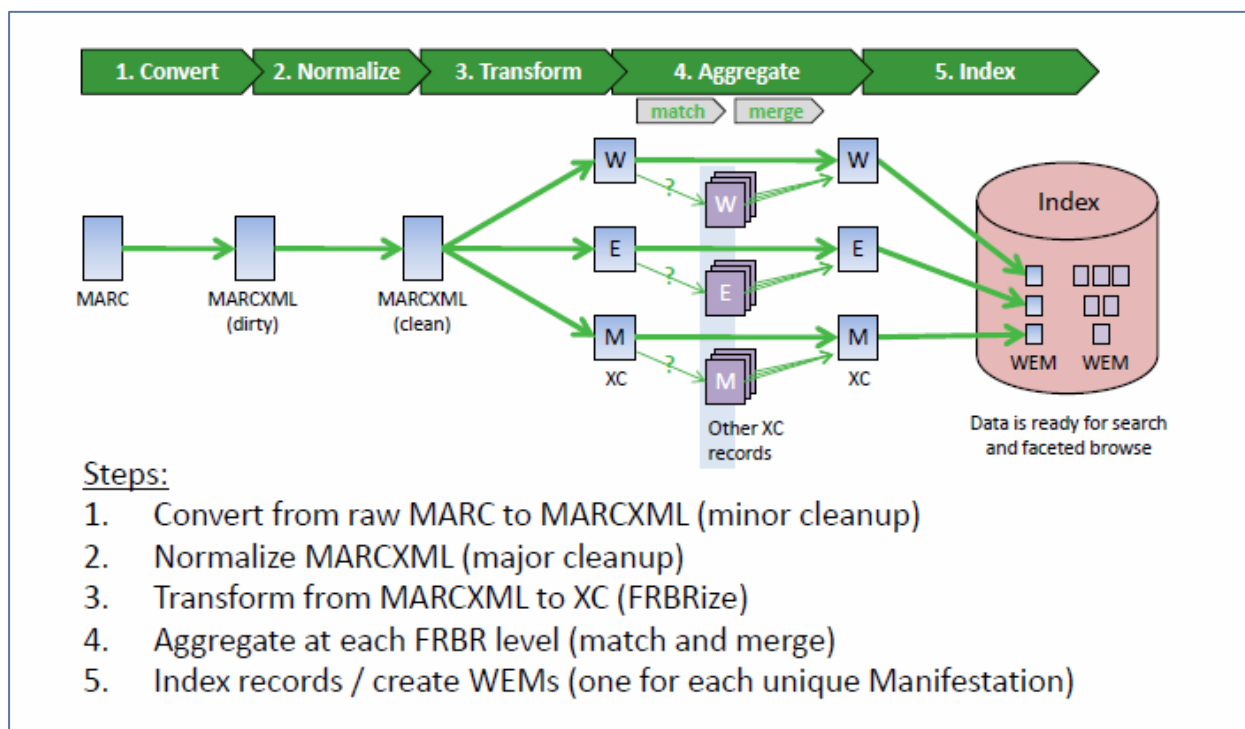


Figure 1: MST functionality

The two functionalities which have been developed and are in operation in the released version are Normalisation and Transformation.

- The **Normalisation** service processes the MARCXML records, resolving common problems and inconsistencies or the normalisation of non-standard data such as the normalisation of the definition of record types, valid values for field 035, normalisation dates, etc., in accordance with the processing rules included in the configurations. This service also prepares the metadata for the faceted interfaces, defining fields for local use (9XX).
- The **Transformation** service analyses the MARC records (bibliographical and holdings) and transforms them into the FRBRised XC schema. This transformation takes into consideration various aspects of RDA, in particular the distribution of RDA elements for FRBR Group 1 entities, the mapping of RDA elements onto MARC and an entity/relationship diagram developed for RDA. The objective was to define the MARC to XC correspondence in order to facilitate the mapping of MARC onto FRBR-RDA. This

eXtensible Catalog

service maps more than 100 MARC XML bibliographical record fields created with AACR2 cataloguing standards onto the XC schema with the influence of FRBR.

- The control of authorities and aggregation are two services which we have not been able to evaluate as they have not yet been released, although their definition indicates that:
 - Authority control: this will be the service which will work with the authority data harvested.
 - Aggregation: this will be the service responsible for grouping together records at the same level (expression, manifestation, etc.); it will handle relationships among records (FRBR entities), the automatic synchronisation of record changes at each FRBR level, and will propose scenarios for NON-MARC cases (application of RDA).

3.2.1.3 XC schema

The XC schema fulfils the need to work with data collected with OAI Toolkit or with MST Toolkit. The architecture of the tool presumes that data will be harvested in a wide range of formats, hence the choice of XML as the standard.

MST requires that the metadata be in XML and processes them to generate new records which will be available to allow other OAI servers to harvest them, although the metadata used in the interfaces both of MST and in Drupal Toolkit must be in the XC schema (faceted, FRBRised, etc.) in order to ensure that the metadata which libraries have already generated are as efficient as possible. The resulting XC schema can, at first glance, appear at first sight unnecessarily complex, involving theoretical aspects which we will not discuss in this report.

The work performed in creating the XC system is of considerable interest, with further information available in the eXtensible Catalog Schema Definition⁷, but it is not the purpose of this report to discuss the detail of the labels and mappings performed. Essentially, the XC schema itself deals with the following issues viewed as fundamental aspects of the project:

- Automated transformation of batches of MARC records into records based on FRBR (work, expression, etc.)
- One MARC record may return several output records
- Maintaining relationships among the initial records

The figure below presents a diagram of the relationships discussed:

⁷ <http://www.extensiblecatalog.org/sites/default/files/fulltext/XCSchemaProperties20090107.pdf>
http://xcmetadataservicestoolkit.googlecode.com/svn-history/r1229/branches/bens_perma_branch/docs/TransformationServiceDocumentation.docx

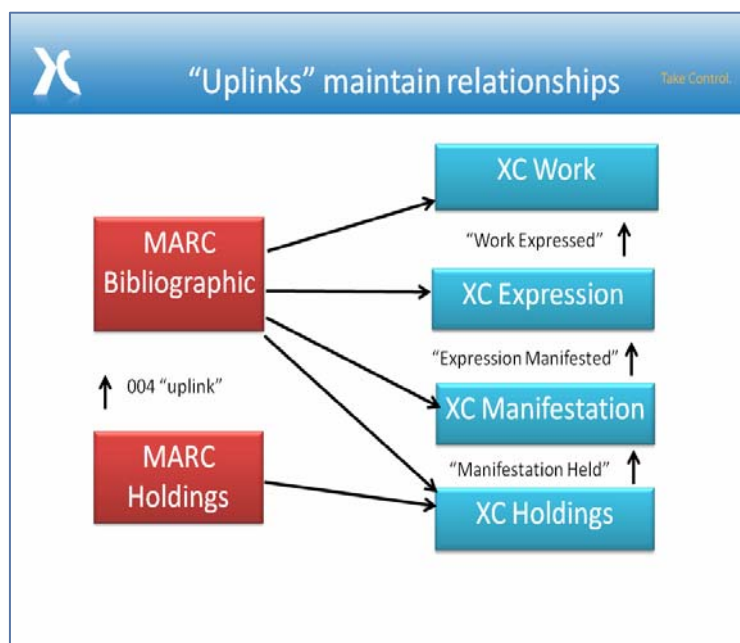


Figure 2: record transformation. XC schema

This schema is for internal use by the application, not for use by the cataloguers or metadata experts. The schema is designed as a container for metadata from various schemas, allowing the metadata to function both within XC and also to be shared with other systems. In other words, an XC user will not see a complete record in XC.

The creation of the XC schema also fulfils the need to maintain the richness of records, since the vast majority of the metadata harvested are drawn from library catalogues with MARC environments, even if they are combined with metadata from less rich schemas, such as Dublin Core. Meanwhile, within the XC schema some metadata are presented as facets, but both the labels of each facet and the facet definition itself can be configured. The development of the XC schema takes into consideration both the description schemas and standards which we will find in libraries in the future, such as RDA (Resource Description and Access), although these have not been 100% implemented, as developments are running in parallel. This influence is found in the FRBR structure.

What must in particular be borne in mind is that the schema has been optimised for XC functionality and not created for the purpose of general use.

This schema takes into consideration:

- Dublin Core terms
- RDA elements
- RDA roles
- MARC vocabularies
- And new elements known as XC elements

It comprises:

eXtensible Catalog

- 22 RDA elements
- 11 RDA roles

It maintains the granularity of MARC data in terms of:

- Frequency
- Series numbering
- Cartographical content coordinates
- Number of tracks (music)

3.2.1.4 Application in the BAGE Common Access Point

The MST or similar service is fundamental in the creation of the Union Catalogue or Common Access Point since:

- It allows for the programmed and automatic harvesting of metadata (from OAI toolkit or other OAI servers)
- It allows for the handling of different metadata schemas, employing MARCXML or DC as the basic structures
- It offers flexibility for defining the treatment and transformation of metadata (creating services and processing rules, as indicated) in terms of normalisation and transformation to the XC schema if the aim is to take advantage of the work performed or to another schema, while allowing this to be handled as a schema container.
- Control of authorities and aggregation are presented as two basic services in a Union Catalogue with quality data, as they offer processed data of real usefulness for the user and allow for "future" value-added services such as the implementation of faceted systems, RDA, FRBR and linked data.

3.2.2 Drupal Toolkit

3.2.2.1 Objective and definition

Drupal Toolkit manages the content and functionality of the website in one simple package with Drupal (as content manager for the portal/consultation interface). This thus involves the management and creation of the interface which will be handled and operated by the end user.

This toolkit integrates the metadata, ILMS services such as the circulation and content of the library website, in one single interface with enriched functionalities (value-added).

eXtensible Catalog

The Drupal Toolkit includes:

- A single interface for the search for data drawn from different integrated management systems, digital repositories and/or the content of the website collection.
- Facets: It includes faceted search functionalities.
- Web services: It offers powerful tools to create web applications connected with the content of integrated library management systems and their functionality.
- Circulation: It is fully integrated with the ILMS to display the circulation status and to file requests with the circulation system.
- It functions with authentication systems, such as the ILMS database and LDAP servers.

