

Grant Agreement ECP-2007-DILI-527003

ARROW

Analysis of bibliographic resources and clearing mechanisms existing in Europe

Deliverable number/name	<i>D5.1</i>
Dissemination level	<i>Public</i>
Delivery date	<i>31 July 2009, review April 2010</i>
Status	<i>Final</i>
Author(s)	<i>Manfred Gravelius (MVB)</i>



eContentplus

This project is funded under the *eContentplus* programme¹, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

I. EXECUTIVE SUMMARY	3
II. METHODOLOGY	3
III. ANALYSIS OF BIBLIOGRAPHIC RESOURCES AND CLEARING MECHANISMS	8
1. INTRODUCTION	8
2. EXECUTION OF THE INTERVIEWS.....	9
3. SECTION 1: LIBRARY CATALOGUES.....	12
3.1 Survey of amounts	12
3.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment.....	13
3.3 Technical and business assessment, digital content assessment.....	15
4. SECTION 2: BOOKS IN PRINT DATABASES.....	17
4.1 Survey of amounts	17
4.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment.....	18
4.3 Technical and business assessment, digital content assessment.....	19
5. SECTION 3: RRO DATABASES	21
5.1 Survey of amounts	21
5.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment.....	22
5.3 Technical and business assessment, digital content assessment.....	24
6. SECTION 4: OTHER ORGANISATIONS AND/OR COMMERCIAL DATABASES	26
6.1 Survey of amounts	26
6.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment.....	27
6.3 Technical and business assessment, digital content assessment.....	28
CONCLUSIONS	31

I. EXECUTIVE SUMMARY

In general, the goals of this document are to:

- (i) Provide an overview of the current existing bibliographic resources and clearing mechanisms in Europe with regard on data coverage and quality, and in terms of technical and business assessment concerning accessible data.
- (ii) Provide consolidated findings with respect on the relevance of data in relation to ARROW objectives and deadlines and achieve basic information on the systems' attributes concerning the potential use and interoperability in the context of ARROW.

II. METHODOLOGY

The current section outlines the methodology adopted for D5.1 with the purpose of illustrating the main steps of the work done.

The Analysis of bibliographic resources and clearing mechanisms existing in Europe presented here (see Section III) result from the analysis and examination of complementary sources of information which have been collected throughout the first phase of the project:

- Desk analysis and processing of WP 4 – WP 5 questionnaire findings²: the analysis provided extensive data for the assessment of existing databases in Europe in terms of quality and coverage of the information, relevance for ARROW objectives, standard identifiers, metadata and technical protocols implemented thus enabling (1) the identification of main *technical areas* to be addressed by the D4.2 guidelines in order to enhance actual interoperability among the different resources involved in Arrow information infrastructure; (2) the identification of standards applicable for the effective management and exchange of bibliographic, rights and rightholders information within Arrow framework, as reported in D4.1; (3) the identification of best sources of information available at country and cross-country level, according to the type of information needed to fulfil ARROW purposes, as reported in D5.1³

² WP4 – WP5 questionnaire aimed at gathering background information for WP4-5 activities. It was acknowledged that the typologies of information needed for the two WPs were complementary and they could be collected within the same target group of professionals dealing with the management of information resources in stakeholders organizations, WP4 and WP5 leaders agreed on working together to harmonize WP4 and WP5 set of questions related to individual workpackages in a single questionnaire thus avoiding the duplication of efforts.

³ The methodology here illustrated is shared by also D4.2 deliverable as far as the preparation of the questionnaire and the first reworking of results are concerned.

- Face to face meetings at national level with key stakeholders: the purpose of the face to face meetings was to follow up on any issues that needed further clarification from the questionnaire results and from the examination of intermediary reports per country. As a result, a subset of organizations to be involved in the first phase of Arrow piloting has been selected. In addition, meetings with the representatives of the relevant databases which will be involved in Arrow pilot have been arranged to discuss technical and business issues related to the integration of individual databases. The National meetings enabled the refinement of D5.1 analysis in terms of domain at cross-country level to be involved into ARROW infrastructure, as well as technical areas to be approached by D4.2 analysis thus facilitating the identification of key requirements for individual databases to meet in order to achieve actual interoperability within the Arrow network.

Here follows a step by step synthesis of the methodology adopted:

Outline of Arrow objectives and preliminary identification of relevant information resources to be involved within the system infrastructure: a preliminary outline of the Arrow system was drafted in order to identify, for each task to be performed, the type of data needed and which of the organisations' databases should be considered as relevant information resources to be investigated and possibly integrated into Arrow information infrastructure. A pre-questionnaire has been circulated among partners to provide a first overview of the resources available from partners in the first place and guidance for a further refinement of the questionnaire. The outline of the Arrow objectives was also intended to provide guidance for project partners in order to identify the relevant databases in each country, so that the relevant organisations could be contacted and interviewed within the framework of WP4-5 questionnaire⁴.

Furthermore, the document provided interviewers with clear criteria to evaluate the relevance of the information resources investigated during the survey and was annexed to Guidelines for the Interviewers⁵.

WP4-WP5 questionnaire design: The Arrow system outline provided the basis for designing the WP4-WP5 questionnaire. For each type of resource (bibliographic, rights, parties information) within the Arrow information infrastructure, key issues to be investigated through the questionnaire were identified and appropriate questions prepared accordingly⁶.

⁴ For the complete list of stakeholders interviewed, please see paragraph 2 Execution of Interviews

⁵ See Annex II– *Guidelines for the Interviewer and Glossary*

⁶ See the table illustrating Arrow objectives in ANNEX II – *Guidelines for the interviewer and Glossary*. For each tasks to be performed by Arrow system, relevant questions included in WP4-WP5 Questionnaire are tracked.

