Grant Agreement ECP-2007-DILI-527003

ARROW

Report on business models - Edition 2

Deliverable number	*D-3.6*
Dissemination level	*Public*
Delivery date	*30th July 2010*
Status	*Final*
Author(s)	*Federation of European Publishers*

*e***Content***plus*

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

Report on business models - Edition 2

Executive summary

One of the objectives of Work Package 3 is to define the business models enabled by the project and general framework for enhancing innovative business models and digitisation initiatives by external participants in the e-book sector.

Chapter 1 briefly introduces the aims of the Arrow project - to create an interoperable search infrastructure, a distributed network of national databases containing information about the rights status of works and right holders, and a set of tools for both public and private organisations who wish to contact active rightholders to seek copyright clearance for the reuse of content - and explains the report's objective to present information on current and emerging business models and digitisation initiatives in order to assess their potential relationship with Arrow and thus the potential demand for Arrow's services.

Chapter 2 builds on the results of the report's first edition and concentrates on those findings related to subjects that are likely to have an impact on the development of Arrow (as potential users), or on which Arrow is likely in turn to have an influence (as an enabler). It provides an overview of the e-book market and digitisation initiatives to illustrate how copyrighted works are currently aggregated and offered on the market, concentrating on areas in which a search for rightholders is needed and on the subsequent information requirements.

In a picture of increasing interest for books in electronic format and of emerging new actors, one main conclusion is that difficulties in retrieving rights information and insufficient claritiy on rights clearing mechanisms are among the main obstacles to the development of digital library initiatives, Libraries with digitisation projects are therefore assumed to be one of the main user categories of Arrow. A large proportion of business models in the publishing sector have no specific interest in a tool like the Arrow system, since they do not have particular rights information requirements. However, other kinds of private actors, now entering the e-book sector, could have an interest in the service offered by Arrow as well:

Chapter 3 illustrates the value offered by Arrow to players wishing to engage in digitisation initiatives, both public and private. The common trait between the digitisation initiatives by libraries and Arrow's potential private customers is their purposed large scale. Arrow's services are useful when a subject, for reasons of public function – such as for libraries – or business (trying to exploit, one way or another, the "long tail") seek the maximum information completeness about a collection, with a view to digitising large numbers of works.

Arrow is set to facilitate the search for rights information and rightholders, and thus also the clearance of rights for digitisation and making available of works. It will represent a cost saving factor and an enabler for those involved in digitisation initiatives, including by contributing to avoiding duplication of efforts.

Arrow will therefore be able first of all to enhance digitisation programmes by libraries; it will also enable business models based on the digitisation and reuse of orphan and out of print works, by facilitating both their identification and the obtention of licences for their exploitation.

Chapter 4, the main one, presents concrete cases to support the theoretical findings of the previous ones. It further analyses the kinds of players that can have a need for a system like Arrow for their digitisation initiatives and their motives: preservation and making available of cultural heritage for libraries, and profit for commercial entities, either linked directly to sales (of e-books or of related devices) or to the possibility to offer a wider range of results to users performing internet searches.

Libraries are seen for the time being as the main potential users of Arrow, especially considering the political will behind the Europeana project and the online accessibility of copyrighted material (in full respect of copyright laws). An analysis of the concrete digitisation plans by libraries confirms this huge potential, while raising the issues of funding and of the actual inclusion of in-copyright works in such plans. Whereas most projects currently regard mostly public domain works, there are signs that the picture starts changing.

The interest of private players can be to offer content commercially or to provide access in the form of search results; in both cases, the availability of a very large pool of resources is paramount, hence the push for digitisation. This also explains the monopolistic tendencies of such a market.

One main category of these players is therefore represented by search engines. While several of them tried to enter the field, the most advanced so far is Google, which has digitised millions of books and plans to continue, in partnership with a large number of libraries, for its Google Books project. These plans are seriously limited by the possible large copyright infringements they have implied so far, but a turn into legality could entail a large scale need for services such as Arrow.

The other category is commercial players entering the e-book market to take advantage of a dominant position in other sectors somehow related: books retailing, or consumer electronics (with a view to adding the offer of content to that of devices). Amazon and Apple are currently the main actors in this field, though other may enter the market or try to improve their position in it; for all of them, any plan to enlarge their catalogues could involve digitisation plans and the use of Arrow. In addition, also non-commercial private entities engage in digitisation initiatives, with aims close to those of libraries. The potential of Arrow to lower the entry barriers and contrast the monopolistic tendencies is also highlighted.

Public-private partnerships are becoming very important in the field of digitisation; they allow libraries to overcome budget constraints while offering their partners reputation gains or more direct returns (as for the cases mentioned before). Still, PPP can also consist of public support for the digitisation of backlists by publishers. A number of examples are given; when the works digitised belong to a library, the need for a tool like Arrow is evident, especially when commercial entities are involved.

The possible role of RROs is also explored, as they could be tasked not only with the issuing of licences but also with the conduction of diligent searches, verifying the availability of licences and the administration of orphan works and orphan works registries. Arrow could then represent a cost saving factor for RROs.