The software repository code for the Drupal Toolkit is <http://drupal.org/project/xc>

The project documentation pages are available at <http://drupal.org/node/499770>

Drupal Toolkit API documentation is available at <http://api.xcproject.org/drupal>

3.2.2.2 Functional characteristics

The Drupal Toolkit is a modular system responsible for the creation and manipulation of metadata in accordance with the XC schema format within the Drupal content management system through:

- Definition of the metadata.
- Importation and exportation of metadata.
- Storage and retrieval of metadata.
- Metadata search.
- Expansion and modification of metadata.
- Control of user access to metadata.
- Generation of fully customisable XHTML and CSS output templates.

Some modules form part of the kit but can be used independently. They are:

- OAI harvester.
- NCIP integration.
- Integration of Syndetics solution services.
- Rewriting of URL addresses in the EZProxy server.

eXtensible Catalog

All these basic functionalities will be handled by XC, defining modules and extensions for Drupal from the system administration.

3.2.2.3 Application in the BAGE Common Access Point

Drupal Toolkit has all the extensions and modules prepared for its integration with possible ILMs, repositories and sources. Drupal, furthermore, as a content management tool, already has in place, for example, a Millennium Integration module, which is very widely used and allows for integration with this ILS, along with considerable participation and projects run by libraries on two main websites, within the Drupal "groups", that dedicated to "Libraries" and outside the community, "drupalib".

Although, as with all the components of XC it may either be used or not, its use is highly recommended, or at least the development of a similar, more basic system based on the same technology, in order to take advantage of the considerable volume of modules or value-added services which can subsequently be implemented and which have already been developed. One could begin with query forms and basic results and gradually incorporate other modules as the Common Access Point achieves progressive consolidation as a project and service.

3.2.3 OAI Toolkit

3.2.3.1 Objective and definition

This Extensible Catalog module offers the possibility of adding an OAI-PMH server to those ILMs which do not have such a utility, serving to reveal records to other harvesters and also to other XC models, allowing them to be gathered by them and processed by their own tools. This module is not essential for the ILM, as it already has its own OAI server, namely Koha.

The tool is implemented with Java technology, so it requires a Web applications server which supports Java Servlets. The most commonly used system, on which functionality was developed and tested, is Tomcat.

3.2.3.2 Functional characteristics

In order to be able to export records from an ILM to OAI Toolkit, a file must be generated in MARC21 format (ISO 2709 or MARCXML). The types of MARC21 records must belong to the Authorities, Bibliographical or Holdings group.

There are three fundamental steps in order to import data into the OAI Toolkit: conversion, modification and the loading of data. They can be applied jointly with a script, or individually in order to generate intermediate results to be processed as required.

Conversion will always be from MARC21 to MARCXML, meaning that one knows if his record file is valid and follows the MARC21 schemas.

Modification by means of transformation sheets serves to adapt our own MARC21 records (9XX) to other fields or character coding systems required by the Toolkit.

Data loading processes the MARCXML records and adds them to the internal database. This process can handle insertion, update and deletion operations. The unique record identifier is harvested using the Control Number (001) field, and if there is a Marc Repository Code (003) it is linked with this.

Below is a diagram illustrating the functionality of the terms described above:

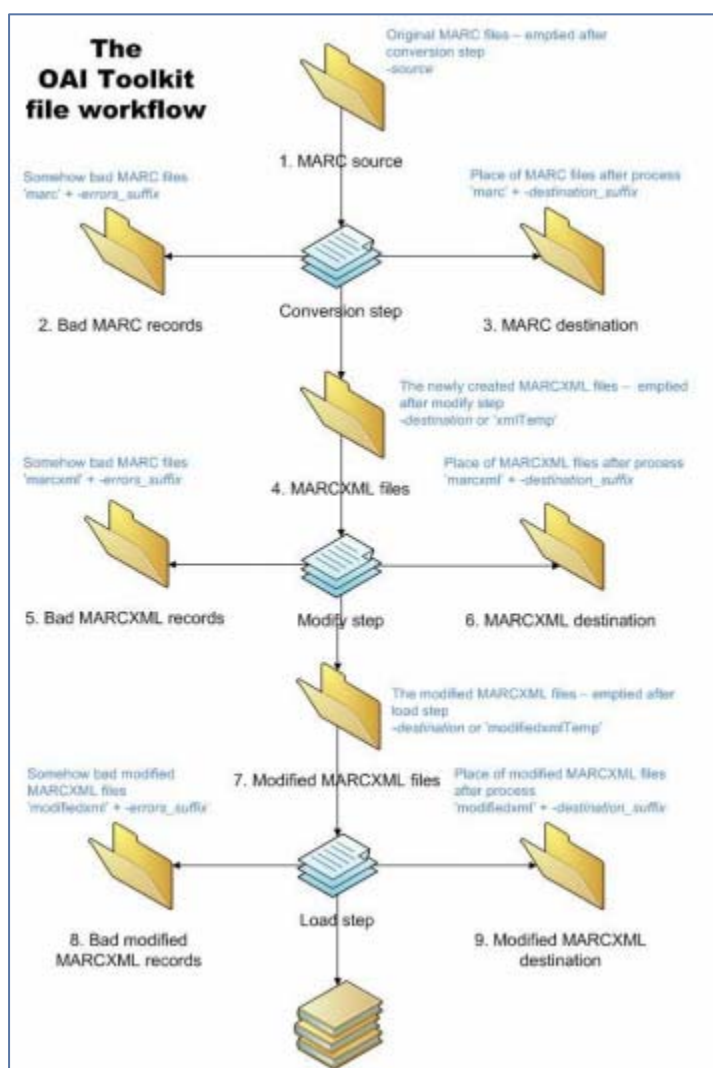


Figure 3: OAI Toolkit. Workflow

Once they have been inserted, the Toolkit can provide MARCXML records, or they can be duly transformed into OAI_DC (Dublin Core), OAI_MARC or MODS.

It should be pointed out that the Toolkit employs two databases: MySQL as a relational database and Solr as a textual database. MySQL must be installed because, when the records are displayed, the Resumption Token (paginated) of the OAI protocol are handled from MySQL. There are three storage options:

- Everything to MySQL: for searches in major repositories this may generate a bottleneck because of its slowness and processing burden.
- Storing the XML representation in Solr and the XML metadata in MySQL.

eXtensible Catalog

- Storing all metadata in Solr.

The third option is recommended, as it yields the highest performance levels.

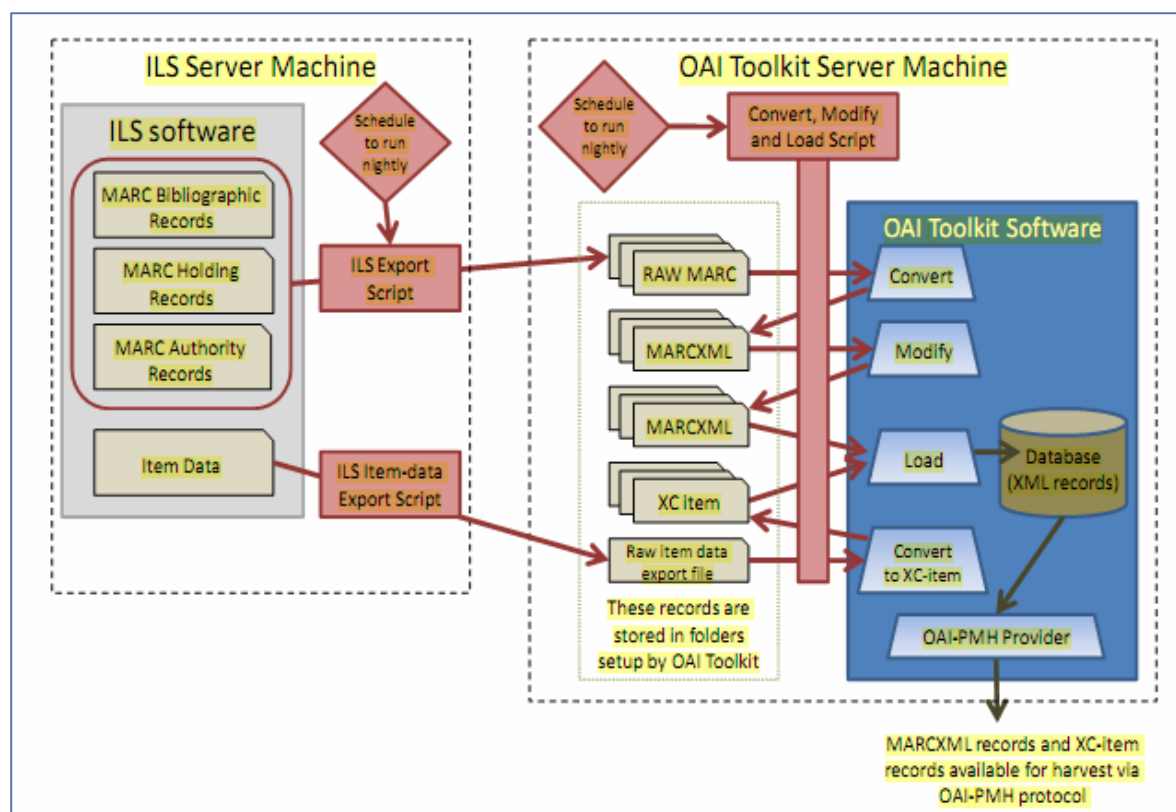


Figure 4: General functional Toolkit schema

The toolkit has a web interface which can be used to manage and configure the OAI-PMH server:

3.2.3.3 Application in the BAGE Common Access Point

This is a fundamental module within the project, as it resolves problems such as the OAI servers in those ILSs which do not have them and transformation from MARC21 to OAI_DC (Dublin Core), OAI_MARC or MODS.

Even for those which do have them, it may be desirable to incorporate them within the OAI Toolkit in order to standardise metadata.

