BERNSTEIN ⊞ THE MEMORY OF PAPERS

ECP 2005 CULT 038097/Bernstein

BERNSTEIN

Deliverable no. 9, ref. D4.2 Integrated bibliography v. 1

Deliverable number	*D4.2*
Dissemination level	*Public*
Delivery date	*31 July 2007*
Status	*Final*
Author(s)	*The Bernstein Consortium*

*e***Content***plus*

---

[1] OJ L 79, 24.3.2005, p. 1.

The Bernstein Project
D4.2 Integrated Bibliography v.1

This document will present the first version of the Bernstein Integrated Bibliography deliverable. This deliverable has been implemented with bibliographical data from the German National Library (DNB) in Leipzig and is shown as screen shots in this document and also available on-line at http://dnb.cheshire3.org.
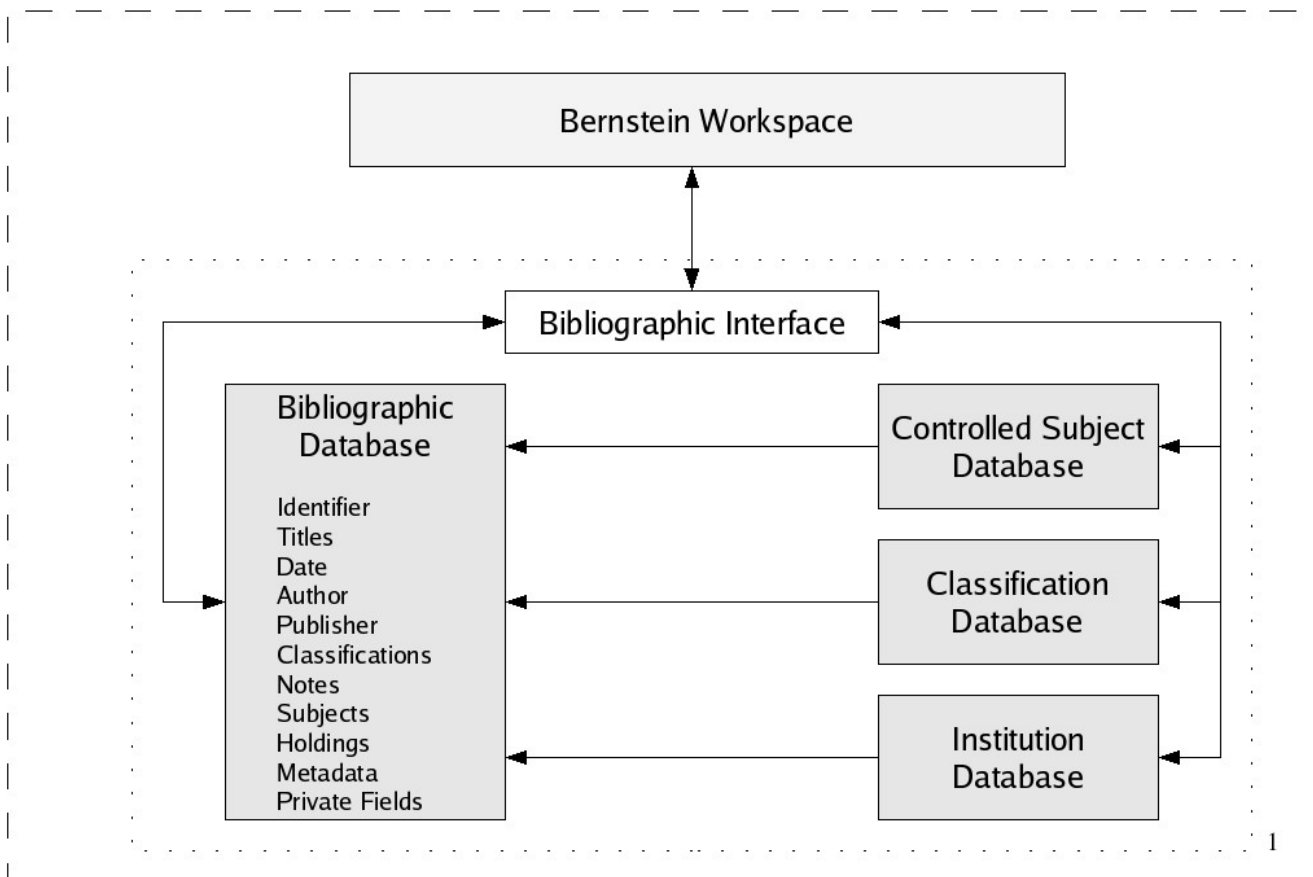
Background

