

ARROW – steps towards resolving the ‘orphan works problem’

One of the more difficult tasks to be accomplished if we are to digitize the ‘physical legacy’ is that of getting the necessary permissions for digitizing works that are still in copyright. A significant part of the challenge lies in identifying and contacting the rightsholders. The ARROW project, funded under the European Commission’s eContentplus programme, has brought together an array of partners to try to develop mechanisms which will help connect those seeking to digitize books with those who are able to give them the necessary permission. This article provides a high level overview of a project which could provide an essential element of the infrastructure for communicating rights and permissions information.



MARK BIDE

Executive Director, EDItEUR

There are many aspects of the laboured transition from the physical to the digital distribution of content which will doubtless be very perplexing to our successors in 50 years’ time. How could it have taken so long? What could all the fuss have been about?

But to those of us who are caught in the midst of these revolutionary events – the beginnings of which can clearly be traced back to around the time that I arrived in the publishing industry in the early 1970s – the challenges are very real and fill our daily horizon. A major current challenge is that of digitizing the physical print legacy – all that content that fills the shelves of our libraries, increasingly isolated from a world of e-books and instant delivery to the user. (It is books that are the focus of this article – there are obviously parallels in other media.)

Even with new generations of high speed automated scanning equipment, scanning all those books is costly – not to mention the logistical struggles involved in getting the books from shelf to scanner and back again.

However, there is a more intractable problem. How do I get permission from the copyright holder (or holders) to make the copy in the first place?

One of the possible answers to this question is, of course, that no permission is required. The author may have died more than 70 years ago, in which case the work is outside the term of copyright; public domain works have formed the

basis of some substantial digital conversion efforts. In some jurisdictions, libraries may also have a limited ‘digitization exception’ in support of their preservation responsibilities, although this right typically does not extend to anything beyond an extremely limited ‘making available’ right – which is hardly the primary point of digitization.

Google, of course, decided that permission was not necessary for its ‘Google Library Project’, because their use of the books was within their interpretation of the scope of ‘fair use’ – a view of US copyright law that remains untested to this day. Although Google has chosen to negotiate a class action settlement with representatives of authors and publishers (which at the time of writing is still before the US courts), it has definitively not accepted that what it was doing infringed the rights of copyright owners (at least in the US).¹

Nevertheless, in Europe, there is a general recognition of the requirement to clear the rights in in-copyright works with rightsholders in advance of digitization. But that recognition is just the beginning of a process. How do you get permission from a rightsholder? How do you find out who the rightsholder is, and how do you make contact with them to request permission?

Finding a rightsholder (or at least a route to contacting the rightsholder) may not be so hard with a recent book – the publisher at least should be easy to find in most cases. But if a book is out of print, then rights *may* have reverted to the author – and individual authors (and their heirs) are much

harder to find than publishers. And, of course, publishers come and go (and are frequently swallowed up by larger publishers) so even finding the publisher may be testing. This all gets more difficult with older books – bearing in mind that books published over a century ago may still be in copyright if the author lived for a few decades after publication of their work.

This is the problem that ARROW was established to address. Like all good European projects (this one under the eContentplus programme), ARROW is (almost) an acronym – Accessible Registries of Rights Information and Orphan Works towards Europeana (Europeana being the name for the pan European digital library). ARROW² has an impressive and eclectic array of partners from across Europe – publishers, national libraries, authors' organizations, collective rights management organizations – under the leadership of the Italian Publishers Association (Associazione Italiana Editori, AIE)³.

Piero Attanasio of AIE characterizes the primary role of ARROW not as creating a registry of orphan works (that is, works whose rightsholders cannot be traced) but rather as "reconnecting works with their parents"⁴ – making sure that rightsholders so far as possible *are* traced. This is the territory of the collective rights management organizations for reprographic rights – or Reproduction Rights Organizations (RROs). Until recently the worldwide network of RROs has primarily been concerned with licensing photocopying, but, along with everyone else, these organizations are now hurtling into the digital age. Distributing money to individual rightsholders from photocopying licences requires that you can identify them – so, many of the RROs have unparalleled databases of information about authors and publishers, and the works in which they have rights.

ARROW's task is to connect those who want to digitize works (and libraries are seen as the most likely candidates for this role) with those who can give them the appropriate permission – and to automate that process so far as it is possible (bearing in mind that the volume of transactions will potentially be very large).

Essentially, the intention is that a library will be able to make a request to digitize a batch of books. (For the time being, at least, ARROW's focus is entirely on books). This request will be either in the form of a message sent to ARROW, or the same information entered through a web interface.

ARROW will undertake its identification and checking processes and then reply to the library with one of a number of messages:

- We believe that this work is still commercially available from Publisher X, and it is therefore unlikely that you will get permission to digitize it.
- We believe that this work is in the public domain, and that you do not need permission to digitize it.
- We believe that the rightsholder to this work is Publisher X and/or Author Y, and you will need to request permission from them to digitize.
- We believe that the rightsholder to this work is Publisher X and/or Author Y, and your local Reproduction Rights Organization is able to offer you a licence to digitize this work .
- We are unable to trace any rightsholder for this work, and it is therefore candidate to be considered to be an orphan work . (The fact that no rightsholder can be traced does not, of course, mean that there *is* no rightsholder – if the author died less than 70 years ago, someone owns the rights even if that someone cannot be traced).