Chapter 5 draws the conclusion that once copyrighted works start being included in large scale digitisation projects, a need will arise to find information on rights and rightholders; the following step will be to put in place rights clearing mechanisms. These will be voluntary for out-of-print works, and possibly based on legislation in the case of orphan works; Arrow will be able to play a role in this framework.

1   Introduction

Within Arrow, one of the objectives of Work Package 3 is to define the business models enabled by the project and general framework for enhancing innovative business models and digitisation initiatives by external participants (authors, publishers, RROs, libraries, e-retailers).

In particular, as to business models, the managing strategy is described as follows:

- analyse and understand the existing European players' (authors, publishers, e-retailers, RROs, libraries) business models and general initiatives with relation to e-books (defined as any kind of text based e-content), in terms of digitisation and commercialisation;
- consider current and prospective initiatives for the digitisation and making available of in-copyright works by libraries;
- identify the main information requirements of the stakeholders for such business models;
- verify the criteria, framework and capacity of coexistence and interoperability among different models (commercial, public sector, etc.) in the field;
- define a role for Arrow in the current market situation, as neutral as possible with relation to the existing business models.

The Arrow project will create an interoperable search infrastructure, a distributed network of national databases containing information about the rights status of works and rights and right holders, and a set of tools for both public and private organisations who wish to contact active right holders to seek copyright clearance for the reuse of content; it will also collect information on current and emerging business models and digitisation initiatives to assess their potential relationship with Arrow.

The main contribution of the Arrow infrastructure, in its first phase, will be to facilitate the identification of books as orphan, out of print or in print and the identification of the relevant right holders; this in turn will be determinant in defining the role of Arrow with relation to the possible business models. With this distinction in mind, Arrow will: on one hand, enhance the evolution of digital libraries in full respect of copyright through the creation of databases to identify orphan and/or out of print works and the testing of clearance mechanisms for the latter; on the other hand, help identify business models enabling access to copyrighted works. Arrow will in fact facilitate the digitisation (and use in general) of out of print works (including orphan works) first by clarifying their status, and then by making it possible to identify the right holders, if any exist.

Following the identification of the work, it could also make recommendations to the users, for instance as to whom they should first contact. A further possibility to be explored will be to set up clearing mechanisms that use Arrow for authorising of the digitisation and and making available of out of print works for which the rightholders intend to permit such uses, and for orphan works according to the relevant legal framework.

The update of this deliverable has built on the results of its first edition and concentrated on those findings related to subjects that are likely to have an impact on the development of Arrow (as potential users), or on which Arrow is likely in turn to have an influence (as an enabler). A particular focus was still on the existing or emerging initiatives related to management of orphan works and out of print works, which were found lacking in the first survey. The deliverable collected information from partners and other relevant stakeholders (publishers, online bookstores, libraries, RROs, etc.) on initiatives and business models related to the digitisation and making available of copyrighted works and on the agreements in place between different stakeholders (publishers, authors, etc.) in order to provide online access to such works; it however concentrated on areas in which a search for rightholders is needed and on the subsequent information requirements.

The understanding of the existing European initiatives and business models in the field of e-books in fact allowed identifying a series of requirements for information on behalf of the stakeholders; this in turn will contribute to designing the best kind of rights information infrastructure for the Arrow system. It will also allow understanding how Arrow could contribute to solving some of the current problems (that slow down digitisation initiatives) and to making the system more open and competitive, face to the potential for creating dominant positions in the field.

In addition, by identifying the needs by players involved in digitisation initiatives and in the e-book market (for rights information, rights clearance mechanisms, interoperability, standards, etc.), the report helps to define business models enabled or enhanced by the project and set the framework for ensuring Arrow's neutrality with respect to the different business models applied. Against this background, the interoperable search infrastructure that is expected to be the outcome of the Arrow project, based on a network of national databases, could represent a huge step forward in the direction of facilitating further initiatives of all kinds.

In this perspective, the report also makes hypotheses about other possible business models evolving into developing a need for Arrow's services from a position already within the market, or about the chances for newcomers to make a move into it.

Ultimately, we tried to concentrate on how the multifaceted picture previously presented – including in the first report – can translate into a demand for Arrow's services and envisage a collocation for Arrow in the context of digital books circulation, in terms of relationship with the existing actors and the role to play in the mechanisms established by the current and prospective business models. In other words, the aim of the study is to present all kinds of digitisation initiatives and business models that can have an interest in the use of a service such as Arrow.

Arrow shall be as neutral as possible with respect to the existing business models; it will not become another player in the e-book business arena, but it will manage information for the actors that have already put in place business models and for the potential new ones. The report therefore analysed and presented different opportunities without making specific recommendations.

2 Confirmation of relevant findings from the report's first edition

The first edition of the report already illustrated how copyrighted works are currently aggregated and offered on the market, with different possibilities regarding the licensing schemes, the standards used by publishers to provide the works, the storage and access facilities for the digital files, and the mechanisms available to end users in order to retrieve the books; it presented a very varied picture, and also highlighted possible difficulties in the current market, in terms of competitive, technical or other issues.

