Grant Agreement ECP-2007-DILI-527003

ARROW

Annual Report

1st September 2008 – 31st August 2009

www.arrow-net.eu

Deliverable number/name	*D1.5*
Dissemination level	*Public*
Delivery date	*26th February 2010*
Status	*Final*
Author(s)	*Maria Loi (AIE)*

*e*Content*plus*

---

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

Table of content

Project Objectives -

ARROW aims at creating a rights information infrastructure to facilitate diligent search of rightholders in large digitisation programmes and the identification of public domain works, orphan works, out of print works and other copyrighted works, thus helping to manage a key risk in the digital library initiatives – the so called "black hole" of XX century works that are impossible to digitise and make available because they are in copyright but could be released for inclusion and access if only the rights information infrastructure existed.

Digital libraries initiatives both at national and international scale will benefit from the infrastructure and in particular the development of Europeana will be enhanced.
In creating the infrastructure ARROW builds on and sets out to implement the conclusions and tools reached by the EC established Digital Libraries High Level Expert Group (HLEG) and its copyright subgroup.

ARROW system will consists of two core components:
- a distributed network of resources to provide the users (primary libraries that wish to scan and make available a book) the best information available on the right status of European books
- a Registry of Orphan Works (RoW)
If rightholders for a certain books are not found, the title will be stored in a RoW searchable by authors, publishers and other rightholders for future claims.

The collaborative approach that sees all stakeholders (libraries, collective management organizations, rightholders) involved in the project gives an added value to the project that demonstrates how copyright issues in the digital environment can be approached through cooperation between parties involved and the innovative use of technologies.

Consortium

The Consortium setting responded to the following criteria:

➢ representation of all the stakeholders involved;
➢ involvement of the most significant experiences already developed or under development in Europe;
➢ presence of high level expertise referred to the work to be done;
➢ inclusion of a significant number of Member States.

ARROW involves in a pan-European consortium key representatives of stakeholders in the book value chain (national libraries, publishers and collective management organisations, also representing writers – working through their main European associations)
In addition to ARROW contracting partners, several national organisations became officially supporters of the project so expressing their adhesion to ARROW objectives and contributing to project activities and results.
Partners and supporters from 13 countries of the European Union are actively committed with the project: Italy, France, Germany, Spain, United Kingdom, Austria, Slovenia, The Netherlands, Norway, Finland, Denmark, Belgium, Sweden.

Libraries - National libraries in Spain[2]*, France*, UK*, The Netherlands*, Germany*, Slovenia*, Finland*, and the University Library of Innsbruck, Austria*

Publishers - Publishers Associations of Italy*, Spain*, France, Sweden, Germany through its service company MVB*, and the Publishers Licensing Society (PLS) in UK.

Reproduction's Rights Organisations (RROs) in UK*, Spain, France, Italy, Denmark, Norway, Finland, plus the Author's Licensing and Collecting Society (ALCS) in UK.

International Organisations - Federation of European Publishers (FEP)*, International Federation of Reproduction Rights Organisations (IFRRO)*, The European Digital Library Foundation (Europeana)

Technology Developers - Cineca (Italy) and Numilog (France)

The role of the different stakeholders represented in the consortium guarantees that the rights information infrastructure can benefit from the best information sources available to facilitate identification of rightholders and right status of a book.
Libraries, publishers organisations and RROs are key metadata providers, respectively bibliographic, publishers' and rights ownership metadata.
At the same time, they act also as end users of the system.
Libraries will be able to use the system in their digitisation programmes to conduct diligent search, RROs will be able to issue licences according to national frameworks and will check the Orphan Works Registry on behalf of rightholders, publishers (and authors) may have their products available in the digital environment in full respect of copyright.
International Organisations will ensure that the project is known among communities of the different domains (libraries, RROs, publishers and authors) and the results could be shared and scalable into a wider environment further to the project duration.
A high level of expertise is guaranteed by the involvement of several partners in important digital libraries experiences such as Libreka (Germany), Gallica-2 (France), Enclave (Spain) and the Bookshelf project (Norway).
The technological provider CINECA in charge of the set up and implementation of the system architecture provides a solid framework for exploiting state of the art technologies for innovative services of Arrow.