The problem we find is that it needs the records incorporated to be MARC21, and we may encounter, as generic groups:

- Libraries which are not automated.
- Libraries with computerised catalogues which do not use MARC.
- Automated libraries which use IBERMARC.

eXtensible Catalog

In the case of those which use IBERMARC, the records can be modified before being loaded, using xsl sheets for adaptation to the MARC21slim_rochester.xsd schema.

3.2.4 NCIP Toolkit

3.2.4.1 Objective and definition

This module of the XC package is designed to serve as an interface between the NCIP compatible client and an ILMS. Its functionalities are authentication, circulation record searches to establish their status, and circulation operations.

The NCIP protocol is implemented in order to transmit information by means of the HTTP or HTTPS protocol in XML format. It is a protocol which does not require session status to be maintained.

The tool is implemented with Java technology and thus requires a Web applications server which supports Java Servlets. The most commonly used system, on which functionality was developed and tested, is Tomcat. It is possible to have more than one NCIP Toolkit on the same web server in order to satisfy the needs of more than one ILMS.

3.2.4.2 Functional characteristics

A plug-in compatible with the NCIP Toolkit is required to serve as the intermediary between the ILMS and the NCIP Toolkit. This plug-in is responsible for gathering the Toolkit requests, applying the relevant queries to the ILMS's own database (users, circulation, cataloguing, etc.), before ultimately returning the information to the Toolkit for it to be transformed, normalised and properly sent to the NCIP client.

Client requests to the NCIP Toolkit are coded by means of the HTTP POST action, although GET may be employed on a limited basis. The tool maintains authenticated user session information as many ILMSs specify this, meaning that the NCIP client must support cookies. The NCIP Toolkit will otherwise not maintain information about the user between requests. If, for example, a request is made for an NCIP operation with no specific authentication provided, the NCIP Toolkit will employ the logged-on user information for this unless a different session username is provided in the protocol.

The requests which can be performed in Toolkit are fairly common and necessary in the circulation of any ILMS, and are implemented in the NCIP protocol:

- Authenticate User.
- Lookup Item.
- Lookup User.
- Lookup Version.
- Request Item.
- Recall Item.

eXtensible Catalog

- Renew Item.
- Cancel Request Item.
- Cancel Recall Item.

As a supplement to the above, the Toolkit can perform four more exclusive operations, in other words, those not specified in the NCIP protocol:

- XC Lookup User, similar to that compatible with NCIP, while allowing more information to be requested.
- XC Get Availability.
- XC Open URL Request Item.
- XC Open URL Renew Item Services.
- Any request for an operation or service not listed above will result in an error message.

This is a basic tool for the management of users and the circulation of items in accordance with the latest standards, and allows for the possibility that the ILMs, irrespective of their data structure and functionalities (although this is not in fact the case as they do not in themselves need to support the NCIP protocol, though they must have within their data structure all the information required by the NCIP) can be interoperable with other ILMs using the NCIP protocol. The greatest disadvantage is the need for a plug-in which (normally this will need to be implemented if this is not been performed by the community) can communicate between the ILM and the NCIP Toolkit.

3.2.4.3 Application in the BAGE Common Access Point

The implementation of NCIP offers particularly appealing and necessary options for the integration of physical libraries (where there is a need for circulation control) within the Common Access Point. This module, therefore, is presented as a basic element in the provision of a holding access service, more than the creation of a catalogue or Common Access Point.

For the implementation of this protocol, one must overcome not only technical and technological issues, such as the drivers and applications allowing for communication with each of the ILMs in use, but must also have defined other issues such as the use (public or private) of collections, and the conditions for access to certain holdings, the criteria or characteristics of the relationships between the libraries which make up the catalogue in terms of inter-library loans, and all issues dictating physical access to holdings. These aspects do not lie within the scope of this report.

Design of the Common Access Point

The PCU is made up of three sections operating jointly:

- The individual library catalogues and repositories.
- This is identified as the Union Catalogue (in the centre of the figure, in yellow).
- The various consultation interfaces for users and other systems.

4.1.1 The individual library catalogues

The integrated library management systems (ILMS) and the repositories of digital libraries reveal their metadata via OAI-PMH to be harvested in the Union Catalogue. Those ILMSs which include OAI provider functionality on a native basis will reveal their metadata directly; those which do not have this will employ an OAI layer, for which we will use the first XC tool, the OAI Toolkit.

Likewise, the NCIP Toolkit, the second XC component, offers an interface for authentication requests, real-time status reports and availability and requests for loans for which the ILMSs do not have a native NCIP facility.

4.1.2 The Union Catalogue

The metadata MARC21 and DC (and, potentially any other schema), are harvested by Metadata Services Toolkit, a powerful XC Project tool for the handling of metadata, comprising a suite of automatic, configurable and modular services.

Metadata from different sources vary widely: versions of MARC21, non-standard DC qualifications, local cataloguing practices, etc. These are handled by means of the normalisation service, and the metadata are then transformed into a complex schema which includes FRBR and RDA (transformation service), with the possibility to subsequently employ other services such as authorities control, or aggregation for all aspects regarding the gathering of various connected records of the different FRBR identities.

Drupal Toolkit is the tool used to create the different consultation interfaces with the facet richness needed for the prior handling of metadata. It provides powerful tools to build applications establishing close links with the content and functionalities of the ILMS.

For example, it integrates with the ILMSs to display the real-time loan and availability status and dispatches loan requests (via NCIP).

4.1.3 User and systems interfaces

This includes various possible interfaces with users or with other applications which reuse the data to create value-added services.

4.2 Methodological requirements

The Working Group has worked with solutions fulfilling the following requirements:

- **Interoperability:** Maximise interoperability, understood as the capacity of information systems and the procedures which they support to share data and allow for the exchange of information and knowledge among them as efficiently as possible. This is achieved by employing open standards at the organisational, semantic and technical levels. It will in particular impact the interoperability of the Common Access Point with other information systems: of the AGE itself, such as the ISO 2146 directory ⁸, or external systems, such as all those tied to the movement of Linked Open Data.
- **Open sources:** There is an agreement that the development of open source software is the best way to encourage interoperability. The implementation of open standards, as the basis for interoperability, is achieved to a greater extent through the use of open source software. In this regard the indications of the National Interoperability Schema are followed, as discussed in section 3.5.
- **Software control:** Maintain the administration and control over the software public and within the user community in order to decide when, how and who will implement new functionalities, error correction, etc.

eXtensible Catalog, as seen earlier, can handle these principles.

4.3 Documentary aspects

The Common Access Point will employ documentary analysis standards and practices including the latest trends. Their implementation will gradually be introduced in various phases of the project, as described in detail in section 5 of this report.

The ultimate aim will be to put in place a metadata schema: **The XC schema or an adaptation of it**, built in accordance with the theoretical model provided by the Singapore Framework for Dublin Core Application Profiles⁹.

⁸ *ISO 2146:2010 Institutions and Professionals Directory Project*, information about the project at <http://www.mcu.es/bibliotecas/CE/RecursosProfesionales/DirectorioISO2146.html>

⁹ Karen Coyle and Thomas Baker, *Guidelines for Dublin Core Application Profiles*, 2009 <http://dublincore.org/documents/profile-guidelines/>

Design of the Common Access Point

Within the provision of metadata promoted by the Singapore Framework, a native metadata schema is built for the PCU project, employing those elements of the relevant related domains.

4.3.1 Legacy formats: MARC, DC

MARC21 is the format which will mainly be used for the records harvested. The various versions and local practices will be normalised following harvesting, with the possibility of the adaptation of the MST Toolkit normalisation services.

As mentioned earlier, the need to work with MARC21 for harvesting forces the libraries working with ILMS and employing IBERMARC to convert their records prior to harvesting. The appropriate OAI Toolkit functionality will be available for this purpose.

The DC employed in digital libraries and repositories will likewise be normalised.

4.3.2 Metadata input control

The PCU will have in place a tool to validate the input of MARC 21 and DC records in accordance with the defined schemas. This tool will have a web interface for inspection of the participating libraries.

4.3.3 Domains employing the XC schema

The domain model is provided by the FRBRs¹⁰. The record typology in accordance with Group 1 FRBR entities will be drawn from them: Work, Expression, Manifestation, Item; with one native aspect added: Holdings.

XC Schema employs 22 elements and 11 roles of those defined in Resource Description and Access (RDA).

Meanwhile, the XC schema preserves the level of detail of the original MARC record descriptions, saving this information as explained in section [3.2.1.3](#).

4.4 Technological aspects

Section 3 of this report has already analysed in detail the eXtensible Catalog software which lies at the heart of the Common Access Point. Here we simply set out the main technical standards employed:

¹⁰ Functional bibliographical record requirements: final report <http://hdl.handle.net/10421/422>

Design of the Common Access Point

- **OAI-PMH:** The OAI-PMH protocol¹¹ (Open Archives Initiative - Protocol for Metadata Harvesting) has become a part of the web infrastructure because of its relative simplicity and implementation. It is, therefore, the chosen option for the transfer of records in the various system components.
- **NCIP:** NCIP - NISO Circulation Interchange Protocol -is the standard handling the need for interoperability among the different circulation applications, interlibrary loans, self-service, etc. in the library world. The PCU will cover this at a subsequent phase of development in order to build value-added services.
- **Linked Data:** The Common Access Point will be in a position, at a subsequent phase, to handle integration within the semantic web using Linked Data mechanisms. There are plans for a specific service for conversion from XC Schema to RDF, through the development of Metadata Service Toolkit.

4.5 Legal Aspects

The Common Access Point is covered by a national and European regulatory framework.

4.5.1 Royal Decree 1572/2007, on the regulation of General State Authority and public body library coordination bodies.

Royal Decree 1572/2007 governs the coordination bodies of the General State Authority's libraries and its public agencies, establishing their aim, together with those of the BAGE organisational and technical coordination structures, as follows:

"To promote the establishment of a Common Access Point which, by means of a Union Catalogue or equivalent electronically accessible system, will allow for joint consultation of General State Authority and public body library collections".

