# A European solution to rights information management:
## the ARROW project

Louise Edwards [*]

A RROW[1] stands for Accessible Registries of Rights Information and Orphan Works. ARROW will establish a single framework to manage rights information, supporting the development of Europeana[2] – Europe's digital library.

The origins of the ARROW project stem from high-level discussions that have taken place among all the relevant stakeholders within Europe, including libraries, publishers and collective rights organisations. The vision they share is to maximise access to digital content by the user.

All of Europe's libraries aim to increase access to digital materials. To achieve this, libraries need to know the copyright status of works they are planning to digitise and make available publicly. ARROW addresses the problem of the "*black hole*" in digital libraries – works that are impossible to digitise because their rights status is unclear.

A major challenge that must be overcome is the significant fragmentation of the rights information infrastructure that exists at present. The project proposes to create a seamless service across a

ARROW website

distributed network of national databases containing information about the rights status of works. This infrastructure, once established, will provide valuable tools for libraries and other organisations to contact rights holders in seeking copyright clearance for the use of content.

The other key strand of the ARROW project is to increase the level of interoperability between public domain and commercial content, leading to innovative business models for the exposure and use of digital content.

The project focuses on books, but could be extended to other types of material in the future, such as journals, newspapers, images, archives and so on.

* The European Library.
With contributions from Sally Chambers, The European Library;
Elisabeth Freyre, The French National Library; Thomas Jaeger,
The German National Library; Hannah Jenkins, The British Library;
Esperanza Ruiz de Velasco, The Spanish National Library
and Sonja Svoljšak, The Slovenian National and University Library.
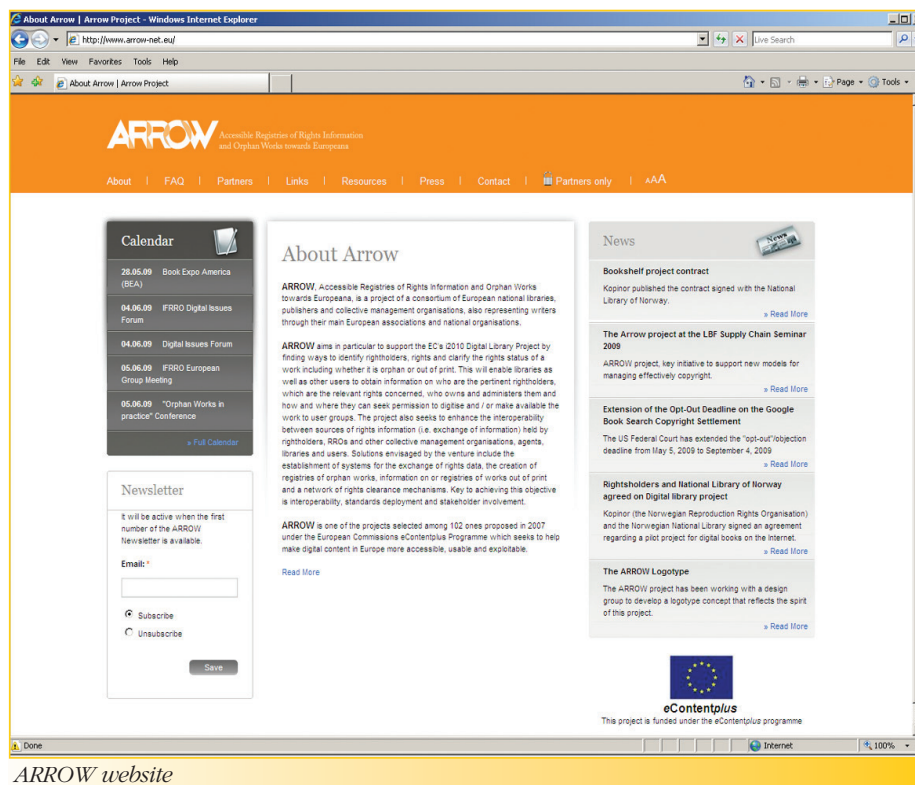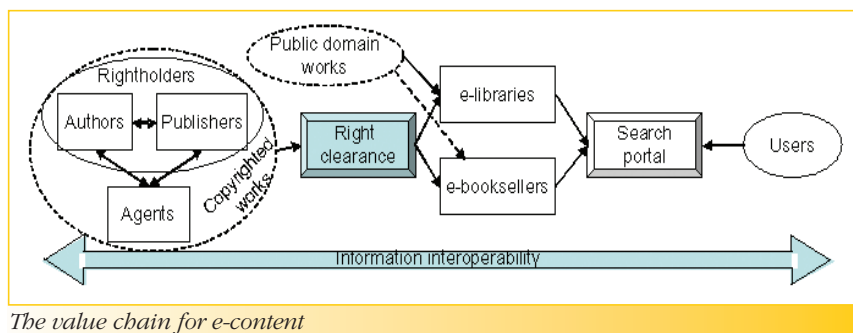
The value chain for e-content