Project Results/Achievements

The first year of the project completed all preparatory work necessary for the launch of the system infrastructure.
A set of **preliminary studies** was completed:

- Report on legal framework (1st Edition)
the report contains an overview of the legislation in place, at international and EU level; an outline of the agreed principles and recommendations at European level developed within the framework of the European Commission High Level Group on digital libraries regarding orphan and out of print works; and an analysis of the examples of existing and planned initiatives in Europe on clearance mechanisms to facilitate the use of orphan works and out of print works

- Report on business models (1st Edition)
It comprises a theoretical background built on previous studies examined; a synthesis on the business models in the e-book market, digital libraries and digitisation initiatives in Europe (with an overview of the e-book market in Europe, the main business models and digital library initiatives, the players in

---

[2] * contracting partners

the e-book world, the digitisation process, the economic results and the relevant licensing models); a series of more detailed country files.

- State of the art and guidelines for standards applicable (1st Edition)
the report provides an overview of existing standards applied and/or available in each domain that were described in their key features highlighting those aspects (e.g. rights coverage) that could be interpreted in terms of strength/weakness with regard to future ARROW applications

- Guidelines for technical interoperability
This analysis allowed the identification of core requirements for technical interoperability within ARROW; these are intended as the guidelines that each database for each single domain (libraries, RROs, publishers/rightholders) needs to follow in order to achieve interoperability within ARROW infrastructure. These guidelines therefore provide the interoperability framework to unite into a coherent system all the individual information sources and systems identified as being relevant for ARROW purposes.

- Analysis of bibliographic resources and clearing mechanisms existing in Europe
A comprehensive set of information on structure, protocols and standards used, quality and quantity of data and metadata stored in databases for all the three domains concerned (libraries, rightholders, RROs).

The studies led to the definition of a coherent **workflow for future Arrow system** that was ultimately translated into core **technical specifications of the infrastructure**.
Key actors in the Arrow workflow and therefore key nodes for information provision/exchange come from all the three domains and are: The European Library (TEL) , the Virtual International Authority Files (VIAF), Books in Print databases, RROs repertoires.
The information needed for the aims of ARROW is therefore stored in the systems of different organisations for their own specific purposes, but it's given added value when made interoperable in Arrow for the objective of identifying rights status and rightholders.
The system design defined the key requirements for the architecture that will:
➢ maintain the bibliographic rights metadata in a distributed network of resources
➢ allow new partners (as metadata providers) to join the ARROW federation with minimum efforts
➢ design a distributed search technique that takes into consideration a decentralized structure where external sources are linked together in a connected network

This would allow guaranteeing an adequate performance, scalability, and flexibility of the system.
In this framework, **ARROW acts as interoperability facilitator**
➢ to access and query different systems
➢ to retrieve the relevant data
➢ to process and exchange this data with other systems
➢ to add complementary data from other systems
➢ to produce new meaningful information

The definition of the Arrow workflow allowed to outline the pilot countries that will be involved in the first testing of the system (Germany, United Kingdom, Spain, France) while the definition of core features of system architecture gave the concrete terms of commitment needed in Arrow system set-up by the different stakeholders in the three domains concerned.

The first year of the project was also important for **raising awareness about ARROW objectives**.
A website was launched to serve as main communication tool and ARROW was promoted in several events at national and international level, within EC institutions and with frequent exchanges with Europeana and the other Europeana-related projects.
The representativeness of the consortium for the entire environment (libraries, rightholders, RROs) related to digital libraries initiatives ensured an increasing familiarity of different actors across various EU countries with the objectives of Arrow and its ongoing work and will provide a good basis for a

wide consensus and interest on the future services, thanks also to the solid network of partners and supporters.

The relevance of ARROW as a European solution for the development of digital libraries and for an effective management of copyright in digitization initiatives has been recognized at EC level by the HLEG and by Commissioner Mrs Viviane Reding, who mentioned ARROW as an innovative project that should be encouraged in the effort to create a European System of Rights Registries[3].

The objective to build a service that could serve for all digitisation initiatives was at the basis of the continuous dialogue and cooperation framework established both with Europeana and with non-European initiatives like Google's Books Search and the forthcoming BRR.