4.5.2 Electronic Access of Citizens to Public Services Act 11/2007

This Act, naturally, applies in full to the Common Access Point, although we will here highlight certain principles, set out in Article 4, which are of particular impact:

- *The principle of **accessibility** to information and services via electronic means.*

¹¹ "Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0"
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

Design of the Common Access Point

- "The principle of **cooperation** in the usage of electronic resources by the public authorities, in order to guarantee interoperability".
- "The principle of **proportionality**, on the basis of which appropriate security measures and guarantees will only be demanded in accordance with the nature and circumstance of the various procedures and actions involved".
- Principle of **technological neutrality**, demanding that the public authorities employ open standards.
- Principle of administrative simplification.

4.5.3 Royal Decree 4/2010, the National Interoperability Schema

The National Interoperability Schema, within the context of electronic administration, aims among other goals to ensure appropriate levels of interoperability among systems.

In accordance with these standards, the main means to achieve interoperability is the use of open standards, the implementation of which is achieved to a great extent through the use of open source software. Article 16 describes this form of licensing applications, which will apply to the Common Access Point.

The licensing conditions for applications and the associated documentation, and other objects of information where public authorities hold the intellectual property rights, and which they could make available to other public authorities and citizens, without payment and without the need for an agreement, shall take into consideration the fact that the pursued aim is their exploitation and reuse, and protection against exclusive third-party appropriation, under conditions releasing the assignor from any liability as a result of possible misuse by the assignee, along with non-obligation to provide technical support or maintenance on the part of the assignor, or any compensation in the event of errors in the application.

The Common Access Point will be distributed with the European Union Public Licence, as expressly recommended in the Royal Decree (Article 16.4)

4.5.4 Royal Decree 3/2010, National Security Schema

This Royal Decree governs the National Security Schema and security policy in the use of electronic resources for access by citizens to public services.

Article 43.1: *"The category of an information system in security terms shall strike a balance between the importance of the information handled, the services provided and the security efforts required, in accordance with the risks to which it is exposed, based on a principle of proportionality".*

The nature of the data fed into the Common Access Point, or which it provides to users, does not represent any substantial risk, and therefore does not need to be covered by the same security protocols as other ministerial data with which it often shares information technology systems.

Design of the Common Access Point

4.6 Key Aspects

Within the PCU project, in accordance with the functionalities described in the previous sections, there is a series of key aspects or issues which the project must resolve as an initial step:

- Identification of sources and types of metadata.
- Metadata harvesting/extraction.
- Processing and storage of metadata.
- Consultation and Service.
- BAGE coordination.

4.6.1 Identification of sources and types of metadata

The Common Access Point to be created draws on a wide range of metadata sources and types, as set out in the automation status report previously drawn up.

The sources on which the Common Access Point will draw are the catalogues of each of the libraries. In this regard there is fairly substantial information in the automation study produced prior to this report. The study reveals valuable data such as the management systems employed and the number of records estimated for each, which will serve as the source of the data required by the Common Access Point.

MANAGEMENT SYSTEM	Total
none	1,355,918
Absysnet	1,288,631
Absys	1,156,753
Aleph 500	1,065,685
SirsiDynix Unicorn	502,585
DIGIBIB	352,314
SABINI	210,921
Inmagic DB/TextWork	120,671
Basis-Techlib	120,253
Own application	89,306
Overall total	6,263,037

It is important to take into consideration the ILMs with which we will need to work, since each of these operate separately not only in terms of metadata schemas (as we will see below), but also in terms of local internal system management and data storage, such as:

Design of the Common Access Point

- Internal handling of authority records and their relationship with bibliographical records.
- Handling of bibliographical and item records (not only conditioned by the management system, but also by local policy: each item has a different record and work is not performed at the bibliographical and item level, but at the individual record level for each item).
- Management of collections of periodical publications.
- Handling and management of digital objects (many of them tied to bibliographical records, others to items).

One must also take into consideration the possibilities which exist within each system in terms of data access (as we will detail in the following section on harvesting), along with the export options.

As for metadata types or schemas, there is not such a great variety as in terms of systems. Most of them work with IBERMARC, and a lesser percentage with MARC21.

The use of MARC21 is a requirement imposed by the OAI Toolkit, as discussed in section [3.2.3.2](#), as normalisation and transformation have already been developed exclusively from MARC21 to MARCXML. Libraries will need to convert in order to be incorporated within the Union Catalogue, at least as follows:

- Bibliographical MARC21
- Holdings MARC21 (both monographs and periodicals)

The type of metadata or schemas employed does not initially represent any major problem, since a transformation sheet from IBERMARC to MARC21 has already been drawn up.

4.6.2 Metadata harvesting and extraction

The method by which records are incorporated within the Common Access Point is another of the project's key aspects. As already suggested in the previous section, one further key aspect is the integration capacity of records in terms of the applications employed. It has already been stated that XC employs OAI-PMH as its Interchange protocol, as this is a standard protocol for the automatic harvesting of metadata, one used by most applications already working with library metadata, and is not a complex protocol.

Those systems which do not have in place an OAI module and which can be harvested will need to have the OAI Toolkit Module and manually load records into MARC21 for them to be converted into MARCXML, thereby making them available for harvesting (using MST, or any other system which may be developed).

The figure below sets out the options which could be considered in order to facilitate automatic metadata harvesting:

Design of the Common Access Point

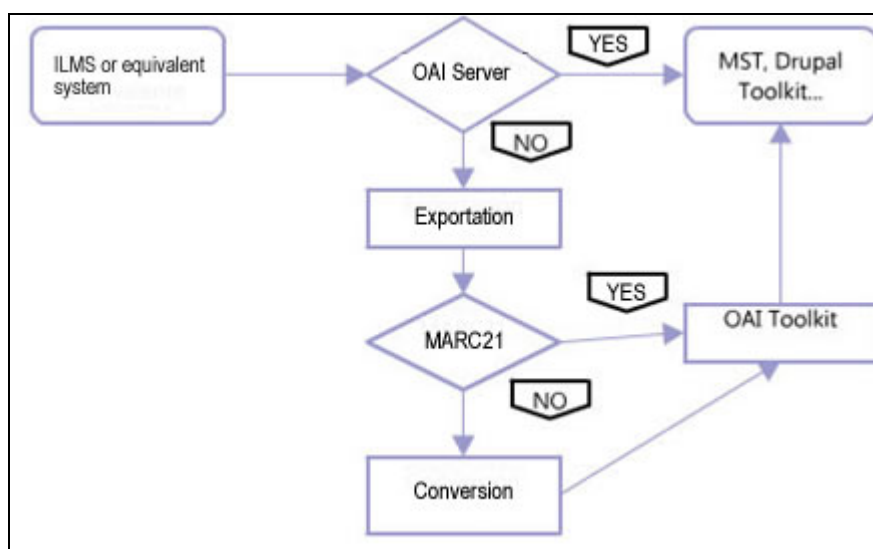


Figure 6: Metadata harvesting automation

Bearing in mind the situation of the AGE libraries and the characteristics of XC, criteria could be established for the incorporation of sources within the Common Access Point in accordance with the characteristics of the systems and the need for intervention for the automatic harvesting of records, as we attempt to set out in the following figure:

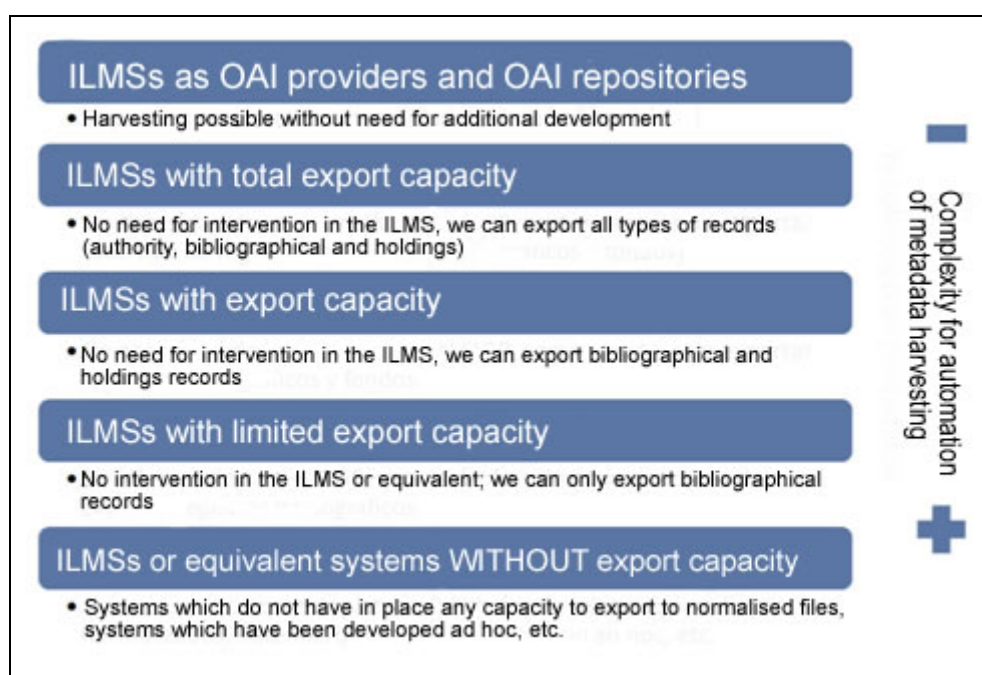


Figure 7: Complexity for the automation of metadata harvesting

Design of the Common Access Point

4.6.3 Processing and storage of metadata

The eXtensible Catalog model requires that, following harvesting, the data be stored. This storage and the subsequent processing depends on the objectives being pursued, as may be seen below.

All the XC models have in place a database used to store the data, with a series of functionalities having been developed allowing each one to process the stored data in different ways, such as:

- OAI Toolkit: loading, storage and conversion from MARC21 to MARCXML, provision as OAI server.
- MST: Metadata harvesting, storage and processing by different services (Normalisation and Transformation) initially, and availability for Drupal Toolkit or some other tool (OAI-PMH).
- Drupal Toolkit: Harvesting and storage.