Making any use of orphan works (such as digitization) without assuming the risks involved in infringing the rights of the copyright holders will still require some changes to legislation, but ARROW at least provides some elements of the infrastructure that will be needed to make the intended purpose of such legislation realistically possible – providing a basis for the type of 'diligent' search for rightsholders which was envisaged by the 'High Level Expert Group on Digital Libraries' established by the European Commission to advise on this issue. The Memorandum of Understanding on the subject of orphan works was signed by a very broad coalition of interests in Europe.⁵

The real challenge lies in attempting to automate as much of this process as possible. While undertaking a diligent search for the rightsholders of one or two books without the assistance of automation may be possible, it becomes an exceedingly daunting task if the requirement is to clear the rights to digitize hundreds of books each week.

'Automation' in this sense means having access to very large amounts of data – data to *identify* the

works which the library wishes to digitize and then data to *identify* the rightsholders. ARROW is working with three distinct communities of data aggregators:

- the national libraries themselves, through The European Library (TEL)⁶ which (among other tasks) is aggregating the catalogues of European national libraries
- the organizations that create comprehensive 'books in print' and 'books out of print' resources across Europe
- the Reproduction Rights Organizations.

The organizations have different sets of information about 'the same works', and it is by steering a path through these complex data sets that ARROW hopes to be able to automate at least some parts of the process of identifying rightsholders. ARROW's primary technical task is to provide a 'switchboard' to direct information to the right place in the right format so that it can be appropriately processed to provide the answers outlined above.

A few minutes of thought by someone with only passing knowledge of or interest in this area will identify a number of the challenges which ARROW faces in making this process work. At the most obvious and straightforward level, books published before 1970 did not have ISBNs, so identification becomes difficult at even the most basic level. Data matching where no unique identifier is available is a much bigger challenge than those we are used to facing today in the automated management of transactions relating to books.

One of the first tasks that ARROW is tackling is to understand and scope those challenges through a series of 'national pilots' which the project is undertaking in the early months of 2010. It is anticipated that the first of these will be implemented in Germany, with a number of other countries (including the UK, Spain and France) completing pilots in the first half of the year. These pilots will enable systems and workflows to be refined.

ARROW is due to complete its work in the middle of 2011. By then, it is targeted to have completed an ambitious programme of work, including the creation of:

- a distributed network of sources for information on copyright status

- a European Registry of Orphan Works (ROW) – which will *inter alia* enable rightsholders to 'claim' works which have been declared to be orphans.

In a few years' time, the type of 'rights information management' infrastructure that ARROW is expected to deliver will be commonplace – for the health of Europe's digital economy it is essential that it should be. In the short term the challenges sometimes look intractable – but ARROW enters 2010 with real enthusiasm and optimism that these can be overcome.

If ARROW works as it should, the problem of rights clearance and orphan works will diminish with time, as rightsholders are discovered or orphans identified. In this way, ARROW has the potential to create a virtuous circle. It will leave a legacy of enabling digitization programmes and better 'rights information' good practice for future developments to build on. But ARROW cannot be the complete solution to 'the orphan works problem'. It will still need primary legislation to empower and give safe harbour to any organization to issue licences on behalf of unknown rightsholders to digitize books (or indeed any other media) that are still within the term of copyright without incurring a potential criminal liability.

References

1. The Google Books Settlement Agreement:
<http://books.google.com/googlebooks/agreement/> (accessed 18 January 2010).
Includes the recital: *WHEREAS, Google has agreed to enter into this Amended Settlement Agreement to conclude finally and definitively all claims brought in the Action and to reduce further expense, inconvenience, and the distraction of burdensome and protracted litigation, notwithstanding that Google does not admit any wrongdoing alleged in the Action;*
2. ARROW:
<http://www.arrow-net.eu/> (accessed 18 January 2010).
3. Associazione Italiana Editori (AIE):
<http://www.aie.it/> (accessed 18 January 2010).
4. See, for example, a presentation to a Meeting of Cinema Expert Group: Subgroup on Film Heritage (June 2009)
<http://ec.europa.eu/avpolicy/docs/reg/cinema/june09/arrow.pdf> (accessed 18 January 2010).

5. Memorandum of Understanding on orphan works: http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/memorandum.pdf (accessed 18 January 2010).
6. The European Library (TEL): <http://search.theeuropeanlibrary.org/portal/en/index.html> (accessed 18 January 2010).
- (www.editeur.org). As a subcontractor to ARROW, EDItEUR is providing consultancy on standards, and is creating the standard XML message infrastructure through which the various stakeholders in ARROW can communicate with one another.

Article © 2010 EDItEUR Limited

Statement of interest

Mark Bide is the Executive Director of EDItEUR, the not-for-profit global industry standards organization for the book and serial communities

■ Mark Bide
Executive Director, EDItEUR
39-41 North Road
London N7 9DP
UK
E-mail: mark@editeur.org

To view the original copy of this article, published in *Serials*, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=23&issue=1&spage=35>

The DOI for this article is 10.1629/2335. Click here to access via DOI:

<http://dx.doi.org/10.1629/2335>

For a link to the full table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=23&issue=1>