One year later, while there has been some progress of course in the e-book market and in the field of digitisation in general, most information - on the qualitative side - remains valid, in particular as far as the trends that may have an impact on Arrow are concerned, which are in turn related to the information requirements of the stakeholders involved. The first report drew a picture of an ongoing process of switch to digital taking place in the book sector, an industry that introduced digital technologies in its production chain early compared to other cultural industries but that initially encountered serious difficulties in adapting the product destined to the end user. Nonetheless, the transition to digital has started also on the consumer end, and - while uneven - it is well on its way.

Certainly, the picture of e-book markets in Europe is very fragmented and diversified: fast expanding markets and initiatives go alongside very underdeveloped markets and businesses striving to find a place in the panorama of digital publishing. Rates of growth in the developed markets are very high, but even in the most developed ones the overall size of electronic publishing compared to the printed book market is very small (in Germany, for example, in the first half of 2009 some 65,000 e-books were sold; while the figure was quite small, analysts considered that with new wireless e-readers and a larger selection of titles on the horizon and distribution becoming more efficient, more relevant figures would come in a relatively short time[1]). The lack of adequate portable reading devices has contributed markedly to this slow development.

Given its characteristics, the STM (scientific, technical, medical) sector this is by large the one that has been involved the most in the digital transition. However, more and more publishers start offering their books in electronic format in other fields. A wide range of different business models for providing access to copyrighted digital book content has arisen after some years of experimenting. One main distinction can be drawn between models targeting end users (digital bookstores, which offer individual licenses to a public who pays for the content) and models targeting businesses (group licence models offered to libraries and commercial entities, thus providing the grounds for creating digital libraries).

New technologies have allowed the creation of hybrid models of digital libraries, blurring the border between library and bookstore: the sale of consultation rights, priced upon direct measurement of consumption. A further step in this sense is the possible role of libraries in redirecting their users to the websites of commercial players which allow the purchase of e-books that have been object of a search on the library's portal. Contractual agreements with libraries, though less frequent, are quite important in economic terms. No significant business models have been reported based on the free distribution of e-books, supported by advertisement or other sources of income unrelated to the end user.

The role of traditional actors in the book value chain (authors, publishers, booksellers, libraries - indeed, e-book stores and digital libraries) tends to be maintained in the digital environment, although some processes need adaptation and some new actors and economic circuits are also likely to appear. In particular, a series of new functions has appeared as the transposition of functions of the printed

---

[1] http://www.boersenblatt.net/336537/

book chain[2]: digitisation and conversion service providers, e-wholesalers, e-distributors, aggregators (a combination of the previous two); other functions are purely related to technological aspects, such as providers of reading software or DRM solutions. In the digital book chain, a single player can hold several functions, while new functions can be held by traditional or new players and the frontiers between the players and their traditional roles evolve.[3]

It is particularly interesting to notice how new players are entering the chain, thanks to their major roles in connection to end-users or customers in other fields:
- search engines or search portals, such as Google, which offers new tools for selecting books with "search inside" technologies (Google Book Search - recently renamed Google Books), and will be opening soon an e-book store (Google Edition);
- pure players of the internet, such as Amazon, but also other kinds of online shops, which develop their e-book stores;
- electronic device manufacturers, such as Apple, which move to the content market for their equipment, thus providing their own e-bookstores as well;
- mobile carriers or internet access providers.

The interest of such a number of players in this field clearly reflects the existence of a substantial demand for books in electronic format, which both stimulates and is reinforced by the evolution of reading devices, which are becoming more and more sophisticated and user-friendly, but most of all available[4]. If the lack of suitable devices was one of the obstacles to the development of the e-book market, now the situation is evolving (for example, in October 2009 Amazon launched its Kindle worldwide, with a catalogue of some 200,000 titles in English; it currently offers over 600,000 titles in copyright and 1.8 million public domain works[5]). All in all, while dedicated e-readers were not widely available in Europe during the time span considered (in many countries they were not commercialised yet and in those where they are they have been introduced only recently), their presence on the market is increasing steadily. The expected downwards trend in the price of e-readers (possibly accelerated by the economic crisis) is likely to further open up and widen the market.

In the meantime, an alternative scenario is developing, which sees a multiplicity of reading devices in use. Smartphones in particular are becoming increasingly the device of choice for electronic reading, although for the time being personal computers are still on top of the list (indeed, toward the end of 2009 book-related applications overtook games in the App Store as a percentage of all released apps, placing the iPhone as a potentially significant player in the electronic book market). And the new iPad,

---

[2] For a more detailed analysis, see D. Zwirn, *Comments and suggestions concerning the Enclave Editores-Bne Project*, June 2009 (www.dilve.es/dilve/getArchivoDocumentacion.do?iddocumento=762)

[3] The report by D. Zwirn (see note 2) illustrates different patterns of distribution of digital functions:
- In the USA, in the U.K. and in Australia, publishers, aggregators and e-book stores are mainly different and independent companies, which contract on the basis of competition between the services offered.
- In Germany, publishers and book stores have launched a common initiative for distributing and selling eBooks (Libreka) via a network of bookstores, through their common association (Börsenverein); besides this initiative, some individual propositions of aggregators and e-book stores are developing.
- In France, the organization of the market tended at first to be structured around e-distributors or aggregators owned by or working exclusively for certain publishing groups, like in the physical distribution industry. This trend has recently changed as more and more various publishing groups are selling their works on common platforms.