The project launched in autumn 2008 and will last for two and a half years. It is funded under the European Commission's eContent*plus* programme,[3] which seeks to make European digital content more accessible, usable and exploitable. ARROW provides tools and information to support policy. Its remit does not include the setting of policy itself.

Background. Many of the issues in the ARROW project have been addressed by the European Commission's High Level Expert Group on Digital Libraries.[4] In 2008, the Copyright Subgroup presented a significant report on the problems of digital preservation, orphan and out-of-print works.[5] ARROW's work is very much in line with the recommendations of the Group.
The starting point is to define precisely what we mean by "orphan" and "out-of-print".
The Group defined an "orphan work" as one where "the rights holders cannot be identified or located." It is recognised that a huge number of printed works are considered to be "orphans". The problem for libraries is the barrier this presents in digitising the works for future preservation and access by users. The solution proposed by ARROW is to create a registry of information on orphan works.
The Group defined an out-of-print work as one that is "commercially not available, as declared by the appropriate rights holders". A significant proportion of works published during the twentieth century are out-of-print. In order to digitise them, each title has to be negotiated and contracted separately – an onerous operation

for libraries. The solution proposed by ARROW is to establish a model licence, a database of out-of-print works, a clearance centre and a procedure by which to clear rights.

A consortium of the major stakeholders in rights information management.
What distinguishes ARROW is that it brings together European national libraries, publishers and collective rights organisations into a consortium. It therefore represents the views and interests of all of the main stakeholders in the book value chain. The project has 16 primary partners from 10 European countries. They include 7 national libraries and significant international organisations such as the Federation of European Publishers (FEP) and the International Federation of Reproduction Rights Organisations (IFRRO), as well as leading technology partners. The European Library as the aggregator of National Library collections in Europe is also actively involved in the project.

New approaches to business models for e-content. The project approach is based on the value chain for discovering and using electronic content. Users look for content
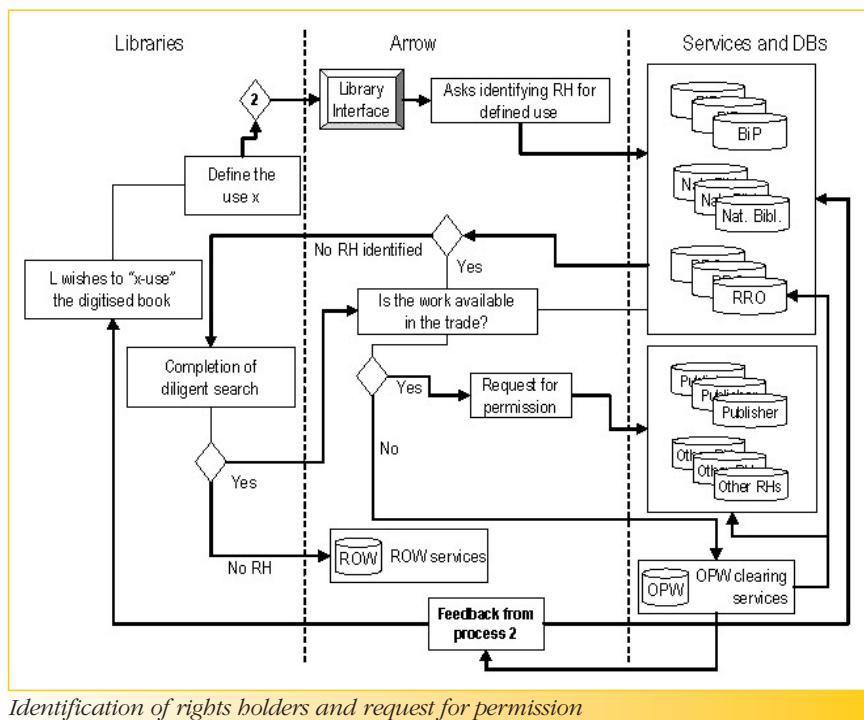
Identification of rights holders and request for permission