Target Users & their Needs

The development of an effective solution for the management of rights information addresses the needs of different stakeholders in the value chain. Here following a summary of target users that could benefit from ARROW, their needs and the country coverage that the project can ensure for those targets through partners and supporters.

Target user description	Needs	Country coverage
Rightholders	✓ To offer their content in the new environment ✓ To maintain control over the content ✓ To receive remuneration from use of their content also through the declaration of rights on orphan works	All Europe through FEP + some countries directly as partners (IT, ES, DE) or supporters (FR, UK, SE)
Rightholders representatives and/or agents (RROs)	✓ To offer new value added services, in particular rights clearance	All Europe through IFRRO + some countries directly (UK) or as liaisons (ES, FR, DK, NW, FI, SL, NL)
Libraries	✓ To reduce costs in rights acquisition and thus include more content at the same budget level ✓ To promote inter-operability for econtent ✓ To avoid duplication of efforts in digitisation	All Europe through EDL Foundation+ some countries as partners (FR, ES, SL, UK, DE, NW), external supporters (FI) or through the EOD network (PT, HU, LI, DK)
e-retailers and other intermediaries	✓ To create commercial supply of econtent collection of copyrighted works ✓ To provide services to rightholders ✓ To reach new potential markets	LIBREKA in Germany (MVB) Numilog in France

---

[3] Viviane Reding Member of the European Commission in charge of Information Society and Media, The Future of the Internet and Europe's Digital Agenda Lunch debate on the future of the Internet and Europe's digital strategy Brussels, 6 October 2009

Underlying Content

The scope of the project is "rights information" about books. One of the first achievements of the project is the definition of "rights information" as a set of metadata including:

➤ the commercial status (in-print *vs.* out-of-print) at work level, as defined by the HLEG;
➤ the unambiguous identification and location of the rightholder(s) and – in case such identification is not possible – the definition of the "orphan" status of a work;
➤ the existing mandates to clearing centers for licensing the work for defined uses (e.g. scanning and making available in the Internet).

Such information are currently spread in a vast array of different sources, usually belonging to the three different communities involved: libraries, books in print databases, and RRO repertoires. The ARROW challenge is to make all such sources interoperable, through the use of appropriate standards. The problems with the pre-existing data are:

➤ In every community, data are not interoperable trans-nationally, with the exception of the library domain, where a key role is played by the TEL project, which already achieved a level of interoperability sufficient for the Arrow system purposes;
➤ There is not interoperability cross-domain: data within library catalogues are not interoperable with Books in print database and both are not interoperable with RROs repertoires;
➤ All data are created at "book" level, while rights information are defined at "work" level: in all the existing resources there multiple records for the multiple books containing the same work (e.g. a novel), and grouping such records is a big and unprecedented challenge.

The added value provided by the ARROW project is precisely to make all those sources interoperable, as far as "rights metadata" are concerned.

The creation of such a rights information infrastructure depends on the availability of existing bibliographic data and rights information. Partners and liaison organisations will provide several million records to be made interoperable for the project purpose.

The consortium will provide to the project data about more than 10 million items, covering around 70% of the European books in print and significantly also out of print, with some information about rights in many cases.

In fact, library catalogues are usually the source covering the highest number of titles and are the first basis to clustering different editions of the same work. The library authority files for authors name are a key step for unambiguous identification of rightholders. Books in print databases are essential to define the status of "in-print" / "out-of-print" and can be used to support the unambiguous identification of publishers. Finally, RROs repertoires are often the best source for further rights information on textual works, and in particular for rightholders identification and existing mandates.

Summary of Activities

The first year of the project consolidated the consortium structure and management and built the basis for the launch of the system infrastructure.

The key elements of **project organization** were defined and a methodology of work (Arrow groupings) established that will be pursued throughout the project, being flexible enough to valorise the contribution from further stakeholders that would like to contribute to the project.

The "Arrow groupings" gather partners and supporters at national or regional level; the rationale underlying this approach is to provide an intermediate layer between partners/supporters on one side and WPleaders/Coordinator on the other. This approach also facilitates exchange between different stakeholders and benefits the project with a coordinated input providing a clear scenario for the area.

Six Arrow groupings (each one with one appointed National Contact Point) were formed (UK, IT, FR, DE, ES, EOD) and coordination was established at national level for the Nordic countries and The Netherlands.