[4] The market is expanding already at a very impressive rate in the US, where reported e-book sales rose nearly 252% in the first quarter of 2010, driven by an ever improving reading experience: new e-reading devices, screen reading rivalling paper, content selection, lower prices, etc.

[5] For non-US customers, content availability may vary; for example, differences in copyright regimes make some works available for free in the US but not in Europe.

launched in the US in April 2010 and in a number of other countries[6] in May, is highly likely to be used for e-book reading, too, and its sales are foreseen to be very strong[7].

This demand on behalf of readers is also potentially quite large (possibly larger) for the availability of electronic books through libraries. Many university and research libraries already offer collections of works to their patrons, but there is scope for widening both the audience and the range of works covered. There is in fact a special case for the inclusion in digital libraries of works under copyright, both for the purpose of their preservation and of making them accessible to the widest possible audience. There is a particularly strong political will in Europe on this point, as demonstrated by numerous declarations and resolutions by the EU institutions (which also, we will see later, highlight the need to maintain the sustainability of the book value chain).

For the time being, the overall offer of electronic books is relatively limited in terms of total number of titles, especially in comparison with the availability of printed books (tens or hundreds of thousands as opposed to some 6 million in the EU alone); nonetheless, in this case too a steady growth can be observed as more and more publishers decide to enter the e-book market. In addition, many works in the public domain are available to the public in digital format thanks to digitisation initiatives by a large number of libraries across Europe, based on public programmes or public-private partnerships.

The presence of copyrighted works in digital public libraries is at the moment close to non-existing, for various reasons: the main interest in preserving cultural heritage and ancient works; legal uncertainty on orphan works; absence of discussions with rightholders to find agreements; cost of finding information on rightholders (in particular for out of print or orphan works) and/or of clearing rights on books to be digitised; little or no interest by library users to access pay-for services. The very few cases existing outside of the commercial models usually rely on ad hoc individual agreements with rightholders. Access to digitised works is allowed only on the premises of the libraries involved, or on institutional intranets/extranets, and the books involved are for the majority out of print.

Some pilot projects exist, which generally involve libraries and broad joint right holders' representations following negotiations among the parties. The most interesting cases are the hybrid models that include copyrighted books in collections that users can search and browse: for viewing excerpts of the in-copyright e-books and purchasing them, users are redirected to the suppliers' websites. The main examples are the French Digital Library Gallica, and the integration of the Spanish Digital Library with the Enclave project. Users can find works in the public domain that can be read and downloaded for free, and they can also access information and excerpts of works in-copyright, which they can further read, purchase, etc. upon redirection to the website of an aggregator, distributor or other supplier.[8]

In the field of digital libraries, a very relevant case is the European Digital Library, Europeana, which aims at providing access to Europe's cultural and scientific heritage though a cross-domain portal. Europeana is a decentralised model: it does not host files of works but redirects to the relevant institutions after displaying search results. For the time being, it does not provide access to copyrighted works, but the possibility is being examined.[9]

---

[6] Including the UK, France, Germany, Italy and Spain.

[7] Already some 200,000 are sold in the US every week and it's estimated that by the end of 2010 some 8 million will have been sold, of which about 4.6 million in the US.

[8] The two cases have been extensively described already in the First Edition of this Report. Now Gallica hosts some 20,000 works in copyright, while Enclave, become operational since a few months, contributes some 1,250 titles.

[9] It is foreseen, for example, that the copyright protected content of Gallica be integrated into Europeana; currently, some technical details are being arranged to this end.

Whatever kind of works readers seek, the way to meet their demand is digitisation. Digitisation methods depend on the available sources for books and the goals in terms of quality and format. Digitisation costs in brief vary widely and can be quite elevated; hence the necessity of high levels of investment for the retro-conversion of backlists and in general for the mass digitisation of large collections; technological developments are countering this phenomenon, especially in the field of conversion/ adaptations, thanks to the development of digital asset management systems[10].

Against the potentially relevant costs of digitisation, available resources are quite limited, in particular as far as public support for digitisation by publishers is concerned. In some cases, public authorities expect the market to drive such initiatives by private actors. In general, plans for mass digitisation are not common, again due to financial constraints; very often library budgets are definitely insufficient.[11]

Decisions about the undertaking of digitisation plans are taken by two main categories of players: rightholders and libraries.[12] Publishers own backlists and have an interest in commercialising them in electronic format, while libraries have collections they aim to make available to their patrons in the digital environment. While apparently an obvious statement, this allows us to focus on the groups that most likely have an interest in the retrieval of information on the rights status of books and on rightholders, since rights clearance arises as an issue whenever plans are made to digitise books.

As finding the rightholder for a book commercially available should not be particularly difficult, it is clear that issues with rights information mainly arise with regards to out of print and especially orphan works (the issue of the ownership of digital rights, which is becoming quite relevant as digitisation initiatives expand, is outside of the scope of this analysis). Moreover, if a book is commercially available there is no reason why the rightholder would allow a third party to digitise it and make it available online.

