

**Grant Agreement ECP-2007-DILI-527003**

**ARROW**

# **Final Report**

**1<sup>st</sup> September 2008 – 28<sup>th</sup> February 2011**

[www.arrow-net.eu](http://www.arrow-net.eu)

<b>Deliverable number/name</b>	<i>D1.14</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>30<sup>th</sup> March 2011</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>AIE</i>



***eContentplus***

This project is funded under the *eContentplus* programme<sup>1</sup>,  
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.



# Table of content

**TABLE OF CONTENTS**..... 2

**PROJECT OBJECTIVES –**..... 3

**CONSORTIUM**..... 3

**PROJECT RESULTS**..... 4

**TARGET USERS & THEIR NEEDS**..... 12

**UNDERLYING CONTENT**..... 13

**SUMMARY OF ACTIVITIES** ..... 14

**IMPACT & SUSTAINABILITY** ..... 18

## Project Objectives –

ARROW (Accessible Registries of Rights Information and Orphan Works towards Europeana) is the European infrastructure of information resources for the management of copyright in support of national and international digital libraries initiatives.

ARROW was created to provide a practical technological solution to bridge the so-called “black hole of the twentieth century” in European digital collections and to facilitate inclusion in the collections of recent works that, being still under copyright, cannot be digitised without permission from their rights holders.

By creating a European network of certified sources of information (bibliographic databases and authority files of national libraries, books-in-print databases, together with repertoires of collective management organisations), ARROW will be able to determine whether a work is copyrighted or in the public domain, whether it is in-print or out-of-print and find the references of rights holders or collective management organisations (RROs) to be contacted to obtain permission to digitise, or declare that the work is an orphan.

The ARROW system created by the project consists of two core features:

**Distributed infrastructure of bibliographic and rights information resources:** it will serve primarily libraries in conducting diligent search for right holders, as well as other actors, public and private, interested in building collections of digital content.

**European Registry of Orphan Works:** when rights holders cannot be traced, the work will be included in a registry of orphan works accessible to authors and publishers so that they can claim authorship

The collaborative approach that sees all stakeholders (libraries, collective management organizations, right holders) 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 set up met the following criteria:

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

ARROW involved, 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 the ARROW contracting partners, several national organisations became official supporters of the project so expressing their support for ARROW objectives and contributing to project activities and results.

Partners and supporters from 13 countries of the European Union were actively committed to the project: Italy, France, Germany, Spain, United Kingdom, Austria, Slovenia, The Netherlands, Norway, Finland, Denmark, Belgium and Sweden.

**Libraries** - National libraries in Spain<sup>2</sup>\*, 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 the UK.

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

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

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

The role of the different stakeholders represented in the consortium guaranteed that the rights information infrastructure could benefit from the best information sources available to facilitate identification of right holders and right status of a book. Libraries, publisher organisations and RROs are key metadata providers, respectively of bibliographic, publishers' and rights ownership metadata. At the same time, they act also as end users of the system. Libraries, in fact, 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 right holders; publishers (and authors) may have their contents available in the digital environment in full respect of copyright.

International organisations ensured that the project was known among communities of the different domains (libraries, RROs, publishers and authors) and the results could be shared and made scalable beyond the duration of the project.

A high level of expertise was 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, provided a solid framework for exploiting state-of-the-art technologies for the innovative services of Arrow.

## Project Results

The key result of the project has been the deployment of the ARROW system as a service to facilitate the identification of right holders (authors/publishers) and the identification of the rights status of works (with particular concern for orphan and out-of-print works). The system was launched in June 2010 in its first release and finally deployed in the second release at the end of the project in February 2011.

The set-up of the technological infrastructure was based on the implementation of the workflow that identified the necessary data providers and data flow to provide a comprehensive set of information on rights holders and rights status. In the following section, we illustrate the main features of the workflow and of the technological infrastructure that is based on it.

