

Grant Agreement ECP-2007-DILI-527003 ARROW

Annual Report

 1^{st} September $2009 - 31^{st}$ August 2010

www.arrow-net.eu

Deliverable number/name D1.10

Dissemination level Public

Delivery date 4th January 2011

Status Final

Author(s) AIE



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

-

OJ L 79, 24.3.2005, p. 1.



Table of content

1	TABLE OF CONTENT	2
2	PROJECT OBJECTIVES -	3
3		
4	PROJECT RESULTS/ACHIEVEMENTS	
5	TARGET USERS & THEIR NEEDS	8
6	UNDERLYING CONTENT	9
7	SUMMARY OF ACTIVITIES	10
8	IMPACT & SUSTAINABILITY	11



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 features:

- 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 Spain²*, France*, UK*, The Netherlands*, Germany*, Slovenia*, Finland*, and the University Library of Innsbruck, Austria*

_

² * contracting partners



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 key result of the period has been the deployment of the first release of the ARROW system.

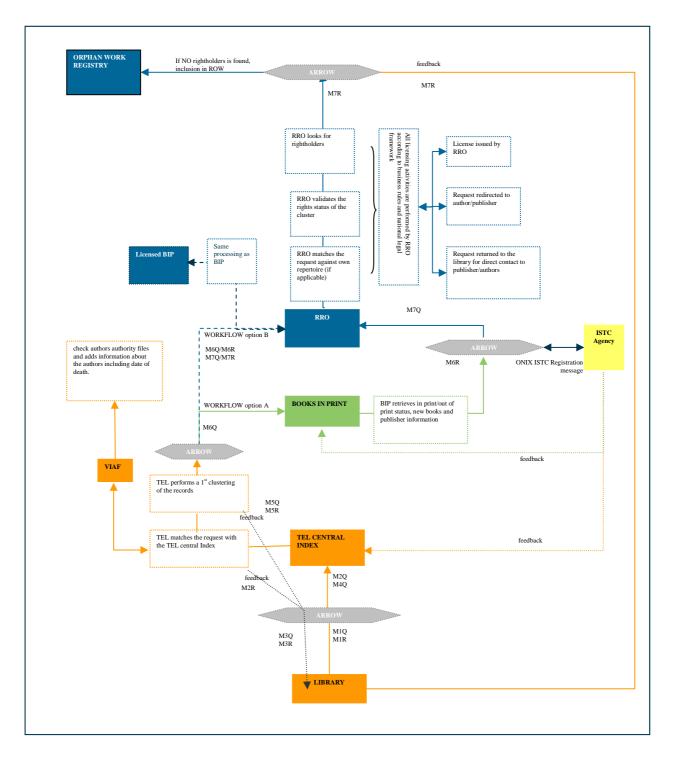
The system, as a service to facilitate the identification of rightholders (authors/publishers) and the identification of the rights status of works (with particular concern to orphan and out-of-print works) has been launched in june 2010.

The technological infrastructure implements the workflow outlined at the end of the first year of the project that identified the necessary data providers and data flow to provide a comprehensive set of information on rightholders and rights status.

Key actors in the Arrow workflow and therefore key nodes for information provision/exchange come from all the three domains of stakeholders for ARROW and are: The European Library (TEL) , the Virtual International Authority Files (VIAF), Books in Print databases, RROs repertoires.

In order to better understand the features of the ARROW system, it will be useful to recall the ARROW workflow, represented in the picture hereunder.





The ARROW System that implements the workflow and elaborate 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 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.

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 elaboration 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 elaborating the gathered results, provides information on the work rights

It is important to underline that the initial library request is performed at manifestation level3, 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 messages4 that has been developed with the standard organisation Editeur.

The first main process takes place in the library domain and involves The European Library (TEL) as main actor and the Virtual International Authority File (VIAF) as source for 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
- 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, their nationality5
- the copyright status of each work: whether the work is in the public domain or copyrighted or whether this information cannot be certainly asserted

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 and adds further information to the output obtained from the previous process in the library domain.