The first year was devoted to the completion of **preliminary studies** necessary to provide the legal, business and technical background needed to design the system architecture.

Through desk analysis and interviews with stakeholders, also with the contribution of national groupings, two studies provided respectively an overview at EU level of legislative framework related to the identification of the status of a work and business models scenarios for text-based digital content in order to provide solid knowledge basis and indications for the future ARROW service on legal aspects to be considered and its possible role in the current market scenario.

The remaining analyses provided on one side a report on existing standards that are relevant to be considered and/or are directly addressed in ARROW workflow and on the other identified core requirements for data sources from the different domains addressed by Arrow (libraries, publishers, RROs) needed to achieve interoperability within the ARROW infrastructure with a specific focus to standards and metadata messaging.

The **system design** included the definition of the workflow of Arrow that was ultimately translated into technical features. A technical group formed by experts from the three domains was formed to conduct the work.

The ARROW workflow runs from the first step of the library request (= Library submits a query to ARROW about rights related to a book they want to digitize) to the final phase where a licence or, in alternative, information about a licensor is given (= RRO provides feedback to ARROW and ARROW to library)

This workflow:

➢ Is based on metadata exchange
➢ Works on step by step metadata enrichment
➢ Involves Libraries, BiPs and RROs

For the libraries side, a specific mention should be made about the role of The European Library (TEL) as single access point to library data sources in ARROW. TEL will be progressively including the collections from 46 from the 48 National Libraries in Europe into The European Library portal. TEL was therefore chosen to be the single access point to library data sources for ARROW during the project phase and into the future. All 7 national libraries who are partners in ARROW already have their collections available via The European Library portal, as do the 4 ARROW associates. Furthermore The European Library is intended to become the libraries aggregator for Europeana, which further strengthens this position.

Particular efforts were devoted to **dissemination and awareness activities**.

A website was launched with the aim to act as reference point for stakeholders and the general public and a working place for Arrow partners and supporters through the design and set up of a cooperative area. The website is frequently updated with news and documents related to Arrow activities, and relevant issues for the digital libraries environment, including libraries, rightholders, RROs and the copyright community. In addition to the website, information material has also been produced.

The promotion of ARROW consisted in the participation of project representatives to national and international events and a constant relationship with the EU institutions, and in particular the HLEG and its subgroup on copyright of which some Arrow partners are members. Clustering with Europeana was implemented through participation to various working groups.

Several articles and news were published mainly in institutional and professional magazines and newsletters and press releases were issued for news of particular importance

Interest for the project grew consistently among all stakeholders and the relevance of the project for the future development of digital libraries was widely acknowledged. Project representatives have been invited to present Arrow in many events of particular importance in the stakeholders communities. Among others, it is worth to mention:

➢ The two conferences on digital libraries organized first by the French and than by the Spanish Presidency of the European Union
➢ The Supply Chain Expert Meeting at the Frankfurt Book Fair
➢ A meeting of the High Level Expert Group on Digital Libraries
➢ A meeting of the Member State Expert Group on Digital Libraries

The project has been presented also in a number of meetings of the European associations of the three key communities involved (libraries, publishers and RROs), as well as in many national initiatives in the book world.

Also beyond Europe, the project gained reputation so that it has been invited for presentation, for example, in a international conference on book standards at the National Library of Korea and at the CERLALC meeting in Bogotá, at the presence of a number of Latin American ministries of culture.

Next steps of the project are:

1st Semester 2010:
- First release of the system
- The system is progressively operative in Germany, United Kingdom, Spain and France

2nd Semester 2010:
- Assessment and validation of the system
- Launch of Registry of Orphan Works (RoW)
- Progressive implementation of the system in other European countries

A second focus will be on sustainability in order to provide solid ground for duration of services beyond the project duration, this will include also a comprehensive treatment of IPR issues related to the system.

Dissemination activities will promote ARROW services fostering implementation in further countries and consolidating knowledge and consensus around the project.

An update of studies will ensure that the results are consistent with the present legal and business framework and technical requirements.

Impact & Sustainability

Impact

The impact of the Arrow project is to be analysed in respect to the general aim of implementing the model for dealing with rights in digital library initiatives, as set by the *Memorandum of understanding* signed in June 2008 by European stakeholders as a result of the work of the HLEG on digital libraries.