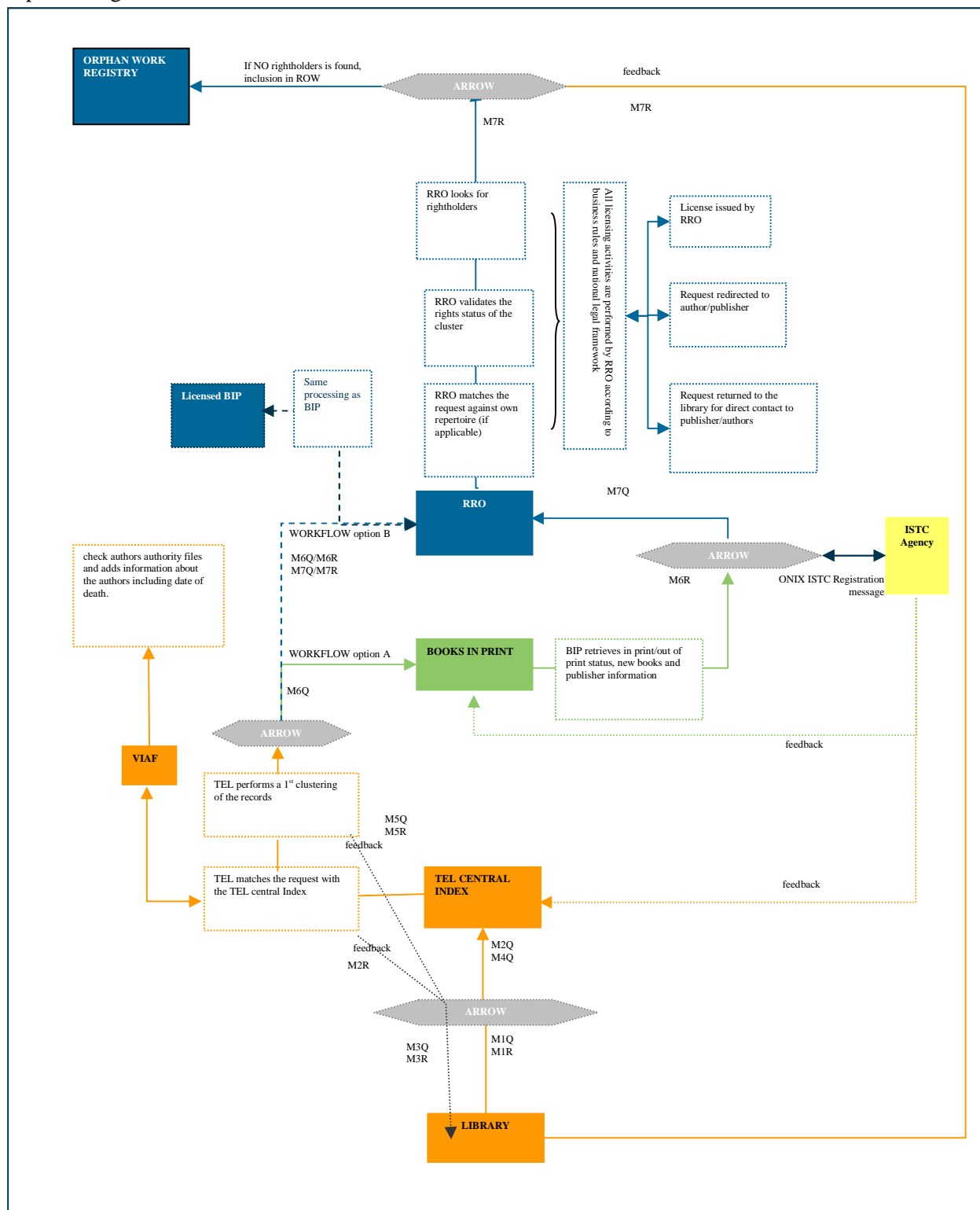
### The ARROW Workflow

---

<sup>2</sup> \* contracting partners

Information needed by the ARROW system to support the diligent search process is stored in the databases of different organisations for their own specific purposes; therefore the role of ARROW can also be defined as that of “interoperability facilitator” to access and query different IT systems, to retrieve the relevant data, to process and exchange these data with other systems and to add complementary data from other systems. In doing so, ARROW provides an open, standard-based network of resources where data not only are gathered but also enriched in successive processes to provide the best information available for the purpose of rights clearance for digitisation.

This network of rights information resources is clearly expressed in the following diagram representing the ARROW workflow.



Key actors in the ARROW workflow and therefore key nodes for information provision/exchange come from all three domains of stakeholders for ARROW and are:

*The European Library (TEL)*: coordinates access to the resources of European national libraries and the Virtual International Authority File (VIAF) and provides resource identification and clustering<sup>3</sup> services to ARROW

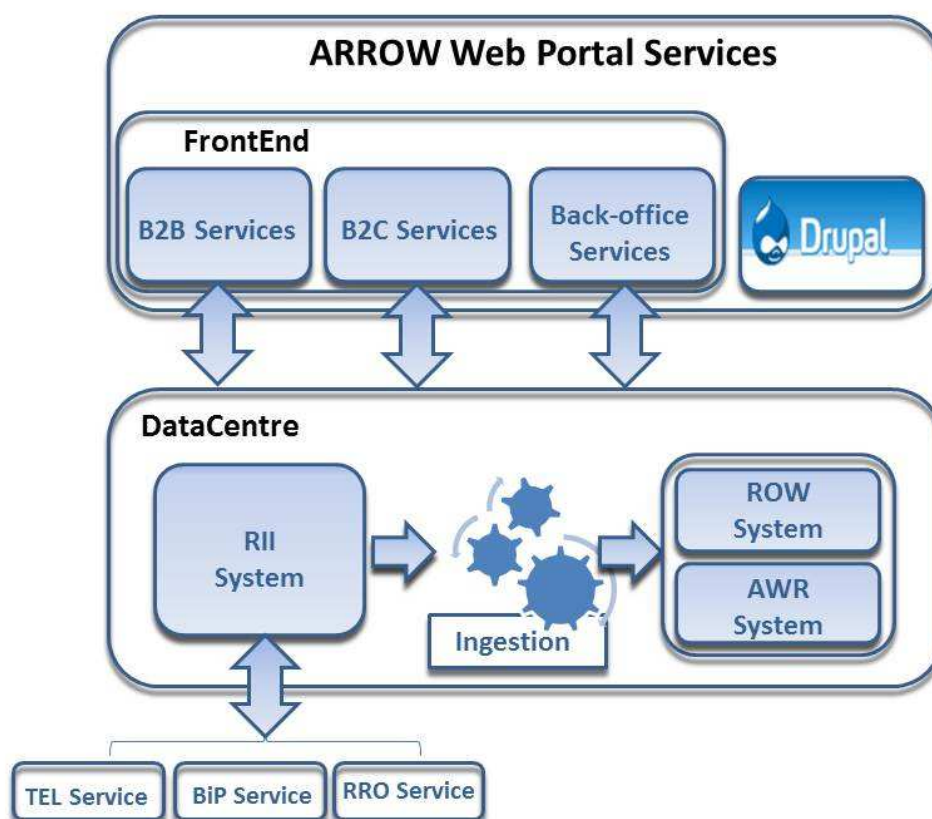
*Books-in-print agencies (BiP)*: match incoming requests against their own records and advise ARROW of any matching ISBNs as well as returning ONIX product records containing in-print or out-of-print status, availability and other metadata.

*Reproduction Rights Organisations (RROs)*: review the consolidated requests from the libraries and the information gathered throughout the ARROW process. On this basis, the RROs then proceed to grant or deny licenses, as appropriate, as well as offering further advice to ARROW and the libraries on how best to bring requests to complete resolution.

The ARROW system that implements the workflow and produces its outputs is made up of the following components:

- The Rights Information Infrastructure (RII)
- The ARROW Work Registry (AWR)
- The Registry of Orphan Works (ROW).