At a meeting on the 2nd April 2007 between the DNB at Leipzig and the University of Liverpool it was decided that Liverpool would provide the technical implementation work for the Integrated Bibliography section of the Bernstein Project. It was also decided the DNB would export their existing Allegro database into a standardized XML version for interoperability. This work is being done by an expert subcontracted by the DNB (for example XML see appendix 1).

Architecture

The architecture of the system and the interaction with the Bernstein workspace is designed as shown below:

Bibliographic Integration Architecture

Each of the records in the bibliographic database will contain the information as listed (Identifier, Titles, Date, etc) either for all or some of these fields as required. The records will contain codes for information for controlled subjects, classification and institutions. These codes will be resolved using information contained in separate databases and can therefore provide the multilingualism desired for the Bernstein Workspace. These records can also be searched through the Bernstein workspace as described in deliverable 1.2 – The Workspace Integration.
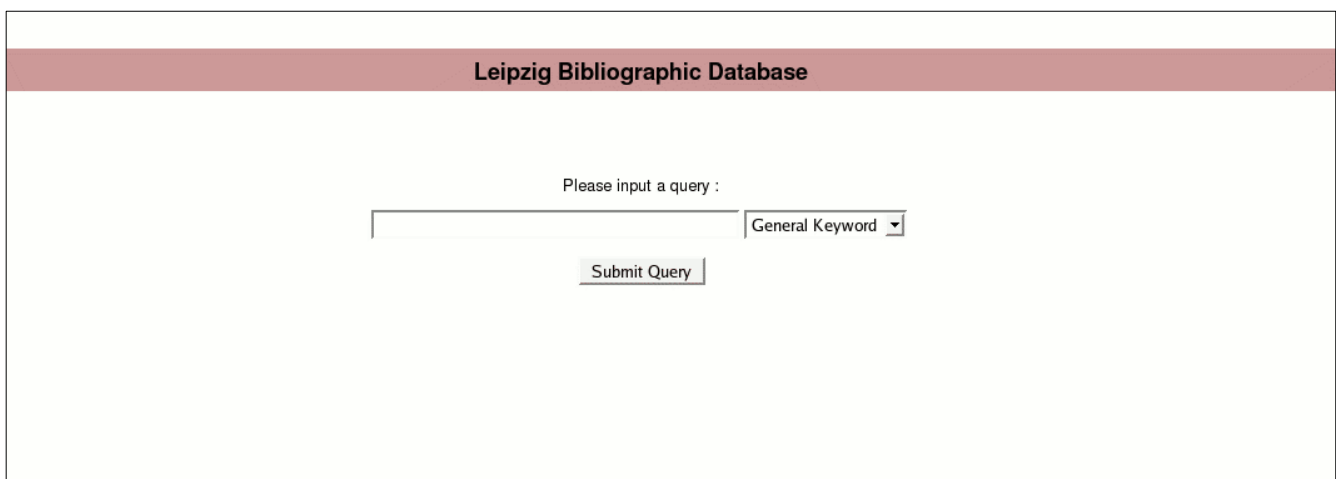
The bibliography is stored as a Cheshire3 database (www.cheshire3.org). This is the third generation of the Cheshire system started more than 10 years ago at UC Berkeley and more recently developed in a partnership between Berkeley and the University of Liverpool. The Cheshire platforms are used by several national services in both the UK and Europe, as well as by several services and projects in the US.

Cheshire3 extracts data from the given set of XML records into indexes. It pre-processes these records with configurable workflows to add normalisation. Once the indexes have been constructed, it supports operations such as search, retrieve, browse and sort. Using Apache handlers an interface is provided directly to this database, it is also accessible to the Bernstein Integrated Workspace via SRU (www.loc.gov/standards/sru/).