Considering again our main business models, in general there is relatively little commercial interest in out of print and definitely less in orphan works. In many cases, rightholders have a reason to let a work run out of print (be it obsolescence, insufficient demand or else) and the potential market for orphan works looks feeble as well. A large proportion of business models in the publishing sector have no specific interest in using a tool like the Arrow system, since they do not have particular right information requirements; their difficulties are more related to the actual negotiation of digital rights, or to the identification of digital rights as a separate entity.

On the other hand, the difficulties in retrieving rights information (and the lack of rights clearing mechanisms, see below) was identified as one of the main obstacles to the development of digital library initiatives. Therefore, of all the cases considered, those most likely to be interested in a solution such as the one proposed by Arrow are libraries (given certain conditions, such as the availability of funds). Their collections contain a good proportion of out of print works, and also a relevant share of orphan works (or presumed so);[13] besides, it is highly unlikely that they be allowed to digitise and make available works still in commerce, as that would seriously prejudice the normal exploitation of works, without an adequate remuneration of rightholders. In turn, this raises the question of what the fee should be to give users unlimited access to copyrighted works by a digital library with a potential audience of millions: the fixing of a fair remuneration in such a situation is in fact unrealistic.

---

[10] As long as closed systems with proprietary features exist - such as Amazon's Kindle - publishers will have to maintain separate supply chains; this is one of the reasons for calling for interoperability in the field of electronic reading.

[11] A possible quite remarkable exception could be the plan of the French Government to devote a large amount of money from a national loan to digitisation of cultural heritage. See below, par. 4.1.

[12] Other commercial players are part of the picture as well, but those were not taken into account in the First Edition of this Report. They have been considered in the present one (see par. 4.2).

[13] The British Library estimates that over 40% of all in-copyright works are orphan. This figure, nevertheless, refers to all kinds of works; the issue is much less relevant in the field of books, where the proportion is deemed to be in the range of 3 to 5%.

Thus, the support of the Arrow system in facilitating the identification of rightholders (made in general more difficult by the non inclusion of digital rights in the large majority of contracts concluded before 1980) to ask for rights clearance can be very useful for the development of digital libraries, making it easier to include copyrighted works in their digital collections (out of print and orphan works are under copyright). In addition, Arrow's potential cost-saving effect (the reduction of the cost of the search for rightholders) is likely to free resources that can be focused on the digitisation process itself.

Another, interlinked factor determining the absence of substantial business models or digitisation plans dealing with orphan works so far is the almost complete lack of legislation on the subject. Also, and in relation to this, almost no licensing models for digitising and making available copyrighted works by libraries or clearing mechanisms for dealing with out of print and orphan works currently exist, especially on any large scale. Initiatives in this field are mainly built upon voluntary agreements set up on ad hoc bases between libraries and rightholders; in most cases, they involve works provided directly by the rightholders (this is the case of the aforementioned Gallica and Enclave).

Some initiatives do actually deal with ways to simplify the digitisation process and handle the issue of orphan and out of print works, like the Bokhylla project in Norway (based on an extended collective licence) and the VOI©E-FOBID agreement in the Netherlands[14]; both are currently subject to restrictions, which in principle could be eased with a simplified system for retrieving rights and rightholder information. However, in some countries some legal solutions are being explored, which should foster digitisation plans and also on the potential demand for services like Arrow's[15]. And besides, of course, there are business models by suppliers of electronic books which provide libraries with access to collections of works under copyright at some fee.

The rights information infrastructure the Arrow project is building would provide a great help in clarifying the status of many works that are potentially orphan and therefore lie in a sort of juridical "limbo" given the absence of established legal mechanisms for dealing with them. Once any kind of legislation is enacted, the Arrow system will be then able to fit in the established framework.

Some countries envisage the possibility of addressing the issue of orphan works through legislative solutions, possibly in combination with collective management of rights, and are reflecting on the out of print side as well. The European Commission is planning a legislative initiative on orphan works; it is expected *inter alia* to provide a definition of orphan works which is assumed not to differ from the one established by the EC i2010 Digital Libraries High Level Expert Group (HLG),[16] outline the kind of diligent search that should be considered necessary to assess the status of a work so as to allow an orphan work to be made accessible across borders, and leave to individual Member States to adopt specific legislation on the details. Arrow will make it easier to implement whatever kind of legislation based on the diligent search of rightholders. With legal certainty, it is foreseeable that other business models be created, and other kinds of private actors could have an interest in the service offered by Arrow; this will be treated in detail in the following chapters.

---

[14] Both cases have been described in detail in the First Edition of this Report. The Norwegian experiment, well under way, allows viewing the books digitized only to Norwegian IP addresses and does not permit downloading or printing of works under copyright. The Dutch initiative has not yet involved books.

[15] For a more detailed analysis of the initiatives aimed at facilitating rights clearance, in particular for orphan and out of print works, and of their legal background, see the Arrow deliverable D3.5. Report on legal framework - Edition 2. Available on www.arrow-net.eu.

[16] *"An orphan work is a work protected by copyright but the current owner is unknown or untraceable by diligent search." Sector-specific Guidelines on Due Diligence Criteria for Orphan Works*, elaborated by sector-specific working groups composed by representatives from cultural institutions and the creative sectors, which reported to the HLG and to the European Commission (http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/guidelines.pdf).