The figure below shows a schematic representation of the Arrow system. The results and the information collected during the RII workflow form the basis for the AWR and therefore for the ROW which is a subset of the above mentioned AWR.



<sup>3</sup> In the ARROW context, a cluster is defined as group of manifestations (books) that share the same underlying intellectual work.

### **The Rights Information Infrastructure (RII)**

Before going into a detailed description of the RII, it is worth mentioning here what we define as **rights information**.

Rights information is defined as a set of metadata that supports the identification of the rights status of a work. This includes

- (i) identification of the book concerned,
- (ii) identification of the work(s) included in that book,
- (iii) identification of the commercial status of a work,
- (iv) identification of the publisher, and
- (v) contributors' name, and finally
- (vi) the location of the right holders.

The **Rights Information Infrastructure (RII)** is at the heart of the ARROW system. The RII is the backbone and the engine that enables ARROW to query and retrieve information from a multiplicity of data providers, in multiple formats, to make the formats interoperable, to process this information and take decisions on the successive processes and finally to exchange information according to the workflow.

Building on the RII, the ARROW system receives a request for permission to digitise and use a manifestation of a work (for instance, a book) from a library and after querying the data providers included in the workflow and processing the gathered results, provides information on the work's rights status.

It is important to underline that the initial library request is performed at manifestation level (4), whereas the response at the end of the workflow is provided at work level. This means that the initial request passes through stages of identification and matching, work and manifestation clustering and the identification of related works and manifestations; each process adds a piece of relevant information towards the identification of the rights status of the work.

To simplify the complexity of the system, the workflow can be divided into three main processes corresponding to the three domains involved, each made up of further processes that contribute to the output. Each process is supported by a well-defined set of ARROW messages (5) that has been developed with the standards organisation Editeur.

The first main process takes place in the library domain and involves The European Library (TEL) as the main actor and the Virtual International Authority File (VIAF) as the source of authors' information. The output of this process is:

- the work to which the original library manifestation belongs
- a list of manifestations that share the same work with the original library manifestation
- any other related work and the list of respective manifestations

---

<sup>4</sup> To be more precise, the initial library request refers to a "resource", where the term "resource" identifies an instance of a manifestation, for example a particular copy of a printed edition of a book. For more information about terms used in ARROW, see *D4.3.2 ANNEX II ARROW Glossary of terms* available for downloading in the Resources area of the ARROW website ([www.arrow-net.eu](http://www.arrow-net.eu))

<sup>5</sup> For a detailed description of the ARROW message suite, see *D4.3.2 Specification for metadata messaging formats* available for downloading in the Resources area of the ARROW website ([www.arrow-net.eu](http://www.arrow-net.eu))

- a set of authoritative information for each author and other contributor of each work, including preferred and alternative forms of their names, their dates of birth and death and their nationality (6)
- the copyright status of each work: whether the work is in the public domain or copyrighted or whether this information cannot be asserted with certainty

The second main process takes place in the Books in Print domain and involves BIP organisations or databases in each of the countries included in the ARROW system. It adds further information to the output obtained from the previous process in the library domain.

Outputs of this process are:

- a list of additional manifestations belonging to the work and related works
- the in-print/out-of-print status and the commercial availability of each manifestation belonging to the work and related works
- the publishing status of each work: whether the work is currently active (in-print) or currently not active (out-of-print) or whether this information cannot be asserted with certainty (7)

The third main process takes place in the Reproduction Rights Organisation domain and involves RROs organisations or databases in each of the countries included in the ARROW system. It adds further information to the output obtained from the previous process in the library and BIP domains. Outputs of this process are:

- a set of information provided by the RRO concerning licensing conditions and reasons supporting the decisions
- the orphan status of the work: whether the work is to be considered probably orphan as its right holders cannot be identified or traced, or not orphan or whether this information cannot be asserted with certainty (8)

As a result of the above-mentioned three processes, the following pieces of information have been retrieved in the message exchange by the end of the ARROW workflow:

- Work information
- Manifestation information
- Relation between each manifestation and the work they belong to
- Relation between works
- Authors and other contributors information
- Relation between each identified author and the work they have contributed to

---

<sup>6</sup> This information is retrieved from the Virtual Authority File initiative (VIAF), considered as the most authoritative source of information of this kind

<sup>7</sup> As the publishing status of the work is deduced algorithmically by ARROW from the information available about in-print, out-of-print status and commercial availability at manifestation level (retrieved by BIPs), there might be cases where the information is missing or may not be sufficiently reliable. These cases produce an “uncertain” publishing status

<sup>8</sup> As the orphan status can be determined only as a result of a diligent search, according to the HLG principles, there might be cases where the search done via ARROW is considered insufficient and needs to be further carried out. These cases produce an “unspecified” orphan status.