through search portals. Such content may be in a library or offered by a commercial provider. The content may be in copyright or in the public domain.

The principle of ARROW is that public domain and commercially-licensed materials co-exist within digital libraries and that the user should be able to discover and access both. Digital libraries should not only deal with free access to public-domain works but should also facilitate access to copyrighted works offered by commercial publishers.

The creation of shared services will enable both rights holders and libraries to exploit the "long tail" of books that have low demand but are still of value to scholars and the general citizen. Digital libraries can be a point of access to these works, directing users to their own electronic collections or to e-retailers.

Solving interoperability issues. A library wishing to digitise a work has to go through a number of steps:

– to identify the underlying work incorporated in the book to be digitised;
– to find out if the underlying work is in the public domain or in copyright, an orphan work or out-of-print;
– to describe clearly the use that is requested for the book, such as digitisation for preservation, electronic document delivery etc.;
– to identify the rights holder(s) or their agent, such as a collecting society;
– to seek the appropriate permission.

Not surprisingly, libraries complain that the procedure is longer, more complex and more costly than actual digitisation of the work itself. ARROW will address the interoperability of rights information along the whole value chain for digital libraries. The infrastructure will support the identification of a work, clarification of its rights status and the identification of rights holders. The project will enhance the interoperability of rights information, facilitating the flow of information between rights holders, agents, libraries and users.

The creation of this rights information infrastructure depends on the availability of existing bibliographic data and rights information. The partners and other organisations will provide several million records for this purpose. There is already an established and generally well-regarded information infrastructure for print material, through national bibliographies, books in print and the databases of rights organisations. The digital world allows us to join up these sources into a distributed but seamlessly accessible network of information. Currently, these sources are not interoperable because of differences in data collection policies and adopted metadata schema. Bibliographic databases rarely include metadata about rights ownership and usage policies. Such information is usually held in a wide array of formats by publishers, collecting societies and authors.

Building on existing interoperability.
Bibliographic data stored in the catalogues of Europe's national libraries will be one of the key data sources in the ARROW system. This is an area where the ARROW project can build on existing interoperability achieved through the development of The European Library.[6] Launched in March 2005, The European Library is a free service that offers a single point of access to

The European Library website

the bibliographical and digital collections of the National Libraries of Europe. By the end of 2009, 46 of the 48 national libraries in Europe – including the 7 national libraries who are partners in the ARROW project – will have included their collections in The European Library. As this bibliographic data is already being aggregated through The European Library, the ARROW project intends to explore how it could be used as a core source of library data within the ARROW technical architecture.

Standards. ARROW will not develop new standards but will use existing, well-established ones. However, the standards landscape is evolving to meet the needs of the digital world and ARROW will develop application profiles within existing services and projects, working closely with the appropriate standards organisations. ARROW will define guidelines on standards that can be applied to the identification and description of content and for technical interoperability between distributed information sources.
As a further step in the process, once the status of a work has been identified, the rights will need to be cleared if the work is still in copyright. The project is testing out rights clearance models for orphan and out-of-print works. Rights clearance centres established nationally will take existing databases as a starting point and interlink them. The result will be a European network of national rights clearing centres.