Beside the overall assessment of the resources investigated and presented in D5.1, specific sections aimed at gathering key information for the technical assessment for D4.2 objectives, have been developed, focusing in particular on:

- Use of standard identification systems, with particular attention to standard identifiers for textual works and manifestations (respectively, ISTC and ISBN)
- Implementation of standard metadata schema for managing different types of information (bibliographic, rights, rightholders data)
- Technical protocols implemented to access and exchange metadata among the different players' databases (libraries, publishers, bibliographic agencies, commercial databases, RROs) in the content value chain

Survey on existing bibliographic resources in Europe and standards implemented: In order to collect exhaustive information, the WP 4 - WP 5 questionnaire has been conceived and designed as a structured grid for conducting face to face discussions with the representatives of the organisations selected for the interviews. The management of the survey was based on the "Arrow groupings" approach, aimed at facilitating the coordination of the survey on a geographical/regional basis and effectively involving both project partners and supporters in the interviews⁷.

Analysis of the questionnaire results: data collected through the WP4-WP5 questionnaire has been further processed in order to obtain more structured reports, tailored for the specific WP4 and WP5 objectives:

1. Summary of individual stakeholders' databases: organisations in charge of conducting the interviews in each country provided a summary of results for each database investigated using a template structured in four thematic areas:
 - Relevance of the database(s) in relation to Arrow objectives
 - Data quality
 - Technical conditions to access and query database(s)
 - Business conditions to access and query database(s)
2. Outline of National scenarios: questionnaires and related summaries⁸ have been subsequently collected by National Contact Points, who were in charge of a further aggregation of results to provide a synthesis of the scenario which emerged at

⁷ The rationale underlying Arrow groupings is illustrated in D1.1 Project Operative Workplan

⁸ All the questionnaires collected and related summaries are available on Arrow website in Partners Only area; WP4 and WP5 sections: <http://www.arrow-net.eu>

national/geographic level⁹, aimed at identifying the relationships among main databases interacting in the book value chain at a national level (national libraries, ISBN Agencies, publishers databases, Books in Prints, RROs).

3. Examination of the country reports: on the basis of the information collected in the previous steps of the analysis, detailed reports were prepared for each country involved in the survey. These country reports were internally used a basis for identifying the key databases to be involved in piloting the Arrow infrastructure.

National meetings: once the key databases to be involved in the pilot phase had been identified, following the assessment of the national scenarios emerging from the detailed country reports, national meetings with key stakeholders from the countries involved in the Arrow pilot¹⁰ were held. Experts with specific knowledge about the structure and content of the databases selected for the pilot were also invited to attend these meetings. The main goal of these national meetings was to gather clear indications for the Arrow system design in terms of:

- requirements (definition of users groups and expectations)
- resources (which databases will be queried by ARROW, in which order/hierarchy)
- use-cases (actual test bed for the system test phase)

Refinement of the methodology and structure of the analysis: input received from the national meetings enabled:

- the identification of three cross-country *domains* to organize D5.1 analysis: Library Domain, Books in Print Domain, RRO Domain
- the refinement of *technical areas* for D4.2 analysis on the basis of domain-specific interoperability issues within the Arrow infrastructure

D5.1 Analysis: the analysis of bibliographic resources and clearing mechanisms existing in Europe focused on the three main cross-country domains identified as the most relevant for the ARROW objectives: National Library Domain, Books in Print Domain, RRO Domain. It considered, however, also the sources of information not belonging to the above mentioned

⁹ See methodology enclosed in ANNEX II - *Guidelines for the Interviewer and Glossary*.

¹⁰ Reports on National Meetings and AIE slides are available on Arrow website in Partners Only area; WP4 and WP5 sections: <http://www.arrow-net.eu>

domains for an assessment of their inclusion in ARROW infrastructure at a later stage of the project. For each domain identified, relevant sources of information have been analysed according to three main general thematic areas:

- Quantity of data available
- Relevance of bibliographic data in relation to ARROW objectives and coverage
- Technical and business assessment, digital content assessment

This analysis provided the framework information for the Guidelines on technical interoperability (D4.2) and the start up of the ARROW System Design.

III. ANALYSIS OF BIBLIOGRAPHIC RESOURCES AND CLEARING MECHANISMS

1. Introduction

The first task in the WP5 of the ARROW project is to give an analysis of the existing bibliographic resources and clearing mechanisms in Europe, by considering the countries and organisations directly involved in the project.

In order to get an overview on the state of art of existing resources, a questionnaire was compiled and then used to interview the partners, as described in Section II METHODOLOGY.

The report is structured into 4 categories: Library catalogues, Books in Print databases, RRO databases, data from other organisations.

Each category is structured into 3 areas of analyses:

- The area “Survey of amounts” indicates the quantity of data available
- Relevance of bibliographic data in relation to ARROW objectives and coverage: a summarized evaluation is provided whether metadata and databases are (or are not) applicable according to ARROW tasks
- In the area “Technical and business assessment, digital content assessment” technical attributes and business features are described in order to gain basic information concerning compliance and interoperability.

Due to data confidentiality, detailed information in the areas “Survey of amounts” of the sections 2: Books in Print databases, 3: RRO databases, and 4: data from other (commercial) organisations are excluded in the public version of the analysis. They are published in Annex III.