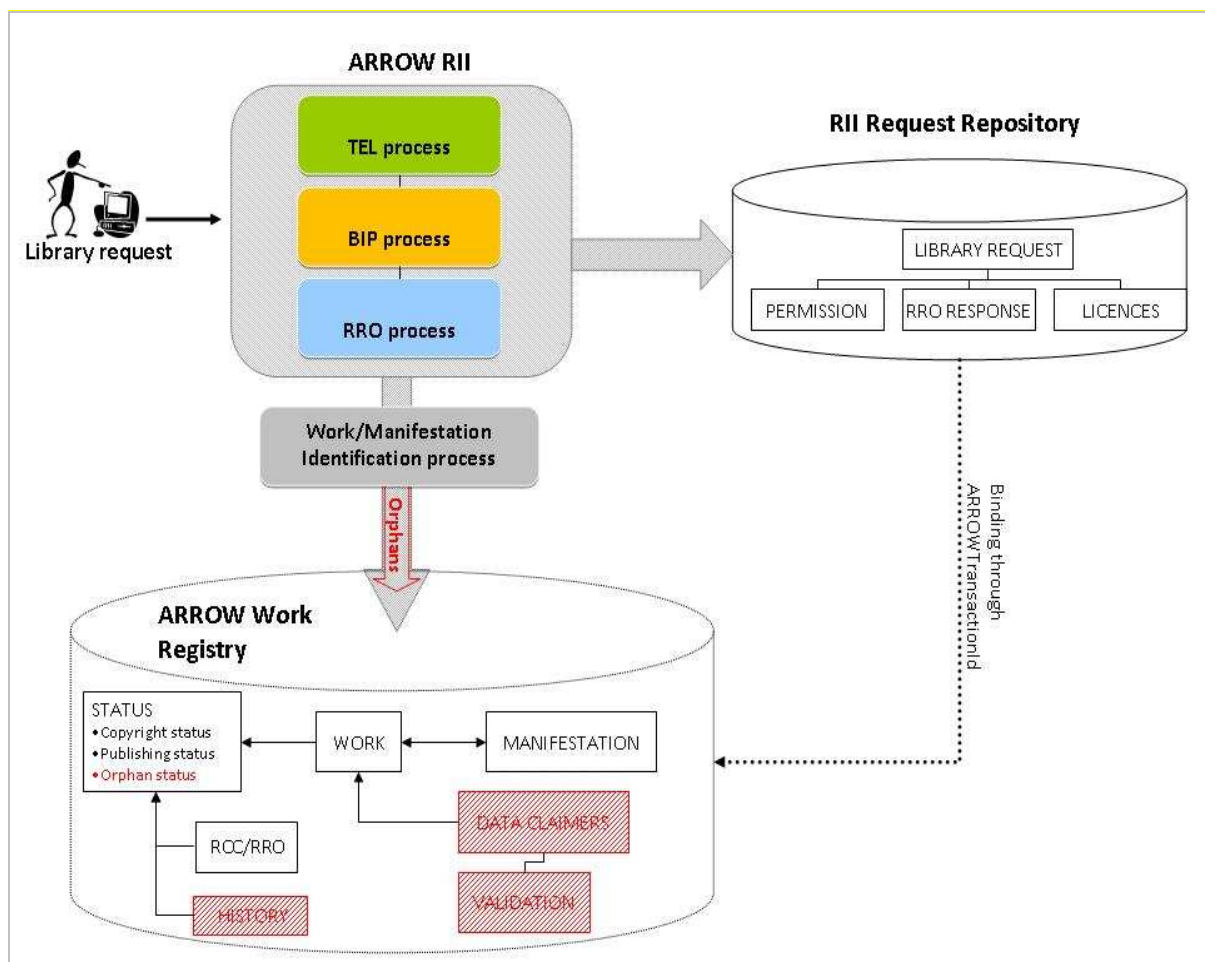
- Relation between each piece of information (work, manifestation, author) and the reference source that provided that information (TEL, VIAF, BIPs, RROs)
- A set of so-called ARROW assertions on each work: Copyright Status, Publishing Status and Orphan Status

Concerning the management and storage of the information, the initial library request, including the permission request, the information gathered and inferred during the TEL and BIP processes and the RRO answer, are stored in the RII repository.

The ARROW Work Registry (AWR) stores and maintains all these pieces of information for every request processed by ARROW.

The Registry of Orphan Works (ROW) is based on a subset of the AWR meeting specific criteria, that will be made publicly available to specific categories of users for specific purposes.

The figure below provides a high level overview of the AWR and its relation to the RII, where also the definition of the ROW as a subset of AWR is evident.



It is worth noting that the ARROW workflow produces ISTC-ready packages of information, meaning that the system is already compliant and ready to respond for future registration of ISTC<sup>9</sup>.

### **A system stemming from stakeholders requirements: ARROW and HLG guidelines for diligent search**

It is important to underline that the ARROW system implements in its workflow and data sources, principles included in the guidelines on due diligence criteria for orphan works (hereinafter referred to as “the guidelines”), as agreed by the EC i2010 High Level Expert Group on Digital Libraries (HLG). A detailed analysis has been conducted, matching the steps and data sources in the ARROW workflow with the principles of the HLEG guidelines.

According to the guidelines, the procedure for due diligent search should be based on a number of principles. The search is done prior to the use of the work; the search is done title by title or work by work; the relevant resources would usually be those of the country of the work’s origin.

The ARROW workflow complies with all of these principles.

In terms of the suggested resources, ARROW covers a large part of those listed in the guidelines, both common and sector-specific, and definitely the most relevant ones.

The ARROW workflow also facilitates compliance with the recommendations on document searches, as the system implements mechanisms to store information, including records of the searches performed. Furthermore, ARROW enhances the adoption of standards (identifiers and metadata) to foster interoperability and facilitates the access to authority files and the use of authoritative data on personal and organisational names and dates. It thus indirectly supports the adoption of some of the measures put forward in the guidelines to prevent future orphan works.

In sum, the ARROW workflow complies with all of the principles and most of the recommendations of the HLG regarding diligent search. It can therefore be considered as a valuable tool for libraries - and other organisations - that need to retrieve information on rights and right holders for digitisation initiatives.

### **ARROW performing live: pilot countries and validation of the system**

For the deployment of the system following the first release, four pilot countries were identified - Germany, UK, Spain and France. For these countries, data providers needed in the ARROW workflow were connected to the ARROW core system so that the workflow could be completed in all the required steps on a country basis, in keeping with the HLG principles (relevant sources of information for ascertaining rights status are those of the country’s work origin).

Based on the infrastructure developed, it will be possible to extend the Arrow Rights Information Infrastructure (RII) to further European countries, including BiPs and RROs organisations.