Output of this process is:

- 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 certainly asserted6

³ 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)

⁴ For a detailed description of 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)

⁵ This information is retrieved from the Virtual Authority File initiative (VIAF), considered as the most authoritative source of information of this kind

⁶ 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



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 and adds further information to the output obtained from the previous process in the library and BIP domains. Output of this process is:

- 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 their rightholders cannot be identified or traced, or not orphan or whether this information cannot be certainly asserted?

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

- 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
- 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

It is worth noticing that the ARROW workflow produces ISTC-ready packages of information, this meaning that the system is already compliant and ready to respond for future registration of ISTC.

As for what concerns 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.

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, respecting specific criteria, that will be made publicly available to specific categories of users for specific purposes.

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

might be cases where those information are missing or do not ensure enough reliability. Those cases produce an "Uncertain" publishing status

⁷ 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. Those cases produce an "Unspecified" orphan status.



In terms of the resources suggested, 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 recommendation to 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. Therefore, it can be considered as a valuable tool for libraries - and other subjects - that need to retrieve information on rights and rightholders for digitisation initiatives.

For the deployment of first release of the system four pilot countries where identified (Germany, UK, Spain, 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 steps on country basis, in coherence with the HLG principles (relevant sources of information for rights status are those of the country's work origin).

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

In the period when the first release of the system was launched, dissemination about ARROW project changed focus from raising awareness to promotion of results, answering to the raising demand of information about the project coming from stakeholders.

The increasing number of events and meetings where Arrow was presented confirmed the large interest and positive feedback to the project by the different communities in the book value chain, in Europe and overseas, both at professional and policy level, while the website became a stable and constantly updated source of information about the project.

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	Needs	Country coverage
description		
Rightholders	✓ To offer their content in the new	All Europe through
	environment	FEP + some
	✓ To maintain control over the content	countries directly as
	✓ To receive remuneration from use of their	partners (IT, ES,
	content also through the declaration of rights	DE) or supporters
	on orphan works	(FR, UK, SE)
Rightholders	✓ To offer new value added services, in	All Europe through
representatives and/or	particular rights clearance	IFRRO + some
agents (RROs)		countries directly
		(UK) or as liaisons
		(ES, FR, DK, NW,
		FI, SL, NL)
Libraries	✓ To reduce costs in rights acquisition and	All Europe through
	thus include more content at the same budget	EDL Foundation+
	level	some countries as



	✓ To promote inter-operability for econtent ✓ To avoid duplication of efforts in digitisation	partners (FR, ES, SL, UK, DE, NW), external supporters (FI) or through the EOD network (PT,
		HU, LI, DK)
e-retailers and other	✓ To create commercial supply of econtent	LIBREKA in
intermediaries	collection of copyrighted works	Germany (MVB)
	✓ To provide services to rightholders	Numilog in France
	✓ To reach new potential markets	

A more detailed analysis of potential users for ARROW has been done in the framework of the second edition of the "Report on business models" produced in august 2010 and available at the ARROW website.

This report confirmed that Libraries and National Libraries in particular are still seen as the main potential users of Arrow system considering the analysis of concrete digitisation plans, though issues of funding and inclusions of copyrighted work in these plans can be raised.

On the other hand, the number of private subjects interested in providing commercial offer or access to content in the form of search results is increasing in parallel with the recent explosion of the ebook market. Search engines, e-retailers and electronic devices producers cannot but benefit by the Arrow service that paves the way for an increased offer of content (information) to the customers once it provides a tool for comprehensive finding on rights and rightsholders. In this scenario, PPPs are becoming vey important in the field of digitisation allowing public and private players to gain reciprocal benefits with the common result of enlarging availability of digital content.

In a framework of this increasing digitisation need, the value of Out of Print and Orphan Works risk to be 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 rightholders and administer Orphan works and orphan work registries.