2. Execution of the interviews

Findings were received from the following organisations:

interviewer	country	organisation	type
AIE	Italy	BNCF (Florence National Library)	Library
		Ediser	ISBN agency
		Ediser (Reprographic Service)	Service provider
		ESAIE	Service provider
		ICCU	Bibliographic agency
		IE (Informazioni Editoriali)	Bibliographic agency (BiP)
		mEDRA (European DOI registration agency)	DOI agency
BL	UK	ALCS (Authors' Licensing and Collecting Society)	RRO
		PLS (Publishers Licensing Society)	RRO
		WATCH (information about copyright holders)	other
		DACS (Design & Artists Copyright Society)	RRO
		CLA (The Copyright Licensing Agency)	RRO
		BL (The British Library)	Library
		Nielsen Book Data	Bibliographic Agency (BiP)
BNE	Spain	DILVE (Distribuidor de Información del Libro Español en Venta)	Bibliographic agency (BiP)
		Agencia Española del ISBN	ISBN agency
		Spanish ISMN Agency	ISMN agency
		BNE (National Library of Spain)	Library
		CEDRO (Spanish Reproduction Rights Centre)	RRO
		FGEE (Federation of Spanish Publisher's Guilds)	Publishers association
BNF	France	BNF (Bibliothèque nationale de France)	Library

interviewer	country	organisation	type
		BNF Gallica (digital objects)	Library
		DILICOM, le réseau du livre commercial database)	Service provider
		Electre / AFNIL (Books in print database, ISBN database)	Bibliographic agency (BiP), ISBN agency
		SUDOC (Système Universitaire de Documentation)	Bibliographic agency
		PERSÉE (digitization and online publishing of scholarly journals archives)	e-retailer
		BIUM (digital library de médecine et d'odontologie)	Library
		CFC (Centre Français d'exploitation du droit de Copie)	RRO
		Lavoisier (commercial database)	Publisher
		Hachette Livre (commercial database)	Publisher
		Numilog (aggregator of digital content)	Service provider
MVB	Germany	DNB (German National Library)	Library
		VG Wort (Collecting society)	RRO
		FRC (Frankfurt Rights Catalogue)	Service provider
		MVB (ISBN agency)	ISBN agency
		MVB (VLB – German Books in Print)	Bibliographic agency (BiP)
		MVB (libreka! – Full text search /e-book platform)	Bibliographic agency
NLN	Norway	DnBB (Den norske Bokdatabasen)	Bibliographic agency (BiP)
		Kopinor	RRO
		LINO	RRO
		NLN (The National Library of Norway)	Library
UIBK	Austria	UIBK (ALEPH, ALO: University and Regional Library of Tyrol)	Library
		Literar Mechana (Collecting society)	RRO
		ÖNB (Austrian National Library)	Library

interviewer	country	organisation	type
NUK	Slovenia	IZUM (Library information service - Conor, Cores, Cobib)	Service provider
		NUK (National and University Library)	Library, Bibliographic agency (BiP), ISBN agency
		SAZOR	RRO
KB	The Netherlands	KB – Royal National Library of the Netherlands	Library
		Centraal Boekhuis	Bibliographic agency (BiP) service provider
		Stichting Leenrecht	RRO
		Stichting Lira	RRO
		Stichting Pictoright	RRO
		Stichting PRO	RRO

With regard on the countries of the organisations taking part on the interviews, we have the following allocation:

	Library catalogues	BIPs / Bibliographic agencies	RROs	Other organisations
Austria	x	see VLB	x	
France	x	x	x	x
Germany	x	x	x	x
Italy	x	x		x
Norway	x	x	x	
Netherlands	x	x	x	
Slovenia	x			x
Spain	x	x	x	x
UK	x	x	x	x

3. Section 1: Library catalogues

3.1 Survey of amounts

	Bibliographic metadata approx no. of records	Parties and/or rights data approx. no of records	Remarks
ÖNB (Austria)	8 Mio	Parties	Authority file: 2.6 Mio rec.
UIBK ALEPH (Austria)	3.5 Mio		
BNF (France)	10.8 Mio	Parties	Authority file: 5 Mio rec.
BNF Gallica (France)	775,000		
BIUM (France)	11,000		
SUDOC (France)	10.7 Mio	Parties	Authority file: 1.9 Mio rec.
DNB (Germany)	9.9 Mio	Parties	Authority files: 4.5 Mio rec.
BNCF (Italy)	2.1 Mio	Parties	Authority file
ICCU (Italy)	10 Mio	Parties	
KB (Netherlands)	14 Mio	Parties	
NLN (Norway)	4,9 Mio	Parties (publ.): 11,200	
NUK (Slovenia)	1 Mio	Parties	
BNE (Spain)	3,4 Mio	Parties	Authority file: 4.5 Mio rec.
BL (UK)	13 Mio	Parties	Authority file (Library of Congress)

3.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment

ÖNB (Austria)	The catalogue of the ÖNB is very relevant covering Austrian literature. It is linked to an Name authority file (PND) which would also be relevant to ARROW objectives. Data are normalized and the input follows clear rules
UIBK ALEPH / ALO (Austria)	The ALEPH catalogue is relevant as it contains bibliographic data of more than 400.000 books published before 1988. ALO is the largest Austrian online digital collection of monographs, but the rights and parties data section of ALO is under construction and will be available not before 2010. Metadata mostly are imported from other sources.
BNF (France)	The BNF catalogue is relevant for ARROW because of its bibliographic and authority data. Information about works is established when necessary for catalogue consistency. Whenever a work is already digitised, links are established from the bibliographic entry to the digital object. BNF performs high quality control on data.
BNF Gallica (France)	Gallica could be made available to ARROW. Gallica services see annex 4.
BIUM (France)	The Medic@ digital library as a local repository of cultural digital documents is focussed on the history of medicine, odontology and health. It contains records from BIUM and other partners and digital libraries (ex. Gallica). Bibliographic metadata are not quality controlled.
SUDOC (France)	The Library Union Catalogue for the academic, specialist and higher education in France contains the holdings of 1,100 libraries. Records are created books-in-hand and picked on external databases. SUDOC is interested in ARROW as for recording manifestations which are not present in other databases such as BNF or Electre.
DNB (Germany)	The information of the DNB catalogue is relevant for ARROW. With regard to the availability status of the publication there is a link from the search results to the title records in the VLB database (buchhandel.de). The data recording and import processes are controlled by using defined plausibilities and field checkings.
BNCF (Italy)	The BNCF catalogue contains relevant information for ARROW purposes. Linkings to digital objects are considered (MARC label 956), but see also ICCU. BNCF collects records from 7 other libraries.
ICCU (Italy)	ICCU manages the network of the National Library Service SBN, insofar the bibliographic metadata and the parties metadata (authors name authority file) could be relevant for ARROW. The links (at manifestation level) to a digital object may be relevant. The data quality is compliant to library bibliographic control rules.