The deployment in the four countries also created the basis for an extended validation of the system aimed at measuring the performance of ARROW in terms of:

- progress in system implementation

---

<sup>9</sup> The International Standard Text Code (ISTC) is a numbering system developed to enable the unique identification of textual works. For more information on ISTC , <http://www.istc-international.org>

i.e. the progress in data quantity (made interoperable and accessible from one single point, also known as the ARROW query system

- level of accuracy

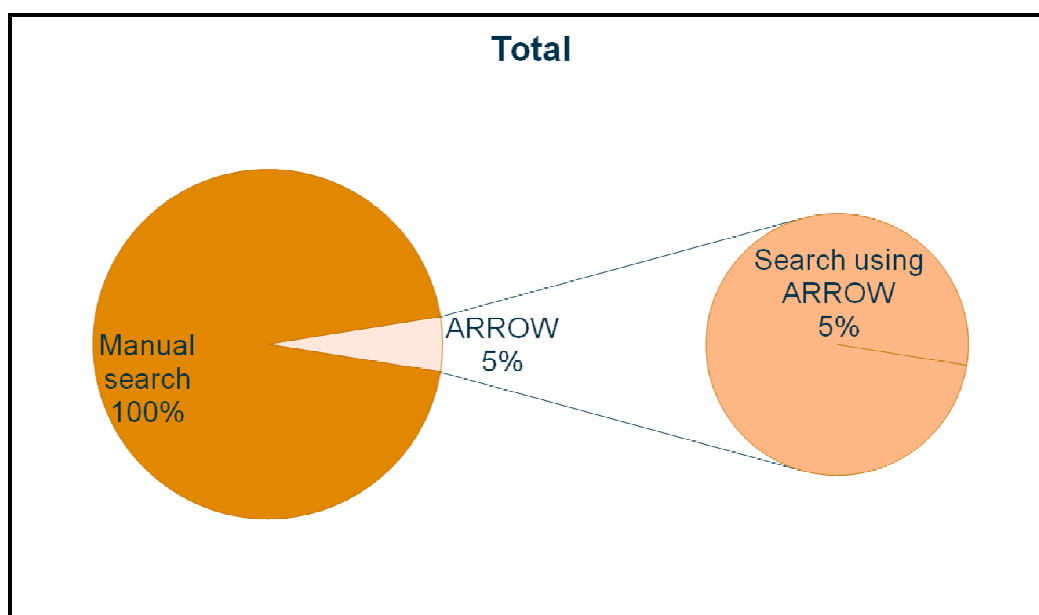
The accuracy of results provided by the system was measured in terms of enhanced performance of matching and clustering processes<sup>10</sup>

- time saving of ARROW versus manual diligent search

A comparison was made on performance and results of diligent search by processing a selected sample of bibliographic records in each of the four countries through the ARROW system and conducting the search manually according the usual practice of the library.

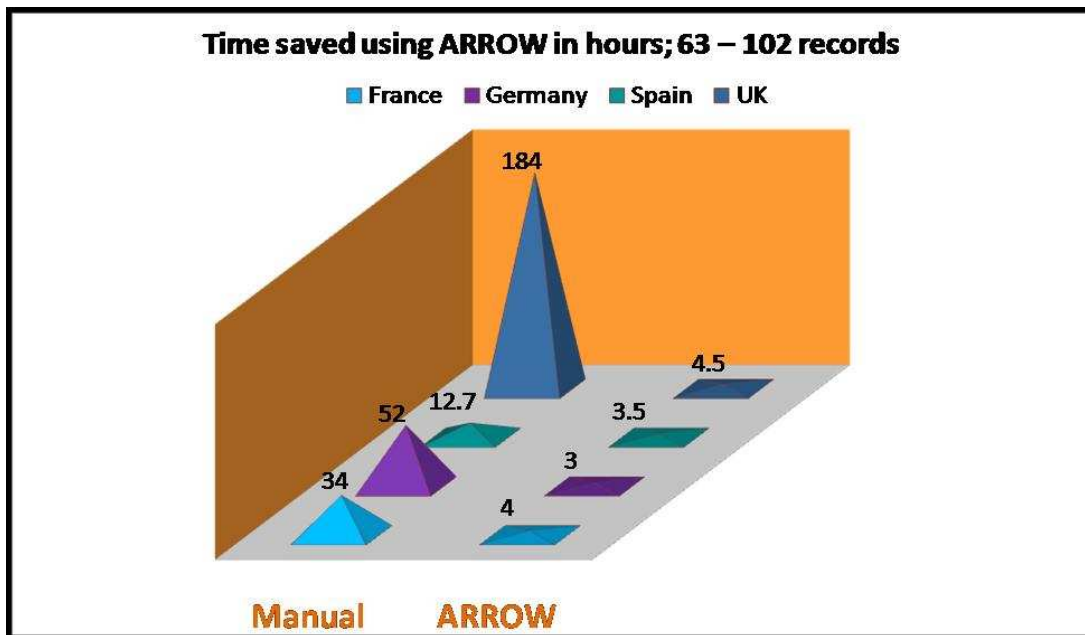
The results were positive overall for all the three indicators. There were positive results for the first two indicators, showing that the ARROW system succeeded in establishing a solid network of data sources. The system was therefore able to gather and produce a critical mass of data relevant to rights status determination, i.e. bibliographic data, data of the main parties (authors' authority files and publishers' data) and rights ownership data.

On the third indicator, the time saved in diligent search by using the ARROW system against manual search showed the advantages that ARROW could provide for any digitisation programme. The chart below illustrates the extent of these advantages; the process of a bibliographic record automatically retrieved through submission of a query to the ARROW system took about 5% of the time employed in average for conducting manual diligent search for the same record.



If we consider each pilot country for their specific sample of records we have the following results:

<sup>10</sup> Matching is defined as the ability to identify the book by matching the query record from the library with record(s) stored in the queried resources ; Clustering is defined as the ability to provide the publishing status of works (in-print /out-of-print status) by clustering the individual editions stored in the queried resources



The results of validation therefore confirmed the efficacy of ARROW as a tool to foster digitization programmes through efficient management of rights information, thus providing an essential service to libraries and in general to all those interested in enlarging the offer of digital content.