From a national to a European dimension.
National pictures vary but the shared view is that a European-wide solution is needed. Given that the rights landscape varies from country to country, a distributed solution will provide the best means of addressing this complex issue. It will define a rights information infrastructure that can be exploited both within and across national borders – a pan-European approach. The project recognises that rights information acquires a new complexity in the international arena, since the meaning of rights terms are expressed in different vocabularies and in different languages and take on different meanings according to national legislation or the legal language in each country. The project will therefore build on existing rights vocabularies. The project involves 10 countries and 9 European languages and is designed to be fully scalable and extendable across the whole of the European Union.

Serving the needs of users. ARROW is serving the needs of the end user by facilitating access to digital content. However, the ARROW resource is primarily a business-to-business service, for example, between rights holders (or their agents) and libraries.

Significant benefits to libraries. There are considerable benefits to libraries from the ARROW project. It will reduce the costs in clearing rights for the digitisation of works in library collections. In particular, libraries will be able to save costs on digitisation of orphan and out-of-print works, as well as directing their users to copyright material in commercially-provided collections.

Key phases of the project 2008-2010. ARROW has three main phases after the initial start-up. Phase 1 will gather data about the current information infrastructure, the use of standards, the current legal framework and business models used. Phase 2 will design and test systems and tools, create the registry of orphan works and the network of clearing mechanisms. Phase 3 will assess the systems and tools implemented within ARROW.

Towards the vision of Europe's digital library – Europeana. ARROW supports the European Commission's i2010 Digital Library[7] vision by providing the means to clarify the rights status of orphan and out-of-print works. These works can then be cleared for digitisation and inclusion within Europe's digital libraries, increasing the amount of electronic content available to users and reducing the "black hole" of twentieth-century European cultural heritage that is missing from today's digital library collections.

National libraries in the ARROW project.
7 national libraries within Europe[8] are playing key roles within the ARROW project. The European Library is representing the National Library of the Netherlands, where the service is based. The role played by 5 of the national libraries is described below.

BIBLIOTHÈQUE NATIONALE DE FRANCE
The Bibliothèque nationale de France (BnF) is one of the largest public and research libraries in the world today. Throughout its history it has always performed its task of collecting and conserving the national heritage entrusted to its care, in

whatever form, for the use of all researchers, students and professionals. Today its patrimonial collections encompass all areas of culture and knowledge in a great variety of languages and illustrate the library's encyclopaedic nature.

The BnF offers access to its digital library Gallica,[9] obtained through the library's commitment to the digitisation of selected items of its collections. More than 800,000 digitised documents (printed and iconographic materials) in French and other languages are freely available online. From the origins of printing to the early 20th century, they cover all domains of knowledge, with a specific focus in literature and history. The current Newspaper Digitisation Programme is making more than 3 million pages available both in image

Copyright Alain Goustard/BnF

Bibliothèque nationale de France

and text format through access tools such as full-text search, quality image viewing and article printing.

In the first phase, only public domain documents were planned to be made available. However, within the context of a study led by French Publishers Association and the BnF, the possibility of making more recent documents available online is being studied in respect to authors' rights. This experiment, launched in 2008 and lasting for one year, allows Gallica to offer copyrighted digitised contents online alongside its Culture Heritage collections. This new offer is the fruit of cooperation between the BnF, the French Publishers Association, the National Book Centre, e-retailers and some publishers.

Thanks to this experiment, the BnF is strongly committed to the ARROW project which is building the first European network for the rights status of works. The BnF is the ARROW National Contact Point for France and the coordinator of the work package entitled "Interoperability" which will provide guidelines in order to facilitate interaction between public and private e-content collections. Besides the BnF, the e-retailer Numilog is also a project partner. In addition, the French Publishers Association and the Centre Français d'exploitation du droit de Copie, the French organisation handling the right of reprographic reproduction for books and periodicals, are ARROW supporters.