These functionalities underpin the idea of being able to work independently with each of the modules or toolkits, or with the complete tool, depending on the aims of each project. For example, one could use solely Drupal Toolkit to harvest and store data and design the consultation interface, but this would not allow for application of the complete faceted search development, because the XC schema would not have been used in the MST. Nor would there be any processing of the harvested metadata, meaning that the data presented would be "raw data". Meanwhile, one could use the OAI Toolkit and MST to harvest and process metadata, develop another consultation interface, etc.

The treatment offered by XC, above all in MST, has major potential based not only on the concepts, equivalences and the development which has already been performed, but also because it resolves many of the problems of ANY Union Catalogue, such as normalisation, transformation (irrespective of the type or model of catalogue chosen) and because this task is automated. Meanwhile, MST is designed in a very modular manner, thereby establishing the repository for which the service is to be applied (normalisation, for example) and the processing rules which will need to be applied. As described in the corresponding section, the XC system has already followed a lengthy process in terms of definition, equivalences and functionality development, and proves itself to be an appealing schema for application to the Common Access Point.

The type of metadata processing performed as part of the Common Access Point project can not only be gradually improved, but also requires that decisions be taken prior to design, new services or the adoption of services and rules undertaken within the XC project. These issues can be redefined in configurations of the MST.

The handling of the metadata which have been gathered is a key issue to be defined at the outset of the project. One could even consider the possibility of, at an initial phase, working without metadata processing, and focusing simply on harvesting and storage. Following storage, in a second phase, analysis and/or improvement of the application of services and schemas defined in XC shall be considered.

Design of the Common Access Point

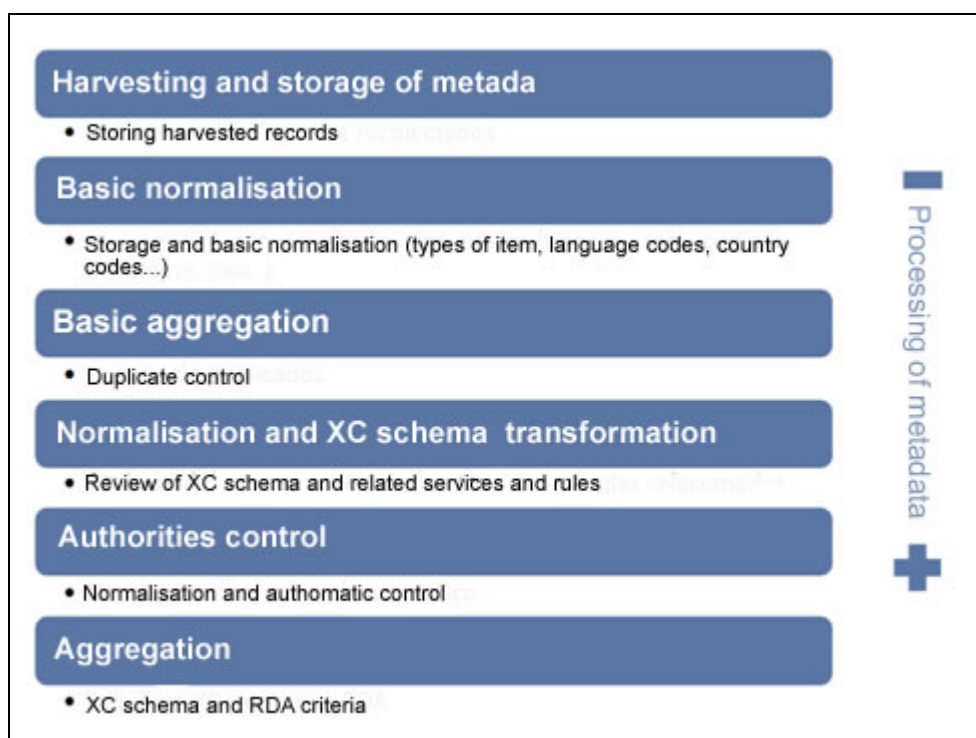


Figure 8: Processing of metadata

4.6.4 Creation of the Common Access Point: Service

We understand creation as the development of the interfaces which will allow for consultation, although the creation of this catalogue refers to all the key issues mentioned above.

The catalogue or Common Access Point refers not only to the harvesting and storage of metadata, but must also be electronically accessible and allow for the joint consultation of all the metadata harvested.

As seen earlier, XC employs the DRUPAL content manager to create these consultation interfaces. It also draws on previously developed search templates, faceted search, etc., and is optimised for treatment and communication with other XC modules. Drupal handles all this as modules and extensions which can be enabled in accordance with the maturity of the project.

This interface must take into account all **services** which the Common Access Point will ultimately offer, and therefore does not eliminate the OPACs or individual consultation systems for each source, while the level of integration can be gradually increased over time. In other words, by way of example, the interfaces initially developed could simply be for consultation and results, and from this point onwards each record would then lead to the OPAC/service portal of each of the sources for other services, such as reservations. As the project is gradually defined and developed, at subsequent stages reservations could be a service provided at the point of consultation, with the integration of both systems being entirely transparent to the user, employing the NCIP module and protocol, allowing user and local circulation information to be provided.

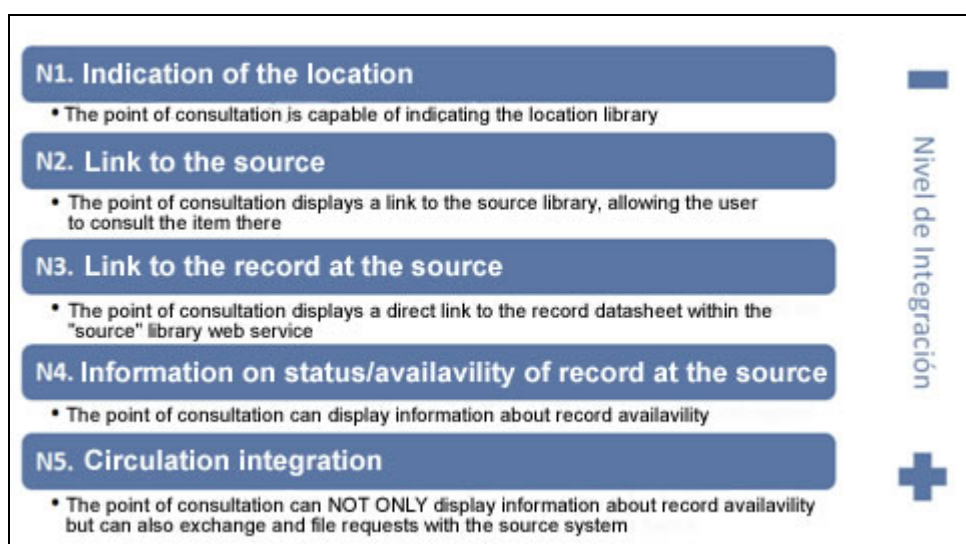
The issues which we identify as key in the development of these interfaces are:

Design of the Common Access Point

- Interface design and creation: Drupal Toolkit as content manager (or otherwise), corporate identity and image, etc.
- Services and functionality: definition of the services to be offered from the Common Access Point and the implications of this at the project development level. If metadata are to be harvested in the initial phase and an analysis begun for their handling in the catalogue, the only initial service would be consultation, and would meanwhile be subject to the limitations that the data would not been processed or transformed.

Regarding both issues it is essential to clearly establish the level of integration which we are willing / able to offer. As already mentioned, XC allows for "independence" of the Common Access Point from other "source" systems, not only if all modules are implemented, but also with the definition of the roles and services to be offered. The level of integration will depend on both issues (complete XC and definition of roles and services), but also on the integration capacity demonstrated by the source systems.

For example, the figure below illustrates the level of integration in terms of the location of a given record:



One could initially consider a low level of integration (bearing in mind the fact that many sources do not even have a web base) with integration gradually being improved as the project matures, not only through its own development but also through the interests and availability of the sources.

5. Planning

5.1 PCU working lines

The creation of the PCU Common Access Point is an ambitious project which depends not only on the implementation of more or less complex technologies, but also on metadata schemas and standards, legal aspects regarding the involvement of libraries, etc. In this section we attempt to map out the working lines which will need to be present throughout every phase of the project.

For development of the Common Access Point three major areas have been identified, comprising blocks or working lines as represented in the following figure

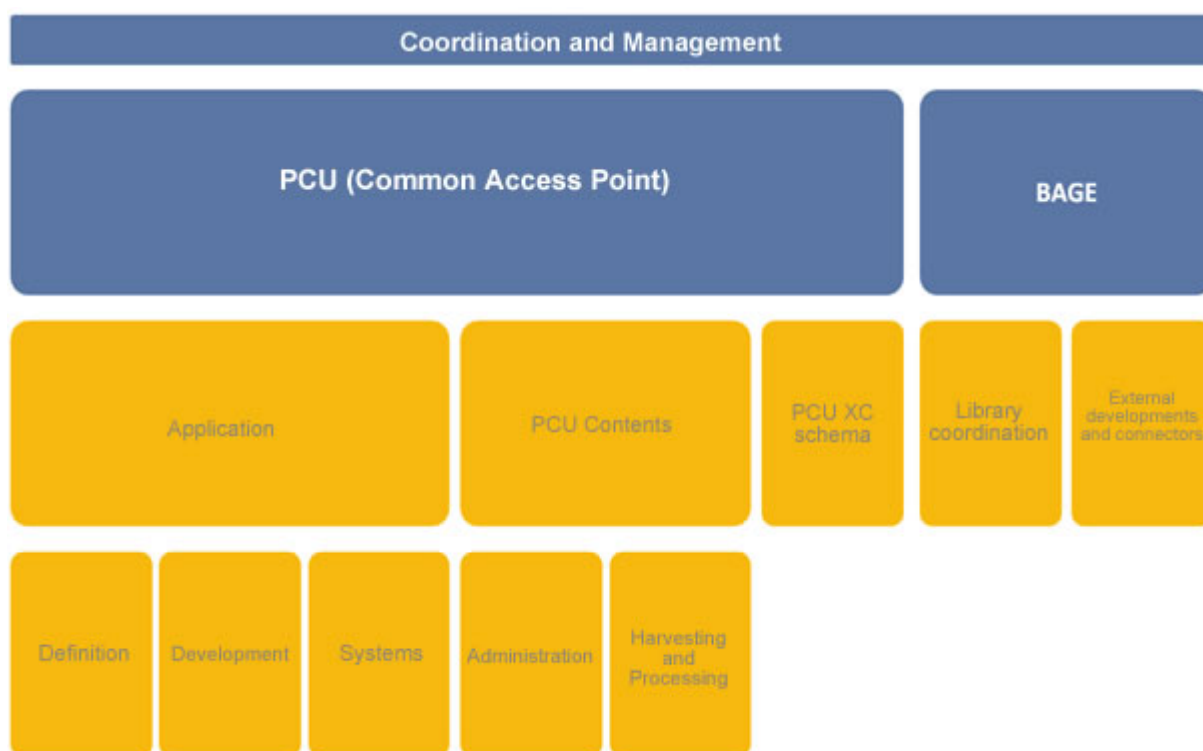


Figure 9: PCU WORKING LINES

Planning

The table below provides details and codes for ease of nomenclature throughout the report:

Area	Line	x ¹²	Description	
Common Access Point	A	PCU application	A1	Definition, analysis and monitoring
			A2	Systems
			A3	Development (New functionalities or modification of those in existence)
	B	PCU content and administration	B1	Configuration, parameterisation
			B2	Harvesting
			B3	Processing
C	Metadata schemas	C1	Definition and modification of XC schema	
BAGEs	D	Library Coordination	D1	Issues regarding library incorporation
	E	Connectors and applications outside the PCU	E1	Analyses and development of services and applications outside the PCU for connectivity with individual systems
Project Management and Coordination	F	Project monitoring and coordination	F1	Project management and leadership
	G	Dissemination	G1	
	H	Assessment and improvement	H1	

Naturally, the three areas are completely interrelated and depend on the phases and maturity of the project, along with the service levels. We below develop in further detail each of the areas and working lines identified as being required within all the phases, although they may be modified or expanded depending on the level of service, the technology, etc.

5.1.1 AREA 1: Platform for the PCU Common Access Point

The working lines identified in terms of development and creation of the PCU are:

A. PCU application: This refers to all platform development tasks and information technology tools which will give it its functionality.

A.1 Definition, analysis and monitoring: This will be the working line dedicated to the analysis and definition of the tool requirements for the PCU, along with follow-up on these developments, bearing in mind the needs defined in other working areas, such as **C**, and the progress and evolution of the XC to itself.

A.2 Systems: This deals with maintenance of the information technology working environments within which the development will take place, along with the analysis and definition of requirements for the definitive environment and PCU production. Some of the tasks identified:

A.2.1 Installation and configuration in development.

A.2.2 Installation and configuration in production.

¹² The codes have been employed in the working lines, and also the various sub-lines to facilitate the corresponding nomenclature process in this report and to highlight the relationships among them.

Planning

A.2.3 Systems maintenance.

A.3 PCU Development (New functionalities or modification of those in existence): This deals with information technology development and update tasks which will be required in order to adapt to the phases of the project, along with the resolution of bugs and/or new developments as defined in **A.1**. These tasks could be subdivided in accordance with the XC components and other PCU design elements (based on the schema set out in figure 6).

A.3.1 Developments in OAI Toolkit

1. ILMs connectors with OAI Toolkit,
2. IBERMARC – MARC21 converter

A.3.2 Metadata Service Toolkit

1. Service and Normalisation: Modification and adjustment (MARC and DC)
2. Transformation Service: Modification and adjustment (MARC and DC)
3. Service Aggregation - Modification and Adjustment
4. Authorities Control Service - modification and adjustment
5. XC to RDF

A.3.3 Drupal Toolkit

1. Web consultation interface
2. Mobile consultation interface
3. Connectivity (OAI-PMH, SRU, Linked Open Data)

A.3.4 NCIP Toolkit

1. ILMs connectors

A.3.5 Developments in the XC Core (contribution to eXtensible Catalog Organization)

A.3.6 Development of a web tool for *metadata input control*, for data clean-up and quality

A.3.7 Developments for the enrichment of records

A.3.8 Real-time integration with other information sources (e.g. SGCB directory)

B. PCU content and administration: We apply the term PCU content and administration to those working lines which, from the outset of the project, need to focus on administration and maintenance of the Common Access Point. As development is gradually expanded, along with the level of service and functionalities of the PCU, this maintenance and configuration will involve substantially different tasks, but will at all times focus on one of the three lines defined below:

B.1 Configuration: With the initial XC infrastructure or with the minimal developments at an initial stage, the system will need a working group and technical staff dedicated not only to registration and configurations of repositories and collections, but also to the definition of schemas, services and rules from the tool configuration archives.

B.2 Harvesting: Input data control and harvesting, as in the case of configurations, will involve an increased workload as repositories and other sites acting as record sources are gradually included. This working line involves the monitoring and optimisation of harvesting, along with the improvement and administration of the harvested data storage.

Planning

B.3 Processing: As in the case of harvesting, a working line is opened up to process the metadata once they have been harvested and stored. This working line refers not only to definition and configuration of MST files and services, but also their monitoring, and is entirely linked to the metadata schema area **C**.

C. Metadata schemas: Independently of the inherent tasks involved in PCU configuration and parameterisation, there needs to be a working line to improve and define the metadata schema employed. As mentioned earlier in this report, the XC Schema is an inherent schema of the XC application, but this project will from the outset need to have in place experts both in the schemas used by the sources (Ibermarc, Marc21, oai_dc) and the standards and new schemas (FRBR, RDA, ESE...) with which we will be working. These tasks could initially be classified as:

C.1 Generation of the schema **MARC21slim_pcubage.xsd** on the basis of **MARC21slim_rochester.xsd**. This will be employed to validate the MARC21 records fed into the system.

C.2 Generation of the **XCPCU** schema based on the XC schema. This will be the actual schema for metadata within the PCU.

5.1.2 AREA 2: Participating libraries and specific needs

In parallel to the technological and technical platform consolidation and development work which will provide a space for the records from the AGE libraries and any others wishing to take part, a series of tasks will be needed to coordinate and analyse the state of the participating libraries for their incorporation, along with work on the needs to be resolved by each of them before they can join the project. Given all the above, we initially establish two separate working lines within this area.

D. Library Coordination: It is essential not only that the participating libraries be coordinated, along with the conditions for participation, but also that there be dissemination and cooperation with other libraries in order to allow them to join in the project.

D.1 Generation of the instruction proposal "Procedure for the BAGEs to contribute their records to the PCU", for approval by the General Coordination Committee.

D.2 Promoting the generation of an instruction on a BAGE Document Access Service.

D.3 Monitoring, presentation of proposals and adoption of service instructions issued by the General Coordination Committee.

E. Connectors and applications outside the PCU: As detailed in earlier sections of this report, there are certain basic issues with which libraries must comply, such as MARC21, OAI, NCIP, etc. One of the principles for their incorporation within the project at each of the stages proposed is specifically that they adapt to these basic aspects. Ultimately, though, they must all take part, and this will require a parallel analysis of those systems and libraries which are non-compliant in order to allow the required connectors or web services to be developed.

Planning

Meanwhile, at subsequent stages of the project, with the incorporation of the NCIP toolkit or other availability services, the connection with local systems will also be vital, and this working line will therefore need to be kept open throughout the entire project.

We include in this working line studies of the different ILMs, with the aim of achieving optimum reconfiguration of the OAI-PMH providers or the OAI Toolkit and NCIP Toolkit connectors. Analysing issues of connectivity or specific factors in local treatment and exportation:

- Metadata schemas employed: MARC21, IBERMARC, etc.
- Data export tools.
- Internal local management:
 - Internal handling of authority records and relationship with bibliographical records.
 - Handling of items.
 - Handling and management of digital objects
- Connections and NCIP implementation

The systems initially considered are: ABSYSNET, ABSYS, ALEPH 500, Unicorn, DIGIBIB, KOBLI. However, all other MARC and non-MARC ILM systems will also be analysed.

5.1.3 AREA 3: Project management and coordination.

The Sub-Directorate-General for Library Coordination is leading this project from the outset. We therefore include an overall project management and coordination line, which also covers dissemination and evaluation. The working lines are thus:

F. Project monitoring and coordination

F.1 Overall coordination and management of all project tasks

F.2 Dissemination: Dissemination campaigns both internally (BAGEs and users) and externally.

F.3 Evaluation and improvement: definition of indicators (qualitative and quantitative) serving to define project milestones and analyse the feedback provided by all participants and other experts, trends, etc.

5.2 Development phases of the Common Access Point project

It would be very difficult at present to define closed phases, with detailed tasks and objectives for the project. As a result, as in the administration of any large-scale project, we initially define three development phases, these phases being open to modification or the emergence of new requirements, although with a defined service level, which means that the planning of the initial phase and its results could modify subsequent phases, while maintaining the idea of a high-value-added service, scalability and a high level of integration underlying definition of the project from the outset.

5.2.1 INITIAL PHASE: Definition and objectives

The initial phase involves start-up of the PCU, and the main objective is thus to provide the general public with a collective BAGE catalogue for consultation, with bibliographical records which have been subjected to basic aggregation and normalisation. For each of the working lines we have defined clear objectives which are detailed and described below:

5.2.1.1 COMMON ACCESS POINT:

Line		Description	Aims	
A	PCU application	A1	Definition, analysis and monitoring	Service level: web consultation
		A2	Systems	
		A3	Development (New functionalities or modification of those in existence)	
B	PCU content and administration	B1	Configuration, parameterisation	Definition of site administration
		B2	Harvesting	ILMS with OAI
		B3	Processing	Basic aggregation and normalisation
C	Metadata schemas	C1	Definition and modification of XC schema	XCPCU schema definition

The above table, in accordance with the working line definition schema, establishes for the PCU application a **basic service level**. In other words, the application will need to harvest those ILMSs with provision for the OAI protocol, work with data at least in terms of the normalisation of codes and other metadata (item types, language codes, country codes...) and the detection of duplicates and the corresponding item generation (basic aggregation). These records must be available for web consultation (identifying the library which provided them and linking to its catalogue).