### 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. Several kinds of players can have a need for a system like ARROW for their digitisation initiatives. Their motives include preservation and making available cultural heritage for libraries, profit for commercial entities, either linking directly to sales platforms or to the possibility to offer a wider range of results to users performing internet searches.

Here is a summary of target users that could benefit from ARROW and of their needs.

Target user description	Needs
Right holders	<ul style="list-style-type: none"> <li>✓ To offer their content in the new environment</li> <li>✓ To maintain control over the content</li> <li>✓ To receive remuneration from use of their content also through the declaration of rights on orphan works</li> </ul>
Rights holders' representatives and/or agents (RROs)	<ul style="list-style-type: none"> <li>✓ To offer new value-added services, in particular, rights clearance</li> </ul>
Libraries	<ul style="list-style-type: none"> <li>✓ To reduce costs in rights acquisition and thus include more content for the same amount of money</li> <li>✓ To promote inter-operability for e-content</li> <li>✓ To avoid duplication of efforts in digitisation</li> </ul>
e-retailers and other intermediaries	<ul style="list-style-type: none"> <li>✓ To create commercial supply of e-content collections of copyrighted works</li> <li>✓ To provide services to rights holders</li> <li>✓ To reach new potential markets</li> </ul>

A more detailed analysis of potential users for ARROW has been done in the second edition of the “Report on Business Models” produced in August 2010 and further analyzed in the “ARROW Business Model” report produced in February 2011. Both documents are available on the ARROW website .

These documents confirmed that libraries and national libraries in particular are still seen as the main potential users of the ARROW system, including the analysis of concrete digitisation plans, issues of funding and inclusions of copyrighted work in digitisation plans.

On the other hand, the number of private organisations interested in providing commercial offers or access to content in the form of search results is increasing in parallel with the recent explosion of the e-book market. Search engines, e-retailers and electronic devices producers cannot but benefit from an ARROW service that paves the way for an increased offer of content to customers once it provides a tool for comprehensive searching on rights and rights holders.

In this scenario, PPPs are becoming very important in the field of digitisation, allowing public and private players to gain reciprocal benefits with the result of enlarging the availability of digital content. The value of Out-of-Print and Orphan Works risks being unexploited, if rights search and management is not administered in the correct way. RROs can have a role as far as they could be assigned the task not only to issue licences but also conduct search on rights holders and administer orphan works and orphan work registries. All these players, in their respective business models, share the need of a rights information service like the one provided by ARROW.

## 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 right holder(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 on the Internet).

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

The problems with the 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 has 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 databases and both are not interoperable with RROs databases;
- All data are created at “book” level, while rights information is defined at “work” level: in all the existing resources there are 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 on 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 step to clustering different editions of the same work. The library authority files for authors name are a key step for unambiguous identification of rights holders. 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 rights holders’ identification and existing mandates.

By the end of the project, the ARROW system ensured the interoperability of a critical mass of relevant data that satisfied the level of progress envisaged at the beginning of the project. The actual availability within the system of interoperable data in respect to targets set for the purpose of performance measuring is articulated as following:

Indicators	Actual Numbers	Expected Progress
<i>Data interoperable and accessible from single point</i>	M30	M30
<b>Bibliographic data</b>	<b>48,2 m</b>	<b>50 m</b>
Data sources: TEL (The European Library) Books In Print databases		
<b>Parties’ data</b>	<b>16,48 m</b>	<b>12 m</b>
(authors authority files and publishers’ data) Data sources: VIAF, Books In Print’s databases		
<b>Rights ownership</b>	<b>3 m</b>	<b>300,000</b>
(data on parties and rights managed in RROs repertoire)  Data sources: RRO (Reprographic Rights Organisations)		

## Summary of Activities

The activities carried out throughout the 30 months of the project can be grouped into three areas:

- Studies
- Technical set up and deployment of the system
- Dissemination

### Studies

During the project several studies and surveys were carried out.

Before the design and set up of the system, these studies were mainly aimed at providing a clear and current overview of the legal, business and technological environment where ARROW would operate once deployed, giving a solid background that guaranteed the compliance of the system not only with the needs of stakeholders but also with already agreed existing key principles for rights clearance in digitisation programmes.

For the same purposes, some of these analyses were updated mid-term so that any relevant development in digital libraries could be taken into consideration for its impact on the ARROW system.

The studies can be illustrated by distinguishing two main areas of analysis: legal-economic and technical.

#### Legal and economic area

- *Report on legal framework (1<sup>st</sup> Edition - July 2009; 2<sup>nd</sup> Edition – August 2010)*
- *Guidelines for clearance mechanisms for out of print works (April 2010)*
- *Guidelines for the definition of Orphan Works (April 2010)*
- *Report on business models (1<sup>st</sup> Edition - July 2009; 2<sup>nd</sup> Edition – August 2010)*

The first three studies provide a comprehensive overview of the EU legislative framework including a specific focus on the agreed principles and recommendations at European level developed within the European Commission High Level Group on digital libraries regarding orphan and out of print works. 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 is also available.

The last report, while illustrating business models in the e-book market, digital libraries and digitisation initiatives in Europe, aims at providing indication to ARROW to define its role in the current market situation and ensuring neutrality in respect to the different business models.

#### Technical area

- *State of the art and guidelines for standards applicable (1<sup>st</sup> Edition-July 2009; 2<sup>nd</sup> Edition July 2010)*
- *Analysis of bibliographic resources and clearing mechanisms existing in Europe (July 2009)*
- *Guidelines for technical interoperability (July 2009)*

The studies were aimed at preparing and supporting the system design with updated information respectively on:

- existing standards applied in the different communities (publishing, libraries, RROs) that must communicate in ARROW and whose databases should be interoperable.
- quality and quantity of data and metadata stored in databases for all the three domains concerned (libraries, rights holders, RROs)

Finally, the *Guidelines for technical interoperability* provided for each database in each single domain (libraries, RROs, publishers/rightsholders), the core requirements needed to achieve interoperability within the ARROW infrastructure.