KB (Netherlands)	KB-MDO integrated metadata repository is relevant to ARROW. It contains bibliographic metadata and links between the different manifestations of the same work are made either by a textual reference, by a URL or pointer. These methods are also used by linking to the digital object where it is available.
NLN (Norway)	The BIBSYS database contains the catalogue for university / college libraries, the National library and a number of other libraries. NORBOOK contains bibliographic records in the national bibliography, and parties data are held in ISBNFORLAG. All catalogues can be accessed on the web by public. Quality control is performed by Librarian staff.
NUK (Slovenia)	The database Katnuk contains bibliographic metadata as well as BIP and ISBN information, as NUK is also running a BIP agency and the national ISBN agency. ¹¹ The dLib database contains bibliographic information about digital material. Bibliographical records are of prime quality.
BNE (Spain)	The bibliographic database (manifestations) and the authority file are relevant for ARROW purposes. The bibliographic and parties data are created books-in-hand and are of adequate quality. Links to a digital version (not under the copyright law) of a non digital work are indicated
BL (UK)	The British Library Integrate Catalogue is relevant to ARROW. It contains the bibliographic metadata (manifestations) and parties data (NACO authority file). Relationship to the underlying work is also given by the use of uniform titles. Concerning data exchange the BL is collaborating with the Library of Congress. For quality control MARC report and FRBR quality measure is used.

1

1

For more information on NUK acting as BiP, see also, 4. Section 2: Books in Print databases

3.3 Technical and business assessment, digital content assessment

ÖNB (Austria)	<p>The database can be queried through Z39.50. In case of ARROW participation the business conditions should be negotiated with the German National Library.</p> <p>The Austrian National Library also manages digital content. (This was not interviewed).</p>
UIBK ALEPH / ALO (Austria)	<p>Both databases can be searched, queried and harvested: Z39.50 for UIBK Aleph, OAI-PMH for ALO.</p> <p>The digital content stored in ALO represents the largest Austrian online digital collection of monographs, but rights and parties data of ALO is yet under construction.</p>
BNF (France)	<p>For the catalogue bibliographic file Z39.50 is used for querying, resp. OAI-PMH for harvesting (OAI-CAT and OAI-NUM repositories). Metadata are freely available on the web, commercial societies have to pay a subscription fee for extracts from the catalogue.</p> <p>Digital objects still under copyright are not available outside BNF. In ARROW perspective the relevant information is on the Gallica platform.</p>
BNF Gallica (France)	Gallica could be made available to ARROW (public domain digital material).
BIUM (France)	<p>Harvesting can be made in OAI-PMH.</p> <p>The Medic@ is harvested by Gallica, public domain works are available.</p>
SUDOC (France)	<p>SUDOC is freely accessible by Z39.50, but cannot be harvested.</p> <p>Digital content: only dissertation abstracts, include in metadata (manifestations).</p>
DNB (Germany)	<p>The protocols Z39.50, SRU and OAI-PMH are used for retrieving and harvesting. The access to the bibliographic catalogue data via the OPAC website (HTTP) is free.</p> <p>The digital collection on the archive server contains monographs and articles. Their metadata are included in TEL+, for Europeana is planned.</p>
BNCF (Italy)	<p>Query protocol for the BNCF OPAC: SRU, for the harvesting of data related to digital collections an OAI-PMH compliant metadata exposition. Access to the whole BNCF OPAC is free, the BNI subset (National Bibliography) is under licensing agreement.</p> <p>Digital content assessment see ICCU.</p>
ICCU (Italy)	<p>Z39.50 and SRU protocol can be used to query data, while there is no OAI-PMH compliant metadata exposition. The access to the whole of the database is free.</p> <p>The digital collection is already included in Europeana (via Cultura Italia). Terms</p>

	and conditions for access are not defined yet.
KB (Netherlands)	<p>KB-MDO is searchable via the SRU protocol. The metadata is made available for harvesting via OAI-PMH.</p> <p>For importing, exporting and exposing metadata about manifestations, works and parties, there are many formats in the NLN databases. For import and export, they include MARC, DC and proprietary formats. MODS is also used for data export. Formats to enable data harvesting are DC, MODS and MARC.</p>
NLN (Norway)	<p>The standard protocols Z39.50, OAI-PMH and SRU are used for searching and querying the bibliographic data. Except of this also proprietary protocols are used.</p> <p>The digital collection (NB digital) is accessible via the web. It contains digital versions of public domain works as well as of copyrighted works.</p>
NUK (Slovenia)	<p>Outside aggregators can use Z39.50 to search and query data from Katnuk, and OAI for dLib.si. The access to the data is free.</p> <p>Digital content is only available from dLib.si, which contains copyrighted and copyright free material.</p>
BNE (Spain)	<p>The BNE databases are freely accessible, the protocol Z39.50 is used for search and query in the bibliographic catalogue. The digital library is OAI-PMH compliant, the access to this database is free, too (except of printing images).</p> <p>The digital collection is already included in Europeana and The European Library.</p>
BL (UK)	<p>For managing metadata in the BLIC catalogue the Z39.50 protocol is used. Anyone can use the catalogue via the website without restrictions.</p> <p>The BL holds digital content in their Digital Library for the purpose of full text search. The access is restricted (there is open access via Internet for the UK PubMed Central collection). BL participates in Europeana.</p>