In order to have consistent parameters to measure such impact it is useful to compare this European approach with the model emerging from the Settlement agreement between Google and the US associations of authors and publishers, pending for approval at the New York Court to conclude a class action lawsuit. The main differences between the two models are two:

1. The European model includes a diligent search of rightholders before using any work, and prior consent any time it is possible to find a party entitled to authorise the use. The Settlement agreement provides instead an apparently simplified approach for out of print works, for which the prior consent is not necessary, and rightholders are asked to claim their works if they wish to actively manage the respective rights.
2. The European model is based on a distributed, standard based infrastructure to search right information in different sources, in order to maintain a distributed control of information, while the Settlement envisages the creation of a single Book Rights Registry (BRR), centralising rights information.

The last aspect suggests that interoperability between the different resources is the key and thus the project results can be assessed through the number of metadata made interoperable within the Arrow system. Two elements are of particular importance:

1. the cross-domain interoperability, i.e. the capacity of the system to exchange information with different types of resources (libraries catalogues, books in print databases, RROs repertoires);
2. the trans-national interoperability, i.e. the capacity to set the basis for a genuine pan-European system

As said, partners and liaison organisations will provide several million records to be made interoperable for the project purpose. The presence of organisations coming from the different

domains meets the first requirement, while the involvement of partners and/or supporters in 13 European countries (up to now) is the prerequisite for the second.

In terms of "figures", the measurability of the project performances will be possible only after the first release of the system. At the current stage, it is possible to analyse the project impact in term of its capacity to influence the debate occurred last year in the international book world. In the trade off between the two models of the European MoU and the US Settlement, the existence of Arrow as a demonstrator of possible implementation of the European model was very often recalled. On this respect, Arrow has been seen as counterpart of the BRR (though this view is imprecise). Two episodes can be cited as key examples of this type of impact: during the public hearing on the Google Settlement organised by the European Commission, Arrow was mentioned by many speakers, also beyond the project partnership, as alternative reliable model to the Settlement (Sept 2009); and even in the "Fairness hearing" at the US Court (Feb 2010), Arrow was cited as the correct tool to deal with European books, which now are excluded from the Settlement.

Sustainability

First of all, Arrow is a system to facilitate diligent search of rightholders in large-scale digitisation programmes, developed by libraries, other institutions or commercial companies.

This defines a value proposition that will be the basis for the design of long term sustainability of the project. From the users viewpoint, "facilitating diligent search" means to reduce its cost. Therefore, the actual launch of digitisation programmes including copyrighted works define the demand for the Arrow services. Since such programmes are announced in Europe as a key priority, there should not be doubts about the existence of such demand.

How to transform this demand in a concrete business model that ensures long term sustainability of the Arrow system is a task that the consortium is currently working on. Some questions are still to be answered. The most important is: when and where (in which European countries) the demand will emerge on sufficient scale to make the project sustainable?

Currently Arrow is piloting its service in the countries where this demand is more mature. However, the objective is to have, in the long run, a real pan-European service. The increase of the number of countries where the project will be fully up and running is a step by step process. The project is designed to be scalable on this respect, through a precise definition of the requisites that every country should meet to join. In theory, the system will be "ready to use" everywhere in Europe at the end of the project, but there is awareness in Arrow that not in all the European countries it will be easy to meet the defined requirements. The system is based on the existence of sources of information that are not developed at the same level in the different member states. Additional work will be necessary to facilitate also countries that are not able to join now. So, there are technical constrains to be considered. However, the path to achieve a real pan-European dimension will be defined also considering the emerging demand, so to have coherence between technical development and sustainability model.

Arrow is expected to provide the communities of interest with additional value. In particular, the concept of creating a distributed "rights information infrastructure" is meant to foster the adoption of standards that are useful also for other reasons. In particular, the launch of the ISTC and (in the next future) the ISNI, can be seen as additional values of the project, and may create side-line business that can contribute to reaching the break even, though not substituting the main service.

Long term sustainability also implies the definition of a governance structure for the system, which will be defined in detail before the end of the funded project. The consortium strength is based on the participation of representatives of all users communities and stakeholders. This will be preserved in the long run, whatever the concrete form of the Arrow governance will be.