At a European level, regarding access issues, the BnF is a founding member of The European Library consortium and is involved in the TELplus and IMPACT projects, where it is looking at high quality OCR, full-text indexing and multi-lingual subject issues and in BHL-Europe*, which aims to bring more content about biodiversity into Europeana. The BnF is also a member of the Europeana v1.0 thematic network.

BRITISH LIBRARY

The British Library is one of the world's greatest research libraries, a unique storehouse of over 150 million items. It plays a vital role in the life of the UK as a cultural heritage resource by managing, preserving, and ensuring access in perpetuity to the national published archive and the national repository of sound.

Like all national libraries, there is a growing expectation on the British Library to use the web to provide access to the collection to everyone. Digitisation of its legacy content on a mass scale would act as a springboard for research and education, for new forms of creativity, and for knowledge creation. This indeed is the thinking behind the establishment of the European Digital Library as well as the i2010 digital libraries programme, of which the ARROW project forms one strand.

To date, the British Library has embarked upon a

British Library aerial view

business models and the validation of the system. In addition to these core tasks, the German National Library is working on a use case to consider the various legal issues that might occur during the rights clearance process. On the basis of a selection of copyright-protected books (approximately 5,000 titles from the *Archive of the Best-Designed German Books*), a title-specific search for rights holders is being carried out, with the aim of entering into licence negotiations with as many rights holders as possible. The chosen collection represents the entire breadth of German book publishing activities during the past 75 years. The wide range and the diversity of the collection make it especially valuable as a use case.

The results will make clearer what percentage of twentieth-century book production can be presumed to be indeed orphaned. It will also show in what percentage of cases a licence agreement with the rights holder can be obtained for out-of-print books. The experience from this process should provide a firm basis for establishing future strategy in dealing with out-of-print and orphan works. The use case will also help to facilitate the integration of commercial content providers by testing different licence models. It will identify a manageable workflow for clearing copyrights that can be integrated within

number of projects to deliver 25 million pages of 19th century books, 4 million pages of 19th century newspapers and 8,000 hours of archival sound recordings. This represents only a fraction of its total collections, and the text-based projects primarily represent out-of-copyright material.

In order to avoid creating an online "black hole of the twentieth century" it is important to establish mechanisms that facilitate the mass digitisation of library holdings. Mechanisms include the establishment of historical databases of rights holder information, which is the role of ARROW. The Library sector believes that this must also be supplemented by legislative change.

The British Library is delighted to be a partner in the ARROW project. It is acting as the lead contact point for the UK, co-ordinating input from the UK partners and encouraging input from as wide a range of stakeholders as possible. The Library believes the establishment of a rights database of European copyright works that clarifies the status of orphan works will be one important step in facilitating the digitisation of European culture.

GERMAN NATIONAL LIBRARY

The German National Library's main role in ARROW is to contribute to a survey of the legal framework and

German National Library in Frankfurt

the ARROW system. The licensed titles will then be digitised by the German National Library and provided as digital content to The European Library and Europeana, as part of the ARROW initiative.

NATIONAL LIBRARY OF SPAIN

The National Library of Spain (BNE) is interested in contributing to the creation of a reference database on the status of individual works in terms of copyright and the identification of orphan works.

The BNE is working as an interoperability expert, providing information designed to enhance business interoperability between public and commercial e-content. The current metadata is in MARC21 and can be converted to Dublin Core or XML. The data is in all languages but mainly in Spanish. The BNE is collaborating with other partners in ARROW to establish the legal framework, to find workable solutions and to define business models in relation to orphan and

out-of-print works. Throughout the project's life, the BNE will be involved in disseminating information about ARROW and its progress, by means of articles, conferences, seminars and other forms of communication.

The budget allocated by the Spanish National Library to the project is €156,000, of which 50% is being funded by the European Commission. The BNE is contributing with 28 man months during the period of the project.

The BNE is also playing the role of coordinator of the Spanish ARROW stakeholder group.