3   Arrow's value for potential users

From our previous analysis, we concluded that among the players of the digital book arena libraries were the ones most likely to have an interest in using the services that Arrow will provide. However, the benefits of Arrow can be relevant also for other kinds of subjects with somewhat different interests.

Arrow will provide a system to identify the right status, rights and the right holders of copyright based works; provide information on rights clearance; and help to establish a European-wide orphan works registry and to build new/interconnect already established out-of-print registries. The Arrow system is set to allow cost and time savings in searches for rights information and so ultimately facilitating any digitisation initiatives that involve copyrighted works. Arrow will be useful whenever a digitisation initiative requires a diligent search for rights and rightholders, especially with a view to clearing such rights (in particular for making available online). This includes mostly plans to digitise library collections, when the works involved could possibly be protected by copyright.

Ideally, digitisation plans for which Arrow would be particularly relevant are those involving mass search on defined criteria that would one way or another entail some kind of search for rightholders, whether to include or exclude (categories of) works, or rightholders, or periods. The Arrow system is in fact designed to manage exactly this kind fo information.

There are actually numerous initiatives to digitise and make available the European cultural heritage. Digitisation programmes have been put forward by public institutions such as libraries and also by commercial players in the market (most notably Google through its Google Books programme, but it is not the only one, and, as we will see more in detail later on, the initiative had and still has serious issues of copyright violation).

The digitisation and making available of our common cultural heritage has also a strong political support. On the European level this is backed consistently by the EU institutions.[17] Many official documents have clearly pointed out the need to increase the availability of works that are in copyright, in particular through Europeana. It has further been stressed as a clear requirement that the digitisation and making available of copyright works must not conflict with the publishers' and authors' interests in commercialising them. Consequently, there is a need to identify whether a work is in or out of copyright, whether a work is offered commercially, who the right holders are and where the rights can be cleared. The Arrow value proposition therefore meets clearly expressed user requirements, as well as those on the political level. In particular, the potential of Europeana as a direct or indirect user of the Arrow services is quite evident.

This is based on the main assumption that the European institutions and the EU Member States will follow up their political statements on digitisation programmes and on the development of Europeana as a portal to provide access to the European cultural heritage, which includes works in copyright, with concrete initiatives. It is also presumed that the institutions and the Member States will encourage and facilitate national digitisation programmes through providing the necessary funding.

Nonetheless, current and planned initiatives for the most part tend to focus on public domain works, rather than on copyrighted works (although, as we will see later, this picture starts changing).[18] There can be many reasons for this, including the prioritising of ancient, fragile books or in general of older collections, which consist of works in the public domain. Still, according to librarians at least part of the

---

[17] For an extensive analysis of the relevant documents by the EU institutions, see below (par. 4.1).

reluctance to undertake digitisation projects involving copyrighted works is related to the difficulties in identifying rightholders and the legal uncertainty that surrounds the use of orphan works.

It is conceivable, then, that the availability of such a tool as Arrow becomes an incentive to increase plans to digitise copyrighted works. Moreover, even within plans for digitisation of only public domain works there could be a need to ascertain the actual status of at least some of the works involved, to make sure they are in the public domain.

The Arrow system can benefit not only customers from the public sector, but also from the non-profit and even commercial ones. Its advantages can apply to all kinds of entities: cost saving, facilitating digitisation of copyrighted material and, very important, contributing to taking works out of a legal "limbo", by finding their rightholders and, coupled with an adequate legal framework, by allowing the exploitation of orphan works.

The common trait between the digitisation initiatives by libraries and Arrow's potential private customers is their purposed large scale. Arrow's services are useful when a subject, for reasons of public function – such as for libraries – or business (trying to exploit, one way or another, the "long tail") seeks the maximum information completeness about a collection, with a view to digitising large numbers of works (aiming at some sort of comprehensiveness). In addition, when the motive is a public task, there is a further interest to turn to a system such as Arrow's, besides the enabling of a due diligent search for rights information: as concerned with matters of sustainability, libraries will also appreciate the cost containment that is to be realised.

As to private players, there can be non-profit entities or foundations that support the goal of making the cultural heritage available in electronic format, and therefore undertake digitisation projects, usually in collaboration with libraries (as ultimately those are the ones that have the books). Such partnership can be limited to financial support or include technical support as well, and in some cases are motivated by the positive publicity that can be gained in sponsoring cultural initiatives. In any case, the goals and mechanisms are basically the same as when projects are managed completely by libraries, especially as far as information needs are concerned.

The relevance for commercial entities most likely does not derive from the economic potential of orphan - and maybe out-of-print - works in terms of sales (although of course there are exceptions, and out-of-print works can become popular again for a number of reasons[19]). As we will see, these players are rather driven by strategic and political motives, mainly connected to the (to some extent) comprehensiveness of their initiatives, and as long as Arrow contributes to making such comprehensiveness easier to achieve it can be an interesting tool for them.[20]

Arrow can facilitate obtaining licences for the commercial exploitation of works under copyright, both orphan and out-of-print. For out-of-print works, it is important to consider how this business sector is characterised by high fixed costs and usually low market volumes, which makes profitability difficult to achieve. This is why it is conceivable that an intermediary can be interested in clearing rights for a number of works that right holders are not interested in exploiting any longer.