4. Section 2: Books in Print databases

4.1 Survey of amounts

	Bibliographic metadata approx no. of records
Austria see VLB Germany	The databases cover approx. 14,2 Mio bibliographic metadata altogether.
Electre / AFNIL (France)	
VLB (Germany, Austria)	
libreka! (Germany, Austria)	
IE ALICE (Italy)	
Centraal Boekhuis (Netherlands)	
DnBB (Norway)	
DILVE (Spain)	
NUK (Slovenia)	
NBD (UK)	

4.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment

Austria see VLB Germany	
Electre / AFNIL(France)	<p>Bibliographic records in the Electre database (BIP) are linked to the publisher authority records established by AFNIL (ISBN agency). Electre contains status information in print / out of print which is relevant for ARROW.</p> <p>The quality of the data is highly appreciated by customers.</p>
VLB (Germany, Austria)	<p>VLB contains bibliographic and BIP information on manifestation level and covers German, Austrian and Swiss publications. The publisher records are created by the German ISBN agency. VLB contains status information in print / out of print and is relevant for ARROW, concerning the German speaking market. The data quality is approved. Linking from VLB records to libreka! are established.</p> <p>As MVB is authorized to act as ISTC agency, the VLB will be enhanced with ISTC in order to create relationships between works and manifestations.</p>
libreka! (Germany, Austria)	The libreka! platform can be available for ARROW in context with VLB.
IE ALICE (Italy)	<p>The IE databases (BIP ALICE and publisher database) can provide high quality metadata for ARROW. The bibliographic data contain the status in print / out of print and the quality is optimised, including duplicate check and data enrichment.</p> <p>IE is planning to enhance the database by implementing the ISTC.</p>
CB (Netherlands)	<p>Centraal Boekhuis manages CB Online a centralized database that serves as the basis for the Books in Print, the tele-ordering service, the ISBN agency and all other added value services offered to Dutch publishers and booksellers.</p> <p>The databases are therefore relevant for ARROW.</p>
DnBB (Norway)	<p>The DnBB contains metadata about all books (manifestations) being published in Norway. Information about the status in print / out of print is included.</p> <p>Quality checks are performed by the staff, the publishers are updating the information themselves.</p>
DILVE (Spain)	<p>The bibliographic and BIP data are provided and maintained directly by the publishers into the database, which is based on ONIX. Status information in print / out of print is available for ARROW purposes.</p> <p>As data quality is maintained by the publisher itself, only consistency checks are made.</p>

NUK (Slovenia)	<p>The bibliographic and BiP data are provided directly by the publishers into the database, when they register an ISBN. Data are confirmed and enriched “book in hand” by NUK staff.</p> <p>It must be noted that BIP in NUK environment is meant to collect the early notifications about new books, not for commercial purposes, therefore the out of print status is not managed.</p>
NBD (UK)	<p>Nielsen BookData provides comprehensive bibliographic data, with availability status plus other relevant metadata for ARROW. Commercial agreements have to be worked out for the use of the DB in ARROW.</p> <p>Moreover manages the ISBN Agency for UK and Ireland and is one of the founding organizations of the ISTC International, acting also as ISTC Registration Agency in the UK.</p>

4.3 Technical and business assessment, digital content assessment

Austria see VLB Germany	
Electre / AFNIL(France)	<p>Access to the Electre website is restricted to subscribers, the search engine is proprietary. There is no access protocol for searching / querying data implemented. The access to the AFNIL website (to search the publisher authority file) is not restricted, but also only a human being may search the website.</p> <p>Electre only manages book covers (JPEG) as digital content. Access is not freely available but might be opened to ARROW under conditions. AFNIL does not manage digital content.</p>
VLB (Germany, Austria)	<p>The VLB Books in print catalogue is available in two versions: freely accessible via the website buchhandel.de (only a subset of the metadata is available), and the complete catalogue with whole information for subscribers. A proprietary XML-based API interface is available for querying bibliographic data. See libreka! For the access to digital content.</p>
Libreka! (Germany, Austria)	<p>The access to the digital objects in libreka! is public and for free, but the view of the full text content is restricted by locking pages according to publishers permission.</p>
IE ALICE (Italy)	<p>The protocols to access BIP and publishers database are based on a proprietary format. Data import/export is possible by using web service or FTP. BIP and publishers data are available under specific licensing terms. IE does not manage digital content.</p>
CB (Netherlands)	<p>Data exchange – import and export – allows ONIX format. Data are available under specific licensing terms. CB deals also with e-books.</p>

DnBB (Norway)	ONIX XML format is applied for the import of data. The database is not freely accessible, only for subscribers or price model on the file. No digital content is managed.
NUK (Slovenia)	See section 2: Library catalogues
DILVE (Spain)	DILVE is accessible freely. Data extractions in different formats can be programmed. An API to download data under program control is implemented. An OAI-PMH server is planned for 2009. DILVE does not deal with digital objects.
NBD (UK)	NBD can import data in ONIX XML, as for export, data are available in several formats: ONIX XML, MARC Format, CSV and proprietary NDB XML format. FTP is used. No digital content is managed.