Two more studies can be considered cross-subject since they compare the ARROW system with the principles agreed by EU stakeholders on diligent search and needs of stakeholders in terms of rights clearance for out of print and orphan works.

These studies are:

- *Analysis on compliance of ARROW workflow with the agreed HLEG guidelines on diligent search (July 2010)*
- *Analysis of relevance of ARROW with emerging clearing centers and users requirements (September 2010)*

The conclusions of these studies in particular are relevant since they proved how the ARROW system was built in accordance with the above mentioned principles and needs and therefore can be an effective tool to serve the development of digital libraries.

All these studies are available to the public on the ARROW website.

### **Technical set up, deployment and validation of the system**

After the preliminary studies were completed in the first phase of the project, the second part of ARROW saw a consistent effort in the technical set-up and deployment of the first and then final release of the technological infrastructure.

The work was carried out in steps: design of the system architecture and release of the system. In the first phase, the system workflow was outlined and translated into requirements and specifications necessary to build the infrastructure.

For the second phase, there were two main ARROW system releases: first release (May 2010) and second release (February 2011).

The main changes between the two concerned the continuous enhancement of the first two components (Right Information Infrastructure –RII and the Web Portal) as well as the creation of the last two (ARROW Work Registry –AWR and the Registry of Orphan Works - ROW). Between the first and second release of the ARROW system, each group of components passed through several intermediate releases that ensured continuous refinements of the system until the final release in February 2011.

The system became operational in all the four pilot countries (Germany, France, UK and Spain).

Despite the focus on four countries, the infrastructure was designed and implemented so that it could be easily scalable to further European countries including BiPs and RROs organisations.

The work on the first release included the completion and release of all the necessary messages (“Specification for metadata message formats”) that were refined for the second and final release. User related aspects were particularly relevant since the activities related to the definition of the system workflow, set up of system architecture and the work to define and implement standard messages saw a direct involvement of stakeholders in different working groups.

Numerous meetings and constant exchanges were the basis of the work conducted and ensured that the results were compliant with the requirements set by users and at the same time consistent with the overall objectives set by the project.

The same user-centred approach taken with regard to technical aspects, has been applied for the validation of the system.

The validation was implemented according to two phases, distinctive both methodologically and in terms of the schedule. An internal validation was conducted, testing the results in the pilot countries against the results of manual diligent search done by national libraries of the four countries. Selected test records were identified accordingly.



This phase was followed by an external validation where libraries outside the consortium tested the system with randomly chosen records focussing specially on usability and general features of the system.

The measurement of performance was also implemented through assessment of test records using established indicators such as “level of accuracy” and “time saving in diligent search” as described in the chapter “Project Results”.

### Dissemination

ARROW activities and results were constantly promoted during the project lifetime.

A website was launched with the aim of acting as reference point for stakeholders and the general public and a working place for ARROW partners and supporters.

The promotion of ARROW was carried out by the wide participation of project representatives in national and international events and a constant relationship with EU institutions. After the launch in September 2008, interest in the project grew among all stakeholders and the relevance of the project for the future development of digital libraries was widely acknowledged both in Europe and overseas, at professional and policy level.

ARROW was presented at the most relevant events for libraries, publishers, RROs and the general copyright community. In Europe, the project was on the agenda of IFRRO and FEP and was presented at International Publishers’ Association (IPA), International Federation of Libraries Associations (IFLA) and at the most important book fairs (Frankfurt, London) and libraries events (IFLA Congress). International organizations like ISBN International, International DOI Foundation, The Conference of European National Librarians (CENL) and EDItEUR, the international organisation coordinating development of the standards infrastructure for electronic commerce in the book and serials sectors, frequently referred to ARROW as a relevant project to follow. Other projects related to the digitisation of European heritage not only have been making references to ARROW but direct contact and exchanges were established with coordinators or specific working groups. Among the projects involved were: Europeana, Athena Europe, Mile Project and the European Film Gateway.

At country level, the range of partners and supporters joining the Consortium guaranteed the constant dissemination of project results and fostered dialogue between stakeholders about possible use of the system for forthcoming digitisation programmes.

Overseas, it is worth mentioning the interest in the project expressed by US organizations such as the Copyright Clearance Center (CCC) that provides international licensing expertise to organizations worldwide, OCLC, the computer library service and research organization that serves libraries in 170 countries and territories around the world, and CERLALC, the international organisation that, under UNESCO, gathers the Culture Ministries of eight Latin America countries.

Specific dialogue was established on the occasion of the so-called “Google settlement” with professionals on the side of Google and US Publishers’ and Authors.

EU institutions have looked at ARROW with particular interest since the beginning of the project. The Coordinator was invited to present the project directly to the former DG Information Society Commissioner Viviane Reding and her Cabinet and on the occasion of meetings of the Member States Expert Group on Digitisation (MSEG) and High Level Experts Group.

During the latter part of 2010 and early 2011, a number of demonstrations of ARROW were organized in Brussels, mainly aimed at showing the system functioning to various stakeholders. Among EC offices visited were DG Information Society, DG Internal Market and DG Education and Culture.

ARROW results were presented officially at a public event held in Brussels on 10th March with an audience of more than 70 people that included participants from various European Commission DGs, the European Parliament and EU Member States' governments.

The conference was opened by Mrs. Neelie Kroes, Vice President of the European Commission and European Digital Agenda Commissioner, who expressed her full support for ARROW.

*“One search in ARROW should be all you should need to determine the copyright status of a cultural good in Europe. If it were embedded in the forthcoming directive on orphan works, ARROW could become the official portal in Europe where you can find essential rights information and do automated searches of rightholders and copyright”.*

The conclusion of Vice President Kroes about the huge potential of ARROW and her praise for the project as an example of how stakeholders can work together towards the digital future paved the way for the next steps of the project through the follow-up initiative ARROW Plus.