---

[18] It is worth pointing out once again that a large number of digitisation initiatives are being undertaken by lots of publishers in many European countries; since they concentrate on their own catalogues, though, they do not have in principle a need for the identification of right holders.

[19] For example, their authors winning a literary prize, or reaching an unprecedented success with a new work.

[20] In her report *"Pour un livre numérique créateur de valeurs"* ("For an electronic book creator of values", April 2010), Christine Albanel, former French Minister of Culture, underlined the importance of comprehensiveness, in particular in the public service perspective. In stressing the need to make available works that are no longer so, such as orphan and out of print works, she mentioned the high attractiveness of an exhaustive offer. Available at: http://lesrapports.ladocumentationfrancaise.fr/BRP/104000189/0000.pdf.

Indeed, the "long tail" economy is an economy that can be very valuable for intermediaries, not only producers; a producer (publisher) may in some cases sell a high number of copies from a small number of titles while also relying on a steady income flow from its wider backlist, while an intermediary will directly go for the widest possible catalogue, in particular if it can be built with limited expenditures. The very width of the offer can in some cases be a strategic weapon to win or retain customers. Others could be interested in offering the largest pool of resources in which to perform searches and thus the largest number of results to such searches.

Arrow's value can in this case go as far as to reduce the entry barriers to the market of electronic books, at least insofar as to the part of it that relates to intermediaries that aim at comprehensiveness or near such. By lowering the cost of retrieving information on rights and rightholders and thus simplifying clearing procedures, and by contributing to giving legal certainty to uses of orphan works, Arrow would contrast the natural tendency to form monopolies that characterises markets based on comprehensiveness.

By providing information on what works have been already digitised and for what works a permission has been already denied, Arrow can also reduce wastes of resources due to duplication of work (in particular, duplication of costly digitising operations) or pointless searches for and contacts with rightholders.

In sum, once copyrighted works start being included in large scale digitisation projects, a need will arise to find information on rights and right holders; the following step will be to put in place clearing mechanisms. These will be voluntary for out-of-print works, and possibly somehow defined or backed by legislation in the case of orphan works; in any case, Arrow will be able to play a role in this framework, offering its services to public and private entities.

4 Arrow's potential users

We have seen what kind of value the Arrow services can bring to players of the electronic book market; in this section we will focus on the various categories of potential users and provide some examples.

The information collected for the first edition of this report and confirmed subsequently revealed that of all the European subjects engaged in digitisation initiatives, the ones most likely to have such information needs as to be able to benefit from the Arrow services were libraries.

The reasons for libraries' interest in digitising their collections is fairly clear – making the cultural heritage available online and maintaining their public function in the digital world – and is also backed by a great deal of political will; the main issues to sort out are whether the digitisation initiatives are going to involve works under copyright and whether there are going to be funds available to this purpose. The plans for enhancing the content of Europeana are likely to be an important driver from both points of view.

Libraries can enter into public-private partnerships to have access to a larger pool of resources. In many cases, such partners are not motivated by a specific interest in entering the digital book arena; their aims are rather the support of culture and possibly the good reputation and in some cases tax reliefs that go with it.

Traditional commercial entities of the book market such as publishers would undertake digitisation efforts of their own catalogues, therefore generally not needing to retrieve right holders information; distributors, aggregators and other similar intermediaries expressed no particular interest in orphan works and sourced their virtual stores directly from publishers (and in some cases added freely available classic public domain titles to their offers). In any case, those kinds of intermediaries did not intend to digitise works themselves, and especially to undergo all the necessary steps to identify rights statuses and clear rights.[21]

Anyway, libraries and their sponsors are not the only potential users of Arrow; commercial players can have an interest to, if their objectives have some similarities with those of libraries. As we said, the main characteristics of digitisation initiatives by such players are their purposed large scale (collections of many thousands if not millions of works), the potentially mixed features of the books involved (in terms of right holders and rights status) and a tendency to comprehensiveness (as an attempt to put as little limits as possible to the number of books that are to be digitised, within those available).

Libraries want to digitise and make available as much as they can of their collections, and Europeana has the ambitious goal of providing access to European citizens to most of their cultural heritage. Commercial players can have an interest in the commercial exploitation of orphan works and also of out-of-print works for which the right holders no longer have such interest (at least not to engage in it directly).

As we already pointed out, though, this might not be the main motive. Orphan works in particular are likely to have a limited relevant economic value, at least as long as they remain orphan; out-of-print works are more likely to, but then again many of them are out of print for a reason: the cost of commercialising them, at least on a short term, may be assumed to be too high; there is little to no demand for that title; an updated version of the same work exists - in some cases the circulation of an

---

21 In case publishers had an interest in digitising out of print or orphan works, their models would be comparable to those of book retailers analysed in par. 4.2.

old version might harm the commercialisation of the new one; the author has decided to retire his/her work from circulation; and more.

Yet, intermediaries interested in exploiting the "long tail" can seek advantages in having the largest possible collections for other reasons. Some can be still interested in actual sales, be it because they run a large electronic bookstore or because they are in the market of the reading devices (dedicated or not) and want therefore to provide content as well.

Having the largest catalogue possible can be a goal for such players, as this would make it easier to attract customers and gain their loyalty. If a number of the titles available were for free, this could even be a further enhancer for potential clients. Hence the inclusion of works in the public domain in the offers, but also possibly of orphan works and out-of-print works licensed cheaply; and hence a possible interest in retrieving rights and rightholders information about some works (to ensure a public domain status, to negotiate licences, etc.).

Commercial entities might as well be interested in digitising the largest possible number of books in order to feed their search services: by widening the base of resources available for search, they aim at having the maximum number of results, which again can be a very attractive feature to customers.

It must be noted, anyway, that the characteristics of this kind of market make it more likely to constitute monopolies, as competition tends to develop on grounds of size rather than anything else. In addition, the use of proprietary technologies and closed standards can further contribute to locking the markets and creating dominant positions.

In the private market, only Google has for the time being undertaken an initiative which because of its large scale can be compared to the comprehensive plans of Europeana and national libraries (although the non-systematic pattern of its digitisation efforts and their being based largely on copyright infringements are serious limits). And while Google's model is also evolving (as witnessed by the new Google's Edition project), they could still try and look for some level of comprehensiveness in their Book Search project (now Google Books), whereas the recent legal setbacks[22] they suffered should mean they will need to look for right holders.

Nonetheless, as we said previously, as Arrow's services become available, they are likely to contribute to lowering the entry barriers in this particular market, thus mitigating the tendency to establishing monopolies. Arrow could indeed be a determinant factor in increasing the number of players in the market.

Other players are actually already moving or planning to move into the field of large-scale digitisation - such as ProQuest or Telefónica[23] - and they could soon express a need for identification and subsequently clearance of rights. An independent, wide-reaching and accessible infrastructure that allows retrieving rights information can make it possible for new such intermediaries to enter the

---

[22] On December 18, 2009, the Paris Court of First Instance condemned Google for violating copyright of books published by French publishing group La Martinière, forbidding the search engine to continue digitising books without publishers' authorisation. This decision was based on the application of the French law, whose application Google had attempted to question, and the acknowledgement by the French court of its competence to deal with such a case. The court also acknowledged that the French Publishers Association (SNE) and the Authors Publishers Association (SGDL) were entitled to join the suit. It stated that "by fully reproducing and making available extracts of books" without the authorisation of rightholders, Google had committed acts of copyright violation. The court gave Google one month to apply the ruling and halt such acts or face a fine of 10,000 € per day. Google will also have to pay 300,000 € in damages to the three publishers owned by La Martinière group and a symbolic sum of one euro to the SNE and the SGDL, thus recognising the damages caused to the whole publishers and authors community.

[23] ProQuest is a US IT company specialised in services for libraries; Telefónica is a large Spanish telecom company (see par. 4.3).

market (as the "long tail" economy is more appealing to intermediaries than to producers). In this perspective, we will also make hypotheses about other companies' business models evolving into developing a need for Arrow's services from a position already within the market, or about the chances for newcomers to make a move into it.

The possible role of RROs will also be explored. RROs are expected to administer rights in relation to orphan works; assist authors and publishers in the licensing of out of print works and alternate format production of works to make them accessible to visually impaired persons; license digitisation projects, etc. This may also entail the use of Arrow to identify right status and right holders; Arrow could then represent a cost saving factor for RROs, for instance, when they have to undertake diligent searches. If mechanisms for clearing the usage of orphan works were eventually put in place, users could benefit from the limitation of liability offered by RROs in schemes that saw them involved.

4.1 Digital libraries and Europeana

Libraries are by definition the main potential users of Arrow, since they are the most likely to be involved in digitisation projects for books (and not only). Libraries hold large collections of books, and have an interest in their preservation and making available, which are among their principal tasks. national libraries have a particularly relevant role, as usually they are the main entities responsible for the preservation of national cultural heritages.

In addition, as anticipated, there is a very strong political will, especially at the EU level, to step up the digitisation of the European cultural heritage and to include books to a large extent. Not only: there is a clear intention to include in the plans for digitisation also works under copyright, to avoid what some call the "20th century black hole", the absence of works from the past few decades in digital libraries.

Said political will led as its main result to the creation of Europeana, the European Digital Library. The initiative was prompted by a letter sent to the Presidents of the European Council and the European Commission in April 2005 by the Heads of State and Government of six EU Members,[24] proposing to coordinate the existing digitisation efforts so as to create a sort of European digital library. The Commission welcomed the proposal and in June 2005 presented the i2010 Initiative, which included plans for the creation of digital libraries.[25]

In September 2005, the Communication "i2010: Digital Libraries" outlined the vision and strategy of the digital libraries initiative; the document highlighted the potential costs of establishing the IPR status of works, in particular in the case of orphan works.[26] In August 2006 the Commission adopted a Recommendation on digitisation and online accessibility of cultural material, in which it put forward measures for Member States to make progress in this field; the Recommendation noted that part of the material held by cultural institutions was protected by intellectual property rights and suggested licensing mechanisms in areas such as orphan