5. Section 3: RRO databases

5.1 Survey of amounts

	Bibliographic metadata approx. no. of records
Literar Mechana (Austria)	The databases cover approx. 20 Mio bibliographic metadata altogether.
CFC (France)	
VG Wort (Germany)	
Stichting Lira	
Stichting Leenrecht	
Stichting PRO	
Stichting Pictoright	
Kopinor (Norway)	
SAZOR (Slovenia)	
CEDRO (Spain)	
ALCS (UK)	
PLS (UK)	
CLA (UK)	
DACS (UK)	
WATCH (UK)	

5.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment

Literar Mechana (Austria)	The databases are very relevant as it identifies Austrian right holders and rights. The database of manifestations is of relevance for arrow, too, and may be used in addition to the library catalogue of ÖNB and the VLB BIP. It has to be considered that there are several databases (per use case), not only one with all manifestations and related rights. Continuous quality check is done, in particular for the right holder's data.
CFC (France)	The database allows to search by rightholder / publisher or by work (manifestion level), and is relevant in the ARROW framework. The Bibliographic data of manifestations allow to see if a work is copyrighted, but does not specify the status in print / out of print. A set of rules is used to check the quality of the data. Data from external sources (Electre, BnF) are also used.
VG Wort (Germany)	The databases are very relevant for ARROW purposes, they identify German rightholders and rights. They contain the copyright status (copyrighted or public domain). The rights information concerns secondary rights. Continuous and regularly quality checks are performed, also on the bibliographic metadata.
Stichting Lira Stichting Leenrecht Stichting PRO	As organisations joining VOI©E foundation, Dutch RROs are currently involved in the agreement signed in 2008 by FOBID and VOI©E within the Digiti©E Committee (Digitisation of Cultural Heritage) which will facilitate the use of orphan works, based on voluntary agreements between RROs and rightholders. Their integrated database is relevant for ARROW purposes.
Stichting Pictoright	Being collective rights and primary rights organization for visual authors is at the moment not included in the ARROW, as so far the project deals with textual works.
Kopinor (Norway)	The KIF database (ACCESS database) contains data about publications, rightholders (publisher and authors), and rights. The data contain a coding wether a work is copyrighted or public domain. Employees collect and receive information and data when needed, but also randomly from various sources outside the organization. The work flow/logistics secures the quality of the rightholder information.
SAZOR (Slovenia)	SAZOR is a non for profit reprographic rights organisation of Slovenian authors and publishers. It was registered in the second half of 2007 when SAZOR became fully operational as a Slovenian RRO, when it was licensed by the Slovenian Intellectual

¹ ² Its founding members, the Slovenian Publishers Association and the Slovenian Authors Association ZAMP, have brought into membership 22 publishing companies, including the biggest which represents 96% of Slovene publishing market, and over 700 authors and translators. Collection is based on the Slovenian Copyright Act and governmental regulation. Photocopying in copy-shops is covered, as well as imports of

	<p>Property Organization.¹²</p> <p>SAZOR recognizes the importance of the challenge that development of digital uses of written works brings, including further development of copyright law.</p>
CEDRO (Spain)	<p>CEDRO keeps the bibliographic metadata about manifestations (monographs, monographic chapters, serials), which are linked to rightholders (authors / publishers) and rights. Because of data coverage and quality (checked by internal procedures) the database is relevant for ARROW.</p>
ALCS (UK)	<p>ALCS uses the database MRS which contains bibliographic data, parties data (writers) and rights data. The data are flagged concerning the copyright status (copyrighted or public domain). Metadata of manifestations are imported from Nielsen and Ulrichs. Quality checks are performed.</p>
PLS (UK)	<p>Bibliographic metadata and rights data are kept in the PLS-e database, publishers data in the MRS database. A few orphan works are existent. Bibliographic data are obtained from publishers or from CLA. Quality control is performed on the data.</p>
CLA (UK)	<p>The OPS database covers titles (bibliographic information), parties and rights, an external database covers excluded works (external accessible). Bibliographic information is collected from Nielsen and Ulrichs, information on licenses, rights and mandates are collected from PLS. CLA works closely with PLS and other partners to ensure the accuracy of data.</p>
DACS (UK)	<p>DACS holds information about artists, their works and uses of their works, their personal contact details and those of their heirs and beneficiaries as part of DACS' copyright licensing, Payback and Artist's Resale Right services. The CRM database contains in particular parties and rights data. DACS does not store systematically bibliographic information. Information on Licenses and Rights and Mandates are collected from the Publishers Licensing Society. DACS performs quality control on the data.</p>
WATCH (UK)	<p>WATCH is managing a website providing information about copyright holders. The parties data contain contact details and related information for copyright and permissions enquiries about writers and artists. Data are collected by research.</p>

copying machines and similar products (scanners, printers etc.). Further in 2008 SAZOR also participated in the private reproduction remunerations, where SAZOR received 6% of the total amount collected – for publishers only. A further 6% (for authors) was handled through ZAMP.

5. 3 Technical and business assessment, digital content assessment

Literar Mechana (Austria)	<p>The data are held in internally used databases (Oracle). No standard interfaces or protocols are currently supported. The databases could be made available to ARROW if this is within a contract.</p> <p>Digital content assessment: not relevant.</p>
CFC (France)	<p>The database (based on an Intranet solution) is only accessible for internal use, no standard interfaces or protocols are currently supported. Information about orphan works could be made available to ARROW.</p> <p>No digital content is managed.</p>
VG Wort (Germany)	<p>The databases are only accessible for internal use. A web service is implemented for rights data (including licenses and permissions). The databases could be made available to ARROW, but with regard on the confidentiality of the members data.</p> <p>No digital content is managed.</p>
Stichting Lira Stichting Leenrecht Stichting PRO	<p>Only internal access on the database is allowed. xml is used to import rightholders and bibliographic data from NBD, the umbrella organisation of public libraries. Export is allowed for internal usages in CSV format</p> <p>Digital content assessment: not relevant.</p>
Stichting Pictoright	<p>Only internal access on the database is allowed.</p> <p>Digital content assessment: not relevant.</p>
Kopinor (Norway)	<p>The ACCESS database is only accessible for internal use. No technical protocols are reported. The data are not available for free or under specific licencing terms.</p> <p>Digital content assessment: not relevant.</p>
SAZOR (Slovenia)	<p>N/A. Further investigation is needed.</p>
CEDRO (Spain)	<p>CEDRO uses a web service protocol for information exchange and information search (bibliographic, parties, and rights data). The protocol for the data access used by CEDRO's database is ADO (resp. OLEDB for parties database). CEDRO could make the database or subsets available to ARROW upon agreement of the terms.</p> <p>Digital content assessment: not relevant.</p>
ALCS (UK)	<p>A web service protocol is used for searching / retrieving the data (manifestations and parties metadata). Data are not available for free, the access is strictly controlled. The database could be made available to the Arrow system subject to data protection restrictions.</p> <p>Digital content assessment: not relevant.</p>
PLS (UK)	<p>Data are available to publishers for searching / querying reasons. ONIX for RROs is used for data import.</p>