## Impact & Sustainability

After 30 months and with the final release of the system up and running, it is possible to give an evaluation of ARROW's impact on the development of digital libraries.

ARROW was truly born as a collaborative project of all stakeholders, united by the common need to find a fair and effective solution to the issue of rights clearing for books digitisation with particular reference to out-of-print and orphan works.

It was the practical outcome of dialogue started by European stakeholders in HLEG on digital libraries that resulted in the model for dealing with rights in digital library initiatives, as set by the *Memorandum of Understanding* signed in June 2008.

This is the basis of the impact that ARROW can have, first of all, on the development of Europeana to become the first point of reference for European culture. Europeana needs to expand its coverage to copyrighted work and also find a solution to include orphan works in its offer.

Dr Elisabeth Niggemann, Chair of the Europeana Foundation, in her speech during the final event of ARROW in March 2011, clearly linked the need in Europeana of “stakeholders working together to build a technical solution and knowledge database to facilitate rights clearance process” with the service that ARROW can offer, thus indicating the ARROW system as a tool to foster the “New Renaissance” triggered by the digitisation of Europe's cultural heritage, as called for by the Comité des Sages, the EU High-level Reflection Group whose recommendations will feed into the Commission's Digital Agenda for Europe.

ARROW would also help to achieve the Digital Agenda's objective of building a common digital market by offering a neutral tool that can serve all players that are, at different levels and with different objectives and business models, interested in carrying out digitisation initiatives.

The comparison of ARROW's truly European approach, based on collaboration with stakeholders, with the US experience of the settlement agreement between Google and the US associations of authors and publishers can give useful hints to understand better the real influence that ARROW can have on fostering digital libraries, not only at EU level.

Curiously enough, the start of the project coincided with the announcement of the Settlement Agreement between Google and the Association of American Publishers (AAP) and Authors' Guild (AG) (that also included the creation of a Book Rights Registry –BRR- as an independent organisation

collecting rights information). The end of ARROW project saw the Settlement being rejected by the New York Court following a class action lawsuit.

It can be said that the different outcomes of the two experiences reflect the difference in the models applied.

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

In its conclusion, rejecting the settlement, the US Court recommended a move from an 'opt-out' to an 'opt-in' system thus implicitly implying a broadening of the tasks of the Book Rights Registry according to a model that is more similar to that developed by ARROW, based on pro-active search of rights holders rather than their claiming of rights in the works concerned.

### **Sustainability**

ARROW is a system to facilitate diligent search of rights holders in large-scale digitisation programmes, developed by libraries, other institutions or commercial companies. This defines a value proposition that has been the basis for the design of long-term sustainability of the project. From the user's viewpoint, "facilitating diligent search" means to reduce its cost. Therefore, the actual launch of digitisation programmes including copyrighted works will define the demand for ARROW services.

At the moment ARROW ends, real use cases are emerging from the countries where this demand is more mature and where ARROW was successfully piloted - in the UK, France and Germany. However, the objective is to have, in the long run, a real pan-European service. The increase in the number of countries where the project will be up and running is achieved by a step-by-step process. The project is designed to be scalable in this respect, through a precise definition of the criteria that every country should meet to join in.

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 all European countries will find it easy to meet the defined requirements.

The system is based on the existence of sources of information that are not developed to the same level in the different member states. Additional work will be necessary to facilitate the participation of countries that are not able to join now. So, there are technical constraints to be considered.

A sustainable mid- and long-term business model and governance have been prepared by the Consortium for the running of ARROW after the first project phase has been concluded. Then ARROW will enter a new phase (currently called ARROW Plus) during which the number of participating countries will be expanded to cover progressively all or the majority of EU Member States and EEA countries. The system set up in the initial phase will be enhanced.

Though several private players are entering the digitisation market, in the short and middle term, libraries and other similar institutions are expected to be main users of ARROW. These are required to digitise their collections and make them available through Europeana, with the aim to make 25 million objects available by 2014 compared to the 14 million at present. Other non-Europeana library digitisation projects and digitisation projects such as Google and Microsoft will have the same obligation to find rights information and clear rights.

For libraries, the benefits of the ARROW system include not digitising books that rights holders will later ask to be removed; the possibility to negotiate with rights holders for permission to make their works available; not digitising books already digitised by other libraries; and all the benefits accruing to increased legal certainty. On a different plane it will be clear that to have a diligent search is more respectful of the copyright principle, facilitates the digitisation process, and is convenient and economic.

Digitising and making available cultural heritage is principally the duty and task of public institutions and that is also where we currently find the majority of the digitisation projects which involve copyrighted works. Digitisation programmes are therefore also chiefly publicly funded initiatives.

After the finance from the European Commission for the ARROW Plus phase, a stable and sustainable flow of revenues to cover costs and a suitable governance model should be established. Although some of this revenue will become the core ARROW service and form its additional value, there is also a rationale for sustained public funding of the system. The ARROW services will enhance the EU's cultural policy and address key EU goals by facilitating digitisation initiatives and offering solutions to issues related to out-of-print and orphan works. Stable funding provided by the public sector will be required in the introductory phase and the first years of regular operation following the project periods.

In the long run, the European Commission and EU Member States should continue their contributions to the financing of ARROW through payment for the use of services so long as institutions make use of them. Governments which plan to fund national digitisation programmes should create a separate budget line for "diligent search" in general or fund ARROW directly in order to support the diligent search in institutions they govern.

Long term sustainability also implies the definition of a governance structure for the system. A new legal entity will be established as the "owner" and manager of the ARROW system. It will be a not-for-profit entity which should comprise the existing partners and be open to new members as well. This is in line with the participation of representatives from all the user communities and stakeholders that has been one of the strengths of the consortium and key to the success and future consolidation of an EU-wide system used in all EU countries.