To continue with the nomenclature and definition we have included in the report, this could be summarised as follows:

Planning

- From the harvesting perspective, we will at this initial stage cover:
 - OAI provider ILMS.
 - ILMS with capacity automatically to export authorities, bibliographical and holdings details.
 - ILMS with capacity automatically to export bibliographical and holdings details.

- From the perspective of **processing the harvested metadata**, at the initial stage we will cover:
 - Harvesting and storage of metadata.
 - Basic normalisation and aggregation.
 - Normalisation and transformation to the XC schema.

- From the data exploitation perspective, including implementation of the various interfaces, the design, information architecture and translation into official languages, where required, with the following levels of integration:
 - Library indexing.
 - Link to the library.
 - Link to the record at the source.

5.2.1.2 BAGEs

Line	Description	Aims
D	Library Coordination	D1 Issues regarding library incorporation <ul style="list-style-type: none"> • Definition of the participation of the BAGEs • Record incorporation procedure
E	Connectors and applications outside the PCU	E1 Analyses and development of services and applications outside the PCU for connectivity with individual systems <ul style="list-style-type: none"> • Connectors for the exporting of records • Normalisation of local policies

At this initial stage, in terms of the participation of the BAGEs, the main objective is the definition and structuring of this participation. We therefore focus the objectives both on the definition of procedures and participation agreements which will need to be reflected in a General Coordination Committee instruction, and the tasks involved in creating the connectors linking the ILMSs to the PCU.

At this initial stage, ordered in terms of the number of records provided by each, the need will be to connect to the PCU application in order to automate the harvesting process (tasks connected with the tasks in line A3, and each of the local ILMSs):

- ILMS with capacity automatically to export authorities, bibliographical and holdings details.
- ILMS with capacity automatically to export bibliographical and holdings details.

These connectors must be based on the inherent ILMS export functionalities and connect to the OAI Toolkit for incorporation within the PCU. Given all the above, the ILMS providers themselves will need to be involved, along with the participating libraries to agree on criteria

Planning

for the normalisation of local practices, which must be applied in the parameterisation and XCPCU schema definition (**A3**, **B1**, **B3** and **C1**).

5.2.1.3 Project management and coordination

Line	Description	Aims
F	Project monitoring and coordination	F1 Execution and coordination <ul style="list-style-type: none"> Project management and leadership Planning and definition Phase II
G	Dissemination	G1 Internal dissemination campaign Design and execution of dissemination actions
H	Assessment and improvement	H1 Analysis and monitoring of feedback on all actions Project assessment reports

As on any project, the corresponding coordination tasks will need to be considered. All these functions are connected with the tasks described in the two previous lines. Their aims are to give support and structure to the project and to monitor the tasks in question. This initial phase will involve dissemination actions ensuring that all libraries are aware of the project and the procedures for their incorporation, along with involvement in working groups.

It is likewise very important to gather all feedback, bearing in mind the fact that the PCU is an open project which, working with the reality of the AGE libraries, is required to achieve their progress towards the most advanced techniques and standards, facilitating interoperability, the reuse of information and the creation of advanced services for citizens.

5.2.2 INTERMEDIATE PHASE

The main objective for the intermediate phase is the incorporation of ALL the AGE libraries, bearing in mind the fact that some of them will need to make efforts to adapt to the project conditions (OAI, MARC21) even if they were included in the prior phase analysing their ILMSSs, along with the need to develop connectors to rectify any harvesting limitations.

This second phase is the project consolidation phase, so the service considered involves a higher level and greater integration than at the previous stage. As in the initial phase, we set out the overall objectives below.

5.2.2.1 COMMON ACCESS POINT:

Line	Description	Aims
A	PCU application	A1 Definition, analysis and monitoring
		A2 Systems
		A3 Development (New functionalities or modification of those in existence)
B	PCU content and administration	B1 Configuration, parameterisation
		B2 Harvesting
		B3 Processing

Planning

				<ul style="list-style-type: none"> • Aggregation (FRBR, RDA) • Services for the clean-up and downloading of records for the participating libraries
C	Metadata schemas	C1	Definition and modification of XC schema	PCU schema consolidation

The above table, in accordance with the working line definition schema, establishes an **intermediate level of service** for the PCU application. In other words, the application will need to harvest those ILMs which provide for the OAI protocol and, with the connectors enabled for this purpose, process the metadata, transforming them in accordance with the PCU schema, and with definition of the aggregation service. These records will need to be available for web consultation, identifying the library which provided them and linking to the record allowing the user to continue using the data). Mobile device interfaces will be adapted during this phase.

To continue with the nomenclature and definition we have included in the report, this could be summarised as follows:

- From the **harvesting** perspective, the second phase covers:
 - OAI provider ILMs.
 - ILMs which have developed connectors.
- From the perspective of **processing the harvested metadata**, as the definition efforts will have been made in the previous phase:
 - Normalisation and transformation to the XCPCU schema.
 - Aggregation (FRBR and RDA).
- From the **data exploitation** perspective, including implementation of the various interfaces, the design, information architecture and translation into official languages, where required, with the following levels of integration
 - Link to the record at the source.
 - Record clean-up
 - Definition of advanced services, such as clean-up, downloading and normalisation of records for the libraries involved.

Web services will begin to be integrated during this phase: APIs or linked data for the enrichment of the data provided by the libraries.

5.2.2.2 BAGEs

Line		Description	Aims
D	Library Coordination	D1 Issues regarding library incorporation	<ul style="list-style-type: none"> • Consolidation and structuring of BAGE involvement • Incorporation and improvement of records
E	Connectors and applications outside the PCU	E1 Analyses and development of services and applications outside the PCU for connectivity with individual systems	<ul style="list-style-type: none"> • Connectors for the exporting of records • Connectors for NCIP integration • Normalisation of local policies

Planning

With regard to the participation of the BAGEs, in this phase the main objective is the consolidation of their participation and progress in the definition of procedures and instructions supporting the evolution of the PCU towards NCIP integration, authentication, users, etc. The libraries will, through the coordination bodies, need to provide and agree on information such as requirements for access to collections, user authentication and validation.

It will be necessary to continue working on the line covering connection with the PCU application to automate the harvesting process (tasks connected with the line A3 tasks and each of the local ILMs), in the case of those libraries which have not participated or which could improve the records provided (holdings, authorities...).

Lastly, this line will also need to work on development of the plug-in allowing for NCIP connection in order to display not only item availability information, but also reservation, loan, application services, etc.

5.2.2.3 Project management and coordination

Line		Description	Aims
F	Project monitoring and coordination	F1 Execution and coordination	<ul style="list-style-type: none"> Project management and leadership Planning and definition Phase III
G	Dissemination	G1 Internal dissemination campaign	Design and execution of dissemination actions
		G2 External dissemination campaign	Design and execution of dissemination actions
H	Assessment and improvement	H1 Analysis and monitoring of feedback on all actions	Project assessment reports

As on any project, consideration will need to be given to project coordination tasks. All these functions are connected with the tasks described in the two previous lines. Their aims are to give support and structure to the project and to monitor the tasks in question. The PCU dissemination actions at the Initial Phase project level are included at this phase, both internally and externally (citizens).

Feedback and definition and monitoring of qualitative and quantitative indicators are a fundamental aspect, allowing us to steer the project planning and definition process.

5.2.3 ADVANCED PHASE

The advanced phase, in addition to maintaining the level of service and increasing the number of libraries and records, is intended to develop all aspects considered as value-added services. The PCU will not only be integrated in real time with the data source systems, but data which are not strictly bibliographical will also be validated and will interoperate, while also working on the reuse of records, interaction with other information systems and user activity.

For this third, advanced service phase, the initial definition, as may be seen in the figure and the PCU schema, defines more ambitious objectives, as listed below.

Planning

5.2.3.1 COMMON ACCESS POINT:

Line		Description	Aims	
A	PCU application	A1	Definition, analysis and monitoring	<ul style="list-style-type: none"> • Service level: web and mobile queries • Information service integration (APIS, linked data) • ILMS integration and management systems • Reuse of user-generated content
		A2	Systems	
		A3	Development (New functionalities or modification of those in existence)	
B	PCU content and administration	B1	Configuration, parameterisation	Site administration
		B2	Harvesting	ILMS with OAI and connectors
		B3	Processing	<ul style="list-style-type: none"> • XCPCU schema • Aggregation (FRBR, RDA) • AUTHORITY CONTROL
C	Metadata schemas	C1	Definition and modification of XC schema	PCU schema consolidation RDF/ linked data

The above table, in accordance with the working line definition schema, establishes an **advanced level of service** for the PCU application. In other words, the application will need to harvest from ILMSs which provide for the OAI protocol and with which connectors have already been enabled, as in previous phases. However, services will also be enabled to work with metadata by performing transformation to the XCPCU schema, aggregation and authority control, equivalences and relationships among authorities harvested. These records will need to be available for web consultation (with integration of service level and availability), although with functionalities and capacities to manage user interaction information (comments, relationships, synopses and any functionality defined), in order to allow this information to be received by the participating libraries, rather than simply remaining within the PCU.

As in previous phases, we summarise:

- From the **harvesting** perspective, this stage will need to continue incorporating those libraries which have not already done so, under the same participation conditions and characteristics as previously defined:
 - OAI provider ILMS
 - ILMSs which have developed connectors
- From the perspective of **processing the harvested metadata**, as normalisation and aggregation efforts will already have been made, this stage will tackle the authorities control service (normalisation of terms, spelling, semantic analysis)
 - Normalisation and transformation to the XCPCU schema
 - Aggregation (FRBR and RDA)
 - Authorities control
- From the perspective of **data exploitation**, this phase will perform all efforts involved in integration with the ILMSs and systems employed at the following levels:
 - User authentication (permissions, access, user types)

Planning

- Circulation and availability information (rules for loans and document access, inter-library lending...)
- Management of data generated within the PCU by interaction between the system and other libraries, with users and information services (e.g. enriched records, SGCB directory), allowing the participating libraries to make use of these data if they wish
- Definition of advanced services, such as clean-up, downloading and normalisation of records for the libraries involved.