	Digital content assessment: not relevant.
CLA (UK)	The access to the OPS database is only for internal reasons (shared with partners). ONIX for repertoire is used for the import of rights data. A Web service to enable data harvesting is implemented. Digital content assessment: not relevant.
DACS (UK)	Only internal access on the database is allowed. There was no response to the questions about technical and business assessment, but DACS declares its interest in the development of the ARROW project.
WATCH (UK)	Public access website. There was no response to the questions about technical and business assessment.

6. Section 4: Other organisations and/or commercial databases

6.1 Survey of amounts

Organisation / Database	Bibliographic metadata approx. no. of records
AFNIL ISBN Agency (France) see Electre	<p>The databases cover approx. 7,9 Mio bibliographic metadata altogether.</p> <p>For detailed information on the specific databases, refer to annex 3.</p>
DILICOM (France)	
PERSÉE (France)	
Hachette (France)	
Lavoisier (France)	
Numilog (France)	
Frankfurt Rights Catalogue (Germany)	
MVB ISBN Agency (Germany)	
Ediser ISBN Agency (Italy)	
Ediser (Reprographic Service) (Italy)	
ESAIE (Italy)	
mEDRA (Italy)	
IZUM (Slovenia)	
ISBN Agency (Spain)	
ISMN Agency (Spain)	

6.2 Relevance of bibliographic data in relation to ARROW objectives and coverage assessment

AFNIL ISBN Agency (France) see Electre	
DILICOM (France)	As a commercial, not bibliographic database (used mainly for logistics), the data do not contain much bibliographic information about manifestations, and no information concerning the underlying work. Also there is no information to rightholders or rights. The data do contain information about the status in print / out of print in good quality and thus maybe useful for ARROW. The commercial data (ex. prices) are quality controlled.
PERSÉE (France)	The databases consist of bibliographic data about journals, issues and articles, authority data (persons), and rightholders data. They are linked together, thus it is possible to identify the right status of a work (copyrighted or public domain), but contains no information about in print / out of print status. Links also exist from the work to a digital collection, if it has been digitised already. Comprehensive quality checks are performed.
Hachette (France)	The database serves mainly commercial purposes, but is used for marketing as well as for rights management reasons. Bibliographic metadata are at manifestation level, it is planned to introduce ISTC in future. Parties and rights metadata are maintained also, but the interface for rights metadata is to be incremented in the next years.
Lavoisier (France)	Data are maintained in 2 databases (authors' rights / information about parties, and bibliographic metadata) which are not linked directly. Information about rights is extracted from the contract, namely the copyright. Bibliographic metadata is regularly updated by the publisher as the work evolves.
Numilog (France)	As an aggregator of digital content on the French market, Numilog manages a bibliographic database, links are established between the underlying work and its manifestations. The use of the ISTC is planned. All works in the database are copyrighted, the status in print / out of print is not indicated currently. Quality control is done by the editorial staff and/or the publisher.
Frankfurt Rights Catalogue (Germany)	The FRC provides data (bibliographic, rights, parties and contact data) in order to serve as a trading platform for rights concerning in print publications (audio, film/TV, merchandising, translation). Though there is no information about digitalisation rights, the database or subsets could be relevant for ARROW. Data are provided and updated by the publishers directly. Quality control is performed by the editorial staff.
MVB ISBN Agency (Germany)	Data could be relevant for ARROW purposes, but could also be taken partially from the VLB, as far as the subset "active publishers with records in the BIP" are concerned. Currently the database containing all ISBN publishers and author/publisher records is only accessible for internal use. The data are also used for the directory of the German booktrade and thus being quality controlled.
Ediser ISBN Agency (Italy)	The agency maintains a title and publishers database. Bibliographic metadata (manifestations) contain the ISBN and core information about books. Being the ISBN standard widely adopted in Italy it could be assessed that the ISBN databases covers a high percentage of the book market and thus being valuable

	for ARROW purposes.
Ediser (Reprographic Service)	Ediser – Reprographic Service databases could provide relevant information to enable Arrow system to uniquely identify (through the ISBN code and bibliographic metadata) a book and to identify the publisher and the contributors of that book and access related contact data.
ESAIE	ESAIE service for textbooks adoption could provide relevant information to enable Arrow system to uniquely identify (through the ISBN code and bibliographic metadata) a subset of Italian books (i.e. books addressed to primary and secondary schools), and partially track the relations between different editions of the same books (Books records include the ISBNs of previous editions, up to max 5 ISBN) and to retrieve data about publishers (name and contact data).
mEDRA (Italy)	Metadata are directly provided by DOI registrants, no quality control is performed later on. The metadata scheme allows to collect well structured information, on manifestation level as well as on works level. But as the DOI registration service doesn't cover a high percentage of the European book market, thus the mEDRA database should not be considered as an exhaustive source of data for ARROW in a first step.
IZUM (Slovenia)	IZUM maintains and hosts the national Slovenian union catalogue Cobib as well as the parties database Conor, which is also used by NUK. See there for further information.
ISBN Agency (Spain)	The agency manages bibliographic data for all ISBNs assigned in Spain (based on manifestations, and including in print / out of print status, reported by the publisher) and keeps a publishers database. No information about the right status of publications is kept.
ISMN Agency (Spain)	The ISMN database allows to distinguish between the status in print / out of print of printed music publications. The music publishers file does not contain rights information. The database may not be relevant for ARROW in a first step.