All the players above, in their respective business models, share the need of a rights information service as 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 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 second part of the project can be described according three areas of activities:

- technical set up and deployment of first release of the system
- analysis of business and legal framework
- user related aspects

Technical set up and deployment of first release 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 release of the technological infrastructure.

Starting from the definition of the workflow identified in the previous phase, the infrastructure was built foreseeing a progressive inclusion of the pilot countries.

The first phase, called "Alpha release", was delivered in March 2010 and had as a result the Arrow system set up and running with all core services.

In this phase it was chosen to work on the German workflow, i.e with data flow and connectors limited to the German BiP and RRO stakeholders.

This phase allowed the partner in charge of technical development, CINECA, to investigate the feasibility and results of the deployment.

The second phase concerned the extension of the functionalities in the Arrow Front End and DataCentre and the inclusion of more data providers from pilot countries.

The output of both phases is illustrated in the deliverable D6.1 Rights Information Infrastructure – 1st release

The second phase ended in May 2010 with the inclusion of the UK and Spain while France would follow shortly after. Based on this infrastructure it will be possible to extend the Arrow Right Information Infrastructure (RII) to further European countries including the respectively BiPs and RROs organisations. The work on the first release included the completion and release of all the necessary set of messages ("Specification for metadata message formats")

Analysis of business and legal framework

During the period concerned the studies on legal framework and business models (WP3) were updated in line with the project's needs in terms of defining the potential use scenarios of Arrow after the project duration and the legal framework that constitutes the project's background, as well as identifying the legal and market developments that could have an impact on Arrow.

In addition to this, a thorough analysis was performed on the compliance of the Arrow system with the HLEG guidelines for diligent search and on the correspondence of the system infrastructure with clearing centres and the needs of stakeholders. The first allowed to evaluate (positively) the



compliance of the Arrow system to principles for diligent search that have been already validated by the wide community of different stakeholders within the EC framework and therefore demonstrate to be the first concrete tool that can serve those principles. The second analysis focused on usefulness/compliance of Arrow system with needs of stakeholders and in particular with existing "clearing centres" and proved to be positive for the evaluation of Arrow in terms of its effectiveness in providing concrete solutions for the right clearance of orphan and out of print works.

Two more documents prepared were "Guidelines for the definition of Orphan Works" and "Guidelines for clearance mechanisms for out of print works", both based on the work of the EC i2010 digital libraries High Level Experts Group (HLEG) chaired by Commissioner Reding and its copyright subgroup which addressed specifically the two issues.

All these analyses are available to the public in the ARROW website dedicated section of resources.

User related aspects

were particularly relevant since the activities related to the validation of the system workflow (WP5) set up of system architecture (WP6) and the work to define and implement standard messages (WP4) saw a direct involvement of stakeholders in the different working groups. Numerous meetings and constant exchanges were at 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-centered approach taken with regard to technical aspects, has been applied for the validation work that also started.

The Validation (WP7) has been planned in methodology and schedule according two phases; an internal validation will be conducted testing the results in ARROW implementation of pilot countries against the results of manual diligent search done by national libraries of the four countries. Selected test records have been identified accordingly.

The second phase consist in an external validation where libraries external to the consortium will test the system with randomly choses records and special focus on usability and general features of the system. In the period considered, the first activities performed concerned in particular performance measuring based on tests records that will be assessed in terms of "level of accuracy" and "time saving in diligent search" and communication with rightholders and collection of relevant data for manual diligent search concerning the complete set of test records previously identified.

On the other side dissemination (WP2) continued to be enforced through the promotion of ARROW in a large number of events both at national and international level while the website was constantly updated with recent outputs of the projects including public reports and news.

The increasing number of events and meetings where Arrow was presented confirmed the large interest and positive feedback to the project by the different communities in the book value chain, in Europe and overseas, both at professional and policy level.

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.

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 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 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) will the demand emerge on a 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. 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 the participation 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 as 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. For example, 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 businesses 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 user communities and stakeholders. This will be preserved in the long run, whatever the concrete form of the Arrow governance will be.