NATIONAL AND UNIVERSITY LIBRARY OF SLOVENIA

The Digital Library of Slovenia[10] at the National and University Library has been a national aggregator of e-content since 2005. Approximately 300,000 textual digital resources (articles, journals, printed books, manuscripts) are now available on the portal. Copyright issues are therefore of the outmost importance for the daily work of the library, as well as for long-term digitisation policy and planning. With the ARROW project, the library will benefit from an international network offering a copyright search and clearing facility. The service will allow the library to sort out the copyright statuses of works without the usual limitations of national content only, and to contribute the relevant copyright information to third parties. During the past couple of years, the library has digitised some of its national and foreign materials for conservation purposes and received quite a significant number of orders for the digitisation of books through the eBooks on Demand (EOD)[11] service, for which the copyright status was unclear. The ARROW registry and its distributed chain of copyright clearing centres will certainly optimise the library's efforts to clarify the statuses of these materials and enable it to offer them to the general public as well. In helping to establish the ARROW registry, the library has managed to broaden the circle of contributing stakeholders at national level. These include the National Bibliography Service and the National and University Library databases, the Digital Library of Slovenia, the National ISBN Agency, the National Publishers' and Booksellers' Association and the National Reproduction Rights Organisation. Information on the role of ARROW within the digitisation and copyright management strategy was included in the

National Library of Spain in Madrid

National Library's white paper on digitisation and associated legal issues.[12] The library intends to enhance its communication with relevant parties in the coming months by attending national and international conferences, where ARROW will be presented to the rest of the interested public.

Conclusions. ARROW's crucial role is to function as a bridge between the available information sources in support of the workflow for determining and clearing the rights on works. It will include a complete set of guidelines of applicable standards for both public and commercial collections. It will also include guidelines for interoperability between public and commercial collections. Access to electronic content will be under the different conditions set by rights holders or libraries, and may be free-of-charge or subject to different payment models. ARROW will propose business models which provide a framework for copyright material, with a range of options for the provision of that content by publishers, such as subscription, pay-per-view etc. There are a number of tangible services that will be created from the project. It will provide a distributed network of information sources on copyright status. It will provide accessible data to assist libraries with the identification of rights holders. It will provide tools to support interoperability between public domain and commercially-available works. It will create a registry of orphan works. It will set up a network of rights clearance centres. By the end of the project in 2011, ARROW should have taken a major leap forward in creating a rights information infrastructure for Europe.

National and University Library of Slovenia in Ljubljana

Photography by Branko Cvetković

References

[1] The ARROW project, <http://www.arrow-net.eu/>.
[2] Europeana, <http://www.europeana.eu/portal/>.
[3] For a list of funded projects in the eContent*plus* programme, see: <http://ec.europa.eu/information_society/activities/econtentplus/projects/funded_projects/index_en.htm>.
[4] i2010 Digital Libraries Initiative High-level Expert Group, <http://ec.europa.eu/information_society/activities/digital_libraries/cultural/actions_on/consultations/hleg/index_en.htm>.
[5] i2010 Digital Libraries Initiative High-level Expert Group, Copyright Subgroup, *Final report on Digital Preservation, Orphan Works and Out-of-Print Works*, June 2008, <http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf>.
[6] The European Library, <http://www.theeuropeanlibrary.org>.
[7] i2010: Digital Libraries Initiative, <http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm>.
[8] The National Libraries of France, Germany, the Netherlands, Norway, Slovenia, Spain and the United Kingdom are contracting partners in the ARROW project. The National Library of Finland is an associated partner in the project.
[9] Gallica, <http://gallica.bnf.fr/>.
[10] The Digital Library of Slovenia, <http://www.dlib.si/dlib_eng.asp>.
[11] eBooks on Demand (EOD), <http://books2ebooks.eu/>.
[12] M. BOGATAJ-JANŽIČ, *Strokovna gradiva in podlage za pripravo predloga sprememb in dopolnitev ZASP za optimiziranje delovanja Digitalne knjičnice dLib.si*, Ljubljana, Inštitut za intelektualno lastnino, 2009.