Version 1 Interface

The Search Interface
The first version of the Leipzig bibliographic database allows the user the ability to input a query to search either General Keyword, Title Keyword and Identifier indexes. The general keyword search will return all records that contain the query word somewhere in the record, the title keyword will return records which have the query word in the title, and the identifier will search for and identical copy of the query term encoded in the identifier tag within the record. The interface is shown below:

Leipzig Bibliographic Database

Please input a query :

[   ] General Keyword ▼

Submit Query

The Results Interface

The results interface displays a list of matching records. For each record the identifier, title and classification number are shown. The identifier is a link which can be clicked to display the full record. When the user enters the search term "papel" and selects "General Keyword" following results are shown below:

Leipzig Bibliographic Database

Identifier	Title	Classification
27103	Los animales en las marcas del papel	8.3.23.
7017	Heraldic watermarks or la heráldica en la filigrana del papel	8.3.24.
7018	La heráldica en la filigrana del papel	8.3.24.
7672	Das Einhorn mit dem dreigeteiltem Schweif	8.3.6.
19222	Los emblemas héraldicos en las filigranas	8.3.24.
7013	La filigrana de la „mano" en documentos de Galicia (España) de los siglos XVI y XVII	8.3.12.
558	Das Wasserzeichen des Pilgers	8.3.21.
8723	Interesantes filigranas de un pliego de papel	8.3.1.

Record Display:

The user can then chose to see the details of a specific record. In this version no attempt is made to make this display user friendly – the initial aim was to display the contents of the record. Therefore, at present this display shows the name of each of the tags marked up in the XML file and the information contained within those tags.

The records are displayed as shown below:

Leipzig Bibliographic Database

Tag	Contents
identifier	27103
author-name-personal origin="#10 " subjId="p2566"	
title type="normalized"	animales en las marcas del papel
title type="full" xml:lang="sp"	Los animales en las marcas del papel
booknumber origin="#02 "	27103
dc:date	1910
collation origin="#56 "	X, 171 S. : überwiegend Wz.-Abb.
format origin="#58 "	28 cm
publisher origin="#54 "	Villanueva y Geltrú : Oliva
subject-matter origin="#98 " subjId="ss599"	
subject-matter origin="#98 a" subjId="ss8813"	
subject-matter origin="#98 b" subjId="ss8814"	
class origin="#90 "	8.3.23.
subject origin="#90 " subjId="ss254"	

Appendix A

1) Example bibliographic record in XML

```xml
<?xml version="1.0" encoding="UTF-8"?>
<l:pgwz xmlns:l="http://www.d-nb.de/namespaces/pgwz/no_valid_uri">
 <l:bibliography>
  <l:record>
   <l:identifier>27103</l:identifier>
   <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:creator>
     <l:author-name-personal l:origin="#10 " subjId="p2566"/>
<!--Bofarull y Sans, Francisco de-->
    </dc:creator>
    <dc:title>
     <l:title type="normalized"> animales en las marcas del papel</l:title>
    </dc:title>
    <dc:title>
     <l:title type="full" xml:lang="sp">Los animales en las marcas del papel</l:title>
    </dc:title>
    <dc:identifier>
     <l:booknumber l:origin="#02 ">27103</l:booknumber>
    </dc:identifier>
    <dc:date>1910</dc:date>
    <dc:format>
     <l:collation l:origin="#56 ">X, 171 S. : überwiegend Wz.-Abb.</l:collation>
    </dc:format>
    <dc:format>
     <l:format l:origin="#58 ">28 cm</l:format>
    </dc:format>
    <dc:publisher>
     <l:publisher l:origin="#54 ">Villanueva y Geltrú : Oliva</l:publisher>
    </dc:publisher>
    <dc:subject>
     <l:subject-matter l:origin="#98 " subjId="ss599"/>
<!--Wasserzeichen \ Abbildungen (in Publ.)-->
    </dc:subject>
    <dc:subject>
     <l:subject-matter l:origin="#98 a" subjId="ss8813"/>
<!--Wasserzeichen \ Motive \ Tiere-->
    </dc:subject>
    <dc:subject>
     <l:subject-matter l:origin="#98 b" subjId="ss8814"/>
<!--Tiere (Wz.) \ #Allgemeines-->
    </dc:subject>
   </metadata>
   <l:classificationInfo>
    <l:class l:origin="#90 ">8.3.23.</l:class>
    <l:subject l:origin="#90 " subjId="ss254"/>
<!--#Allgemeines-->
   </l:classificationInfo>
  </l:record>
```