6.3 Technical and business assessment, digital content assessment

AFNIL ISBN Agency (France) see Electre	
DILICOM (France)	<p>Protocols used for managing data: FTP, AS2 (peer to peer), HTML (web) for export.</p> <p>DILICOM does not manage digital content.</p>
PERSÉE (France)	<p>The manifestation metadata can be harvested by OAI-PMH and searched by Z39.50.</p> <p>PERSÉE provides free public access and could be made available to ARROW.</p> <p>At the present, 70 collections of journals (full text) are available for free public access.</p>

Hachette (France)	<p>OAI-PMH is planned for 2010. The database is not open to the public, there is currently no plan to open access for any use other than cooperate.</p> <p>Hachette Livre takes part in Europeana and TEL, certain titles of the group's publishers are found referenced in Google Book Search.</p>
Lavoisier (France)	<p>Bibliographic metadata is exported, but not harvested, and neither parties or rights data are exported or harvested. The information in the database is not open to the public.</p> <p>Lavoisier takes part in the Gallica2 project and in Europeana.</p>
Numilog (France)	<p>The data on the database are available only internally, but access and query of the database in collaboration with ARROW is welcome.</p> <p>Numilog takes part in Gallica.</p>
Frankfurt Rights Catalogue (Germany)	<p>The database is accessible via web application, but only premium subscribers have full access on the data. Proprietary protocols are used for the query and searching. It is perhaps intended to integrate the catalogue into a German solution "portal for clearing and trading rights", but it is quite unsure if this project would match the ARROW deadlines.</p> <p>Digital content assessment: not relevant.</p>
MVB ISBN Agency (Germany)	<p>Currently there is no access for external aggregators or users. Thus no protocols for query and search are implemented.</p> <p>No digital content is managed.</p>
Ediser ISBN Agency (Italy)	<p>The bibliographic database is available for searching and querying metadata via web application and via web service (XML format). No formats for data harvesting have been implemented.</p> <p>No digital content is managed.</p>
Ediser (Reprographic Service) (Italy)	<p>The metadata schemas implemented in the three databases are proprietary. When possible, an ISBN is associated to a bibliographic record in the bibliographic database. No other standard identifiers are used. No service is available for automatic data export, searching, querying or data harvesting.</p> <p>No digital content is managed.</p>
ESAIE (Italy)	<p>The metadata schema implemented in the textbooks database is proprietary. No service is available for automatic data export, searching, querying or data harvesting. A service to query the database via web application (http protocol) and retrieving a subset of metadata in HTML format is publicly available. Specific search options are available for publishers only as they can search and retrieve all the metadata of all their titles. The request is made via web application, and an ASCII or Excel file is returned, formatted in a proprietary standard, containing complete metadata records.</p> <p>No digital content is managed.</p>

mEDRA (Italy)	<p>No protocols for data harvesting are implemented. A web service to query and retrieve metadata in XML format is available to DOI registrants.</p> <p>No digital content is managed.</p>
IZUM (Slovenia)	<p>Outside aggregators can use Z39.50 to search and query data. The catalogue is available for public searching.</p> <p>Digital content assessment: not relevant.</p>
ISBN Agency (Spain)	<p>The data can be accessed interactively through the web, access is free for any use.</p> <p>No digital content is managed.</p>
ISMN Agency (Spain)	<p>No protocols are implemented for query and search the database. Access through the web is free.</p> <p>No digital content is managed.</p>

CONCLUSIONS

Most of the **libraries** who responded to the questionnaire are sure to match the ARROW deadlines, especially those who are partners in the project. BNCF and ICCU are mentioning some restrictions which should be discussed. Regarding the deadline for ALO UIBK, it depends on the needs of ARROW.

It is common for the catalogues of the National libraries, that they use the standard identifiers (ex. ISBN on manifestation level), the standard data formats (ex. MARC) and the standard protocols for data access or harvesting (ex. Z39.50, OAI-PMH) in a same or compliant way. This will be useful in order to consider the relevant data within the ARROW workflow and architecture.

The **BIP** providers who took part in the interviews generally assume that they match the ARROW deadlines (with the exception of DnBB). But with respect on restrictions concerning the data access or technical adjustments details have to be discussed with the partners.

The BIP systems generally have in common the use of the ISBN and ISSN as unique identifiers, and offer the same or similar functionalities. They mainly use the ONIX for Books standard as the primary format for data acquisition. Some of them are planning to introduce the ISTC as standard identifier for textual works. The quality of the data reflects on the needs of the publishers and is in any case up to date. As the systems are customized in order to meet requirements of the national book markets, they may use proprietary solutions concerning data access more often than libraries.

RROs generally confirm matching with ARROW deadline, but there are several restrictions concerning accessing the data, as RROs also hold confidential information. Besides this, the aspect has to be considered that they – working as collective societies – usually do not offer public access to their data. Therefore the adoption of data in the ARROW framework will depend on the kind of data to be used and will have to be negotiated.

As the databases of the different RROs are available basically for internal use and for membership purposes, they generally hold the same kind of data but are not compliant in a technical sense, as, for example, it is given with regard to the library systems, where mostly common standards are applied. Therefore, as a rule, more technical adjustments might be necessary to consider these data in the technical ARROW framework.

Most of the **other organisations** interviewed confirm the availability of their data regarding the ARROW deadlines, but in some cases it will depend on the type of data required by the project (ex. Lavoisier, Numilog). Some of the databases probably could be included only in a second step. In some cases ownership problems first have to be solved (ex. IZUM).

Due to the fact that these databases usually are built and managed for special commercial purposes (apart from the databases managed by ISBN agencies), they are as a rule not compliant in sense of data contents and functionality. In any case it might be necessary to investigate in detail technical issues (data mapping, proprietary identifiers, interfaces, accessibility, other restrictions.)