5.2.3.2 BAGEs

Line		Description	Aims
D	Library Coordination	D1 Issues regarding library incorporation	<ul style="list-style-type: none"> • BAGE interoperability
E	Connectors and applications outside the PCU	E1 Analyses and development of services and applications outside the PCU for connectivity with individual systems	<ul style="list-style-type: none"> • Connectors for the exporting of records • Connectors for NCIP integration • Normalisation of local policies

In terms of the involvement of the BAGEs, it is very important at this stage that the characteristics and circumstances for access to collections, and the services which the system is willing / able to offer, have been defined. All this information will need to accompany the metadata in order to allow it to be managed from the PCU. The idea is that users will be able to use the Common Access Point to display the availability of items, request access to the document (either physical or electronic), and it must be capable of recognising them as library users with the required permissions. To this end, in this phase we specify as one of the objectives what is referred to as BAGE interoperability.

In terms of external applications and connectors, the aims are shared with the previous phase, based on a belief that if more libraries are to be incorporated, or those which are already involved improve and enrich their systems, ongoing work will be required on the connection with the PCU application in order to automate the harvesting process (tasks connected with the tasks in line A3 and each of the local ILMs) in the case of those libraries which have not taken part or which could improve their contribution of records (holdings, authorities...) and the circulation of items (NCIP).

5.2.3.3 Project management and coordination

Line		Description	Aims
F	Project monitoring and coordination	F1 Execution and coordination	<ul style="list-style-type: none"> • Project management and leadership • Planning and definition Phase III
G	Dissemination	G1 Internal dissemination campaign	Design and execution of dissemination actions
		G2 External dissemination campaign	Design and execution of dissemination actions
H	Assessment and improvement	H1 Analysis and monitoring of feedback on all actions	Project assessment reports

Planning

The main objective in terms of project coordination is to consolidate the PCU as an advanced holdings dissemination and consultation service, and work will therefore need to continue not only on the management of the stated tasks, but also and above all on the dissemination of the service (internal and external) and the constant evaluation and improvement of the project. The individuals responsible for project coordination will need to keep abreast of new technologies, services and standards as they emerge, with a view to potentially improving the PCU.

6. Development of the Initial Phase

In this chapter we focus on the development of the project at the initial phase, expected to last a period of eighteen months, providing details both of the tasks to be performed and the human resources and job profiles required, along with the technological resources and infrastructure.

6.1 Detail of tasks

6.1.1 Detail

We below detail the tasks, as set out previously, and also attempt to offer an overview of project development over time, with eighteen months being considered a reasonable period for execution.

A. PCU application:

A.1 Definition, analysis and monitoring: This groups together generic development management tasks. The development team will need to be involved throughout the development period on the analysis and identification of requirements for each of the specific tools. This is an open task.

A.2 Systems: This deals with maintenance of the information technology working environments, along with the analysis and definition of requirements for the definitive environment and PCU production. The tasks identified are:

A.2.1 Installation and configuration in development (pre-production). Development providers will have their own environments.

A.2.2 Installation and configuration in production.

A.2.3 Systems maintenance.

A.2.4 Partial roll-outs. The idea is progressively to publish in pre-production and production the aspects developed in order to allow group participants to evaluate and work on harvesting, processing and the registration of new OAI providers.

A.3 PCU development: At an initial stage, and following the analysis performed within this project, the following developments have been identified as necessary.

A.3.1 Developments in OAI Toolkit

1. ILMs connectors with OAI Toolkit,
2. IBERMARC – MARC21 converter

Development of the Initial Phase

A.3.2 Metadata Service Toolkit

1. Service and Normalisation: Modification and adjustment (MARC and DC)
2. Transformation Service: Modification and adjustment (MARC and DC)
3. Aggregation Service - Duplicate Detection

A.3.3 Drupal Toolkit: PCU interfaces

1. Information architecture and functionalities (basic level)
2. Design and modelling
3. Development (modification, development, administration and integration of modules and extensions)

A.3.4 Developments in the XC Core (contribution to eXtensible Catalog Organization). Selection and preparation of the developments performed.

A.3.5 Development of a web tool for *metadata input control*,

B. PCU content and administration: All tasks in this phase involve the occasional assignment of technical staff throughout the project period. Naturally, a greater workload will be involved as further repositories and ILMs are incorporated as record sources.

B.1 Configuration.

B.2 Harvesting: This task will be called on to support working line **C**: metadata analysis.

B.3 Processing: Monitoring and testing entirely linked to area **C**: metadata schemas.

C. Metadata schemas: Analysis of harvested metadata, the XC schema and proposed PCU schema. The work performed and the modifications in the ecstasy schema will have a substantial impact online **A.3.3**. as the ecstasy schema is the basis of the functionality of the modules pre-loaded into DRUPAL.

C.1 Generation of the schema **MARC21slim_pcubage.xsd** on the basis of MARC21slim_rochester.xsd. This will be used to validate the MARC21 records entering the system.

C.2 Generation of the **XCPCU** schema based on the XC schema. This will be the actual schema for metadata within the PCU.

D. Library Coordination: :

D.1 Generation of the instruction proposal "Procedure for the BAGEs to contribute their records to the PCU", for approval by the General Coordination Committee.

D.2 Promote the generation of an instruction on a BAGE Document Access Service.

Development of the Initial Phase

D.3 Monitoring, presentation of proposals and adoption of service instructions issued by the General Coordination Committee.

E. Connectors and applications outside the PCU: At this initial stage, these developments will focus on the exporting of records and their communication with the OAI Toolkit and the developments undertaken in **A.3.1.** as well as in **C** because the schema requirements and metadata processing may need modification in the export fields (above all aspects of local usage, more than MARC)

E.1 Needs analysis (in accordance with OAI Toolkit requirements)

E.2 Development and functional adaptation

E.3 OAI toolkit connector

F. Project monitoring and coordination. Transversal tasks

F.1 Overall coordination and management of all project tasks

F.2 Dissemination: Internal and external dissemination campaigns (BAGEs and uses). This campaign in this initial stage, will need to focus on drawing up clear and concise information about the PCU, in addition to access to information about the overall project. The idea is to encourage participation in all communication platforms already developed by the project's coordinators, and to use this channel to provide further information regarding participation, while also collating any contributions made by the participants.

F.3 Assessment and improvement: definition of indicators (qualitative and quantitative) serving to define project milestones and analyse the feedback provided by all participants and other experts, trends, etc.

6.1.2 Timescale

We include a guideline timescale presenting a visual presentation of the tasks in logical sequence, although also where and when efforts will need to be centralised for each task.

The gradual incorporation of repositories or ILMs and the analysis, study and definition involved in certain tasks lead us to set longer durations for all tasks, although the following distinction has been made:

	Main task	Dedication to the task is complete, with individual management and milestones
	Final roll-out	End of project
	Secondary task. MAINTENANCE	Partial dedication. Although the bulk of the task is complete, it may be reopened as a result of requirements or definitions from another main task, or may need to be reconsidered before becoming a main task. Also for maintenance tasks.
x	One-off task	One-off and scheduled tasks
x	Secondary/one-off task	Secondary tasks (partial dedication) involving scheduled planning or monitoring actions.

Development of Initial Phase

6.2 Human resources and working team

The team involved in development of the PCU at its initial stage is as diverse as the inherent tasks defined and the institutions which will be involved.

The human resources dedicated to the project will belong to one of the groups listed:

- Personnel belonging to the Ministry of Culture, of the Sub-Directorate-General for Coordination which, as the project manager, will have in place resources for management tasks, along with technical work. The Sub-Directorate-General for Technologies and Information Systems will also provide consultancy services on the various project milestones.
- PCU working group: comprising representatives appointed by each of the ministries (essentially technical library staff or heads of information systems) who may be assigned particular tasks as specified, or a portion thereof.
- Documentary analysis experts belonging to the institutions involved or outside parties, on highly advanced tasks.
- Provider companies: this group includes providers of systems, developers, consultants.

The following table sets out the main profiles and roles in accordance with the tasks defined.

Line		Description	Profiles and roles	
A	PCU application	A1	Definition, analysis and monitoring	Analysts, developers and library technicians
		A2	Systems	Systems administrators
		A3	Development	Developers, web architects, designers and layout artists
B	PCU content and administration	B1	Configuration and parameterisation	Developers/application administrator
		B2	Record harvesting	Application administrator/Library technicians
		B3	Record processing	Application administrator/Library technicians
C	Metadata schemas	C1	Definition and modification of XC schema	Metadata experts
D	Library Coordination	D1	Issues regarding library incorporation	Legal advisers, librarians, administrative technicians
E	Connectors and applications outside the PCU	E1	Analysis and developments for connections between the PCU and individual systems	Librarians, systems administrator, development companies
F	Project monitoring and coordination	F1	Project management and leadership	Librarians
G	Dissemination	G1	Internal dissemination campaign	Librarians, administrators, designers and layout artists
H	Assessment and improvement	H1	Analysis and monitoring of feedback on all actions	Librarians

The project coordinator will establish a working team covering all the profiles established. This may be expanded if deemed necessary.

Development of Initial Phase

6.3 Technological resources

Another of the key aspects of project implementation is the provision of the infrastructure for the PCU platform.

As mentioned, from the outset servers will be required with the components needed for the application:

- For XC
 - HTTP webserver (such as Apache)
 - SQL database server (such as MySQL)
 - PHP 5.2, or higher
 - Drupal 6.12, or higher
 - File uncompresser (such as 7-Zip, tar, or gzip)

- Drupal requirements
 - Web Server:
 - Apache 2.x,
 - mod_rewrite module for friendly urls
 - Database server:
 - MySQL 5.0.15 or higher. MariaDB perhaps in the future
 - Application server
 - PHP 5.2
 - PHP modules: XML, PDO,GD, ImageMagick

- Additional software
 - Sun/Oracle Java 1.5, or higher
 - Apache Solr 1.3, or higher
 - Tomcat 6

In order to facilitate the partial roll-outs and all tests to be performed by the participants, during the project development phase the PCU will be hosted at a location outside the ministry.