



Supporting rights clearance for digitisation projects with the ARROW service

Nuno Freire

The European Library, National Library of the Netherlands,
nfreire@gmail.com

Gabriella Scipione

CINECA, InterUniversity Computing Centre Information and Knowledge Management Services,
g.scipione@cineca.it

Markus Muhr

The European Library, National Library of the Netherlands,
markus.muhr@kb.nl

Andreas Juffinger

The European Library, National Library of the Netherlands,
andreas.juffinger@kb.nl

Abstract

The process of clearing the rights situation of a work for digitisation is very difficult to be performed for works that have been published during the 20th and 21st centuries. Although libraries hold materials of public interest, which should be made digitally available to a broader public, legal issues make it necessary to determine their exact copyright status before a library

can digitise it. One of the major challenges in rights clearance is the significant fragmentation of rights information across multiple data sources, some of which are not remotely accessible. This makes the rights clearance process very demanding and expensive for libraries. Large-scale digitisation projects can digitise thousands of books per week, and therefore there is a significant need to develop faster ways to clear the copyright status of the books. In the ARROW and ARROW Plus projects (Accessible Registries of Rights Information and Orphan Works), a single framework is being established to combine and access rights information. It proposes to create a seamless service across a distributed network of national databases containing information that will assist in determining the rights status of works. Its goal is to support mass digitisation projects by finding automated ways to clear the rights of the books to be digitised. This paper describes the ARROW service from the perspective of libraries undertaking digitisation projects. It presents the complete ARROW workflow, how national bibliographies are used in ARROW, and the services that ARROW offers particularly for libraries.

Key Words: rights information infrastructures; copyright; digitisation

Introduction

Libraries, publishers and collective rights organisations are discussing possible ways of maximising access to digital content in Europe without harming the rights of authors and copyright owners. It is commonly believed that libraries hold materials of public interest, which should be made digitally available to a broader public. Legal issues, however, make it necessary to determine the exact copyright status of those works before a library can digitise a manifestation of that work. One of the major challenges in rights clearance is the significant fragmentation of rights information across multiple infrastructures, some of which are not remotely accessible. This makes the rights clearance process very demanding and expensive for libraries. Large-scale digitisation projects can digitise thousands of books per week, and therefore there is a significant need to develop faster ways to clear the copyright status of the books.

The ARROW network¹ (Accessible Registries of Rights Information and Orphan Works) is establishing a single framework to combine and provide access to rights information. It proposes to create a seamless service across

a distributed network of national databases containing information that will assist in determining the copyright status of works. Once established, this infrastructure will provide valuable tools for libraries and other organisations to identify and contact rights holders, in seeking rights clearance for the use of content.

The main motivation behind ARROW is to support mass digitisation projects by finding automated ways to clear the rights of the books to be digitised. This rights clearance process is time-consuming, since a library has to go through the following steps for each book:

- Identify the underlying work incorporated in the book to be digitised;
- Find out if the underlying work is in the public domain or in copyright, and whether it is an orphan work or out-of-print;
- Clearly describe the use that is requested for the book, such as digitisation for preservation, electronic document delivery, etc.;
- Identify the rights holder(s) or their agent(s), such as a collecting society;
- Seek the appropriate permission, if necessary.

This process depends on the availability of existing bibliographic and rights data. There are already-established information sources for printed material, in national bibliographies, books in print and the databases of rights organisations. Without ARROW, these sources are not interoperable because of differences in data collection policies and data schemas. Bibliographic databases rarely include metadata about rights ownership and usage policies; instead, this information is usually held in a wide array of formats by publishers, collecting societies and authors.

ARROW addresses the interoperability of rights information along this process. It supports the identification of a work, the clarification of its rights status and the identification of the rights holders. This paper will present the complete ARROW workflow and the main ARROW system, and discuss the particular role of The European Library regarding the use of national bibliographies in ARROW. The paper will also present some results of an internal ARROW validation, and end with the on-going and future work on the national ARROW implementations.

Approaches to rights clearance

The process of clearing the rights situation of a work that has been published during the 20th and 21st centuries is very difficult to be performed. These processes may require information that is not publicly available, or which is not even recorded anywhere (leading to the existence of orphan works, that is works for which the rights holders are unknown or cannot be traced). In addition, the costs of performing such process manually (often called diligent search) may be extremely high, and often it is not even considered a viable solution.

The conjunction of these two difficulties has led to a scarce availability of digitised works from this time period, a problem that is often called the digital 'black hole' of the 20th and 21st centuries in Europeana, and represents a significant barrier to research, innovation, education and culture.

Several approaches to this problem are being discussed, and some are already being applied. Due to the high costs of performing diligent searches, some countries embraced, or are considering the application of Extended Collective Licenses. In this type of legal model, the binding effect of a collective agreement between an organization of copyright holders and a user of copyrightable works is extended to right holders who are not members of the organization (Riis & Schovsbo, 2010). Extended Collective Licenses have been used for rights clearance in the European Nordic countries since the early 1960s, and have been the object of much interest around the world as a means for rights clearance (Riis & Schovsbo, 2010).

Alternative approaches are based on voluntary or contractual agreements between stake holders (van Gompel, 2007; Attanasio, 2010). These kind of approaches must be based on information existing in databases, and rely on registries that enable the processing of transactions, as well as the establishment of agreements between stakeholders based on existing rights information (Varian, 2006).

The rights infrastructure provided by ARROW allows the interoperability of rights information, in a way that is not exclusive for a particular kind of legal model. By making rights information available and interoperable, it enables the information to be used according to the requirements of many possible contractual agreements or licenses currently under discussion throughout Europe.

Other relevant work for libraries undertaking digitization projects is the Public Domain Algorithm² developed in the EuropeanaConnect project³ (Angelopoulos & Jasserand, 2010). The algorithm is intended to assist users in the determination of whether or not a certain work with copyright has fallen into the public domain and, therefore, content consumers, such as libraries, have the right to make use of such works without obtaining permission, and copyright restrictions. This project has created, so far, thirty Public Domain Calculators, each one covering the copyright and neighbouring rights term of protection regime in a separate European jurisdiction (Angelopoulos & Jasserand, 2010).

The ARROW rights clearance workflow

Figure 1 shows an overview of the general workflow of ARROW. It starts from a library as a potential user that wishes to digitise a book, and shows the process that the ARROW system supports to provide a response containing the requested rights information. The process depends on data from several sources:

- National bibliographies' data aggregated in The European Library (TEL);
- Author data from the Virtual Authority File (VIAF);
- Publishing data from Books In Print databases (BIP);
- Rights holders data from Reprographic Rights Organisations (RRO).

The initial steps of the workflow depend on The European Library's system to fulfil the information requirements of the process regarding national bibliography data. Three tasks are carried out:

- Identification of the exact record, from the national bibliography, of the book that the library intends to digitise;
- Identification of other records of books which share the same intellectual work and, therefore, are essential for the rights clearance process;
- Improvement of the data about the contributors of the work.

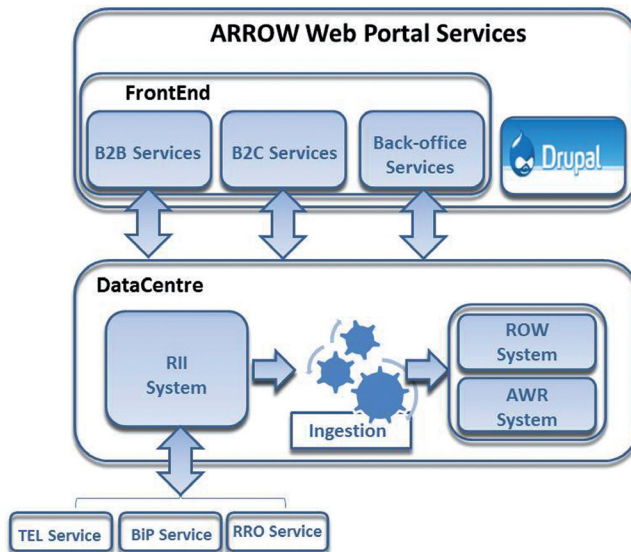
- ARROW Web Portal Services
- The Rights Information Infrastructure (RII)
- The ARROW Work Registry (AWR)
- The Registry of Orphan Works (ROW)

Figure 2 shows a schematic representation of the ARROW System. The *ARROW Web Portal Services* comprises two main components: the FrontEnd, and the Drupal Content Management System (CMS).

The FrontEnd is responsible for collecting input in various forms from the user and processing it to conform to a specification that the DataCentre can use; in other words, it represents an interface between the user and the DataCentre. The interaction with the user can be through the ARROW web portal (B2C services) or directly by querying the ARROW web service (B2B services).

The CMS is a software tool designed to facilitate the management of the website content. ARROW uses Drupal as CMS, a free and open source that is distributed under the GNU General Public License. The Back Office is conceived

Fig. 2: The ARROW system overview.



as a set of services designed to help the management of the entire system, such as the administration of the users and their roles.

The *ARROW DataCentre* constitutes the back end and performs the business logic of the entire system, including both the RII and the ARROW Work Registry workflow. The business logic of the DataCentre is based on the workflow described previously that requires the exchange of information with other external data providers, exposing data via different interfaces and protocols.

The *Rights Information Infrastructure* (RII) is at the backbone of the ARROW System 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 ARROW 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 (The European Library, VIAF, Books in Print, RRO) and elaborating the gathered results, provides information on the work's rights status.

To obtain rights status information, ARROW RII performs three subsequent processes:

- The European Library process in which ARROW exchanges and elaborates the information coming from The European Library.
- The BIP process in which the data coming from The European Library are further elaborated and enriched with the information gathered by the BIP.
- The RRO process that sends to the RRO the library request enriched by all the data at work and manifestation level, collected and processed by the previous data sources, and gathers the RRO response.

The initial library request is performed at manifestation level, 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 (The European Library process), work and manifestation clustering and the

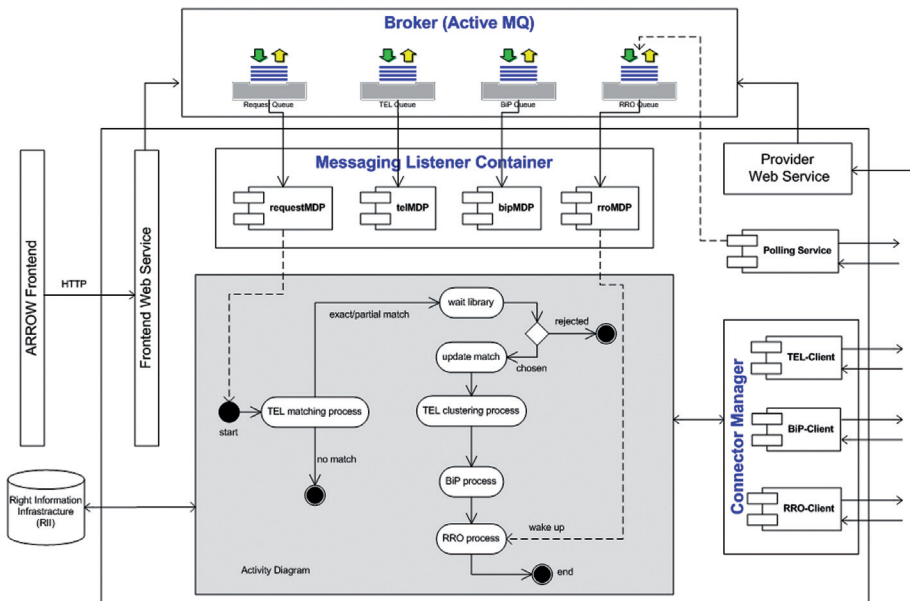
identification of related works and manifestations (The European Library process).

A suite of messages called “ONIX for Rights Information Services” (ONIX-RS) has been designed ad hoc by the ARROW team, in collaboration with The European Library and specialist support of EDItEUR⁴. Its purpose is to support the automatic exchange of metadata between the different sources involved in the project. ONIX-RS relies heavily on the original work of ARROW, but has been extended to accommodate other flows of information in the field of rights, so that it can be used by other organisations or associations working in this field.

Figure 3 shows a simplified architecture diagram of the DataCentre, containing some of the components described previously, as well as the current workflow represented by the activity diagram highlighted in grey.

As the figure shows, all external communications with data providers pass through the Connector Manager component. In case of synchronous

Fig. 3: The ARROW system architecture overview.



communication, the provider client components send the requests according to the provider's protocol and implementation specification, and the retrieved responses are immediately returned to the workflow engine.

In case of asynchronous communication, the provider client components send the requests to the relevant service, and the responses will later be retrieved in two ways:

1. ARROW service performs a polling on the provider's service,
2. External providers invoke the ARROW Provider web service.

In both cases, the obtained responses are moved to the messaging broker.

This asynchronous mechanism has been implemented using an external service (ActiveMQ) which fully supports transient, persistent and transactional JMS messaging. The DataCentre uses JMS listeners to fetch responses from the proper queue and delegates the responses to the workflow engine.

The workflow engine component is implemented using jBPM framework. jBPM manages the process instances described by a process description document. This framework enables us to deal with workflow declaratively and in a more flexible manner.

At the end of the ARROW workflow, the following pieces of information have been retrieved in the message exchange and stored in the RII repository:

- 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 (The European Library, VIAF, BIPs, RROs)
- A set of ARROW Assertions on each work: Copyright Status, Publishing Status and Orphan Status.

The results and the information collected during the RII workflow form the basis for the *ARROW Work Registry* (AWR) and therefore for the *Registry of Orphan Works* (ROW), which is a subset of the AWR. The works stored in the AWR that are at any point in time marked as “*probably orphan*” are added to the ROW. The information about works contained in the *Registry of Orphan Works* is made publicly searchable through a web interface, thus allowing the rightsholders (individually or through a collective representative organisation or agent) to claim their rights.

The design of the AWR and ROW also considers the:

- ISTC metadata for works, in order to guarantee interoperability with the services provided by the ISTC international agency for ISTC registrations;
- Guidelines provided in the High Level Expert Group (HLEG) Final Report on Digital Preservation, Orphan Works and Out of Print Works⁵, which includes recommendations and key principles for rights clearance centres and databases for orphan works.

National bibliographies at The European Library

National bibliographies are one of the key data sources in the rights clearance process. This section presents the challenges related to using national bibliographies to support the ARROW workflow, and how they were addressed in the system of The European Library.

National bibliographies aim to list every publication in a country, under the auspices of a national library or other government agency. Depending on the country, all publishers in that country will need to send a copy of every published work to the national legal deposit or, as is the case in certain countries, a national organisation will need to collect all publications. Given that the publishing domain is very heterogeneous and thousands of publishers might exist in a country, national bibliographies are effectively the single point of reference with which to identify all the publications of an intellectual work by country. However, national bibliography catalogues are created for library management, preservation and, of course, for library users. Due to this fact, they are not immediately usable for the purpose of rights clearance. The necessary information resides encoded, structured and unstructured in

the national bibliographies. For rights clearance, it is necessary to process the bibliographic data to make relevant information explicit, cleaned and normalised before it can be used.

National bibliographies are typically created and maintained by national libraries. Whenever a book is published in a country, it is recorded in the corresponding national library catalogue from where the national bibliography is derived. However, each new publication of a work is recorded independently of all others, and has its own record.

National bibliographies are created mainly with the users of bibliographic data and the management of library holdings in mind. So, although the required information exists within national bibliographies, it is not available in a structured form. Besides the lack of the necessary data structure, national bibliographies also face data quality problems. In spite of many efforts that libraries undertake regarding the standardisation of cataloguing practices and bibliographic data formats, a profusion of heterogeneous data is still found in library catalogues.

Cataloguing rules therefore still leave room for different interpretations, and the information that libraries record in catalogues is often too complex to be encoded in a structured form for machine processing. Librarians often have to resort to general notes fields to record particular information, or they have to work with the limitations of the information systems that are not always up-to-date with the standards, or that do not fully enforce and validate cataloguing practices. As a result, the same information may be represented quite differently from library to library, and even within one library. In addition, these data sources are also subjected to general data quality problems, such as typing errors, misspellings, synonyms, homonyms, abbreviations, etc.

Since each publication of a work is recorded by libraries separately, the work may be represented in different ways. Starting from a particular publication of a work, the task of retrieving all other publications cannot be achieved easily by searching for equal values on titles and contributors (Freire, Borbinha, & Calado, 2007). Comparison and matching of work data has to be carried out by taking into consideration the heterogeneity of data. Without it, the task of rights clearance would be performed on incomplete information, which could lead to incorrect results with legal consequences for libraries, or which could be unfair for rightsholders.

Although national bibliographies are promising data sources for identifying all the publications of a work, this process must be done with some uncertainty due to the characteristics of the data. For this reason, a specialised system has been developed at The European Library to fulfil the information needs of ARROW regarding national bibliographies.

The main use case presented by the ARROW workflow is the *Work Clustering of Manifestations*. It has, as input, a record from the national bibliography, chosen in the previous step of the workflow. The system should identify all other manifestations that potentially share the work, in part or totally.

Even though the clustering is to be done only with records of the national bibliography, elements which must be considered are data heterogeneity and missing data. Based on the results of a previous study (Freire & Juffinger, 2011), the manifestations should be clustered hierarchically in one *Primary Cluster* and several *Secondary Clusters*. The clusters should be formed as follows:

- Primary Cluster: This cluster should be formed by all manifestations that contain the same or very similar title, the same contributors and the same language(s) of text.
- Secondary Clusters: These clusters should provide an organised view of similar manifestations, which may actually share the same work with the input manifestation, but were not included in the Primary Cluster due to data heterogeneity. These clusters should identify manifestations with the following similarities:
 - Same first contributor — because additional contributors are very often not catalogued in a structured way.
 - Similar titles by the same contributors — because titles are sometimes catalogued in different ways, particularly subtitles.
 - Manifestations catalogued without language of text — if the language is missing, it may be the same language as the input manifestation.
 - Manifestations catalogued without country of publication — the missing value may indicate that the manifestation was not published in the target country, but it may also be missing by error.

This organisation of the manifestations in different clusters aims to allow the ARROW workflow to proceed automatically in some cases, or to reduce the manual effort in the verification of the results.

More comprehensive data about the contributors may be of value for other tasks of the ARROW workflow. For this reason, data about the contributors in each cluster should be complemented with data available in VIAF, such as variant forms of the names, dates of birth and death, and nationality.

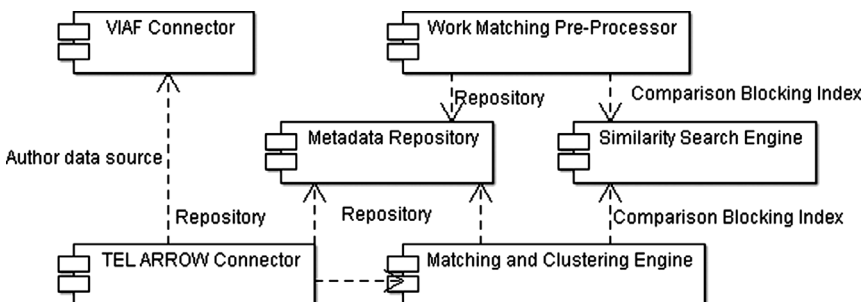
Each cluster is described by work-level metadata so that it can be used in the workflow for interoperability with publishers' bibliographic data and ISTC.

The system of The European Library comprises six main software components, as shown in Figure 4. Together, these components provide the system with a repository for metadata, functionality for information extraction, similarity searching and clustering, and interoperability with external systems.

The *Work Matching Pre-processor* component is the main processor of bibliographic data from the national bibliographies. It addresses heterogeneity and quality issues in the data, so that the rest of the system may function on consistent data. It applies data cleaning, information extraction and data transformation techniques, in order to extract and normalise work metadata from bibliographic records.

The *Similarity Search Engine* allows the processing of similarity queries on work metadata. Its main role is to improve the processing performance of the *Matching and Clustering Engine*, by restricting the processing to a subset of the national bibliography, containing only records with a minimal level of similarity with the input record. Its implementation is based on indexing of character n-grams⁶ of titles and author names.

Fig. 4: Components diagram of the system used at The European Library.



The *Metadata Repository* component provides storage for the bibliographic records in the catalogues of national libraries and their corresponding work metadata.

The *Matching and Clustering Engine* component is responsible for the main processing of the ARROW requests. It applies duplicate detection techniques that do not require data to be encoded in the same way, and does not need the existence of identifiers to detect duplicates. The work clustering of manifestations are a form of duplicate detection, applied with defined criteria based on work metadata. These criteria defined how primary and secondary clusters should be formed.

The *VIAF Connector* component provides the retrieval of authority records about contributors from VIAF. These are used by the *TEL-ARROW Connector* for enrichment of the work metadata provided to ARROW.

Communication and exchange of data with ARROW is carried out by the *TEL-ARROW Connector*. This component implements the web services interface of ARROW and the ONIX-RS messages.

Validation of ARROW for rights clearance by libraries

One of the last tasks of the ARROW project was to conduct an internal validation of the system. It addressed three different aspects of the ARROW rights infrastructure:

- Validation by the stakeholders — Structured interviews were conducted with project partners and stakeholders (including libraries, reproduction rights organisations, collective management societies, and books in print organisations).
- Feedback from the early adopters — Both partner libraries and external libraries ran pilot projects or experiments with the ARROW system.
- System performance evaluation — Measurements were made on several aspects of the performance of the results obtained through the ARROW system.

The validation by the stakeholders involved different kinds of organizations from France, Germany, Spain and the United Kingdom. In total, 12

organizations were involved, including 4 libraries, 4 reproduction rights organizations, 1 collective management society, and 3 books-in-print organizations.

This validation addressed issues such as expectations towards ARROW, the current implementation, market expectations, and the business model of ARROW. The main conclusions of this evaluation related with the process of establishing an effective cooperation between all required partners in each country. National meetings need to be organized to ensure good communication channels, sharing of experiences, and establishing a common understanding of all partners' roles. The validation has also highlighted that since ARROW is a unique and innovative system, and its functionality needs to be frequently demonstrated. It also emphasized the need for good quality databases to be available in order to make ARROW a viable solution.

The first adopters of ARROW included libraries that were project partners, and also external libraries. Thirty libraries participated in the evaluation of the ARROW system performance, and its functionality and usability.

The participating libraries were introduced to the ARROW system, and were allowed access to the system, where they could execute the ARROW workflow for a sample of records. Their feedback was collected by means of a questionnaire, concerning several steps of the ARROW workflow such as the upload process, the monitoring area, the matching process, The European Library work clusters, the BiP response, the RRO response, the registry of works, and also the general user experience.

The participants were in general quite satisfied with the ARROW system and the registry of works, as it allowed for the first time an automated search for the right status of books. The main recommendations from this validation concerned practical aspects of the system. It highlighted issues such as support for additional descriptive metadata formats, handling of large quantities of records, particular user interface functionalities, and better explanation of the main concepts and terms used within the system.

Another relevant outcome of the validation process was the comparison between performing the rights clearance process manually, and by using ARROW. For this validation, four national libraries have recorded the total time spent in performing the rights clearance process manually and, at a later

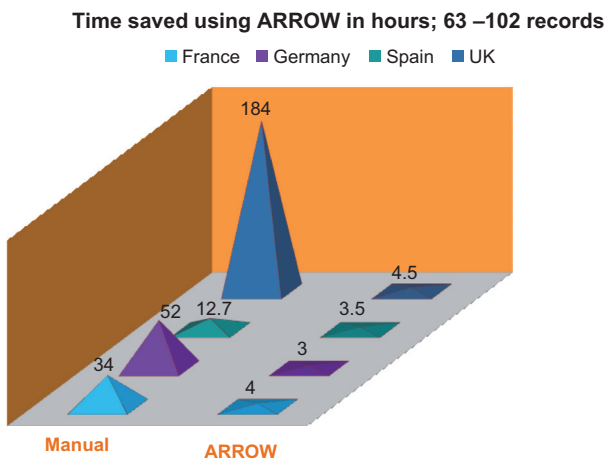
stage, the same process was performed using ARROW. In total, the process was done for 102 books, and the total time spent was significantly lower using ARROW. The number of hours by country can be seen in Figure 5. Performing the process through ARROW required only 5% of the time required by the manual process.

Future work on the national ARROW implementations

The ARROW workflow is currently deployed in four countries: France, Germany, Spain and the United Kingdom. These were the pilot countries of the earlier ARROW project. Currently, the establishment of the ARROW workflow in other countries is being addressed in the on-going ARROW Plus project⁷.

ARROW Plus is a Best Practice Network project selected under the European Commission’s Competitiveness and Innovation Framework Programme, running from the 1st of April 2011 until the 30th of September 2013. ARROW Plus aims at refining the ARROW workflow by increasing the number of countries in which ARROW is used, and broadening the types of works for which it is used to include visual materials.

Fig. 5: Total time spent by libraries to clear the rights of 102 books by a manual process, and by using ARROW.



The establishment of ARROW in other countries is currently being addressed in ARROW Plus for the following 12 countries: Austria, Belgium, Bulgaria, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Poland, Portugal and the Netherlands. These countries are being addressed in different levels of the requirements, since each country presents a unique setting with different levels of compliance of the information resources needed by ARROW. The target countries can be roughly divided into two groups, depending on their level of compliance. The first group includes Austria, Belgium, Greece, Ireland, Italy and the Netherlands; in these countries the book market is covered by a books-in-print catalogue and there are collective management organisations that represent, at least to some extent, the relevant rights holders. Bulgaria, Hungary, Latvia, Lithuania, Poland and Portugal are less compliant with the requirements of ARROW, especially because they do not have books-in-print databases; in some cases, the collective management domain is also less developed or simply non-existent.

For countries where the necessary data infrastructure is already in place, ARROW Plus is cooperating with the relevant national stakeholders to allow the integration of the required data sources. For the second group, the project partners will provide expertise and advice on the characteristics and functions of books-in-print and reproduction rights organisations and on their business models; they will also develop a software system that national stakeholders will be able to use as a basis for their own national services.

Conclusions

The ARROW service is the first system of its kind, providing an interoperability infrastructure that brings together information from national libraries, reproduction rights organisations, collective management societies, and books in print organisations. By making all these data sources interoperable, it allows libraries conducting digitisation projects to perform automatised diligent searches, and also provides informational support for the establishment of nationwide legal models based on contractual agreements or licenses between stakeholders.

The evaluation of the ARROW approach to rights clearance has revealed that, although several improvements are still desirable, the system can fulfil the requirements of all stakeholders and libraries. In particular for libraries

undertaking digitisation projects, the costs of diligent search can be dramatically reduced.

Current work is addressing the establishment of ARROW in more countries, and in the constitution of the ARROW legal entity. The business model for ARROW is also under preparation, since the system will need a stable and sustainable flow of revenues to cover its costs and a suitable governance model (ARROW, 2011).

Acknowledgements

This work was carried out in the ARROW and ARROW Plus projects. The ARROW project was funded under the eContentplus programme of the European Commission, grant agreement ECP-2007-DILI-527003. The ARROW Plus project is a best practice network funded under the European Commission's Competitiveness and Innovation Framework Programme, grant agreement number 270942.

We would also like to acknowledge the work of the participants in the ARROW and ARROW Plus projects, who contributed to all the work described in this paper. In particular, we would like to acknowledge the contributions of Associazione Italiana Editori, the University of Innsbruck, and the Bibliothèque nationale de France, for their support in the design and validation of the ARROW and The European Library systems.

References

- Angelopoulos, C., & Jasserand, C. (2010). *Europeana Licensing Framework and Background Justification*. Europeana Connect work package four deliverable D.4.1.2. Retrieved October 29, 2012 from <http://outofcopyright.eu/research/Term%20Report.pdf>.
- ARROW (2011). *ARROW Business Model*. Retrieved October 29, 2012 from http://www.arrow-net.eu/sites/default/files/ARROW_Business_Model.pdf.
- Attanasio, P. (2010). Rights information infrastructures and voluntary stakeholders agreements in digital library programmes. *Italian Journal of Library and information science*, 1, 237–261. doi: 10.4401/jlis.it-4539.

Caroli, C., Scipione, G., Rapi, E., & Trotta, G. (2012) ARROW: Accessible Registries of Rights Information and Orphan Works Towards Europeana. *D-Lib Magazine*, 18 (1/2). Retrieved October 29, 2012 from <http://www.dlib.org/dlib/january12/caroli/01caroli.html>.

Freire, N., Borbinha, J., & Calado, P. (2007). Identification of FRBR Works within Bibliographic Databases: an Experiment with UNIMARC and Duplicate Detection Techniques. In: *Proceedings of the 10th International Conference on Asian Digital Libraries* (pp. 267–276). doi:10.1007/978-3-540-77094-7_36. Retrieved October 29, 2012 from <http://www.nsl.hcmuns.edu.vn:8989/greenstone/collect/icadl/index/assoc/HASH255a.dir/48220267.pdf>.

Freire, N., & Juffinger, A. (2011). Using national bibliographies for rights clearance. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2011* (pp. 167–174). doi: 10.1145/1998076.1998109.

Riis, T., & Schovsbo, J. (2010). Extended Collective Licenses and the Nordic Experience - It's a Hybrid but is It a Volvo or a Lemon? *Columbia Journal of Law and the Arts*, 33, IV. Retrieved October 29, 2012 from <http://ssrn.com/abstract=1535230>.

van Gompel, S. (2007). Unlocking The Potential Of Pre-Existing Content: How To Address The Issue of Orphan Works in Europe? *International Review of Intellectual Property and Competition Law*, 38, 669–702. Retrieved October 30, 2012 from http://www.ivir.nl/publicaties/vangompel/IIC_2007_6_orphan_works.pdf.

Varian, H. R. (2006). Copyright term extension and orphan works. *Industrial and Corporate Change*, 15, 965–980. doi:10.1093/icc/dtl026.

Notes

¹ Accessible Registries of Rights Information and Orphan Works website: <http://www.arrow-net.eu/>.

² The Public Domain Algorithm website: <http://www.outofcopyright.eu/>.

³ The Europeana Connect project website: <http://www.europeanaconnect.eu>.

⁴ EDItEUR is the international group coordinating development of the ONIX suite: <http://www.editeur.org/>.

⁵ The Final Report on Digital Preservation, Orphan Works and Out of Print Works is available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

⁶ An n-gram is a subsequence of *n* characters from a given character sequence.

⁷ ARROW Plus - European Registries of Rights Information and Orphan Works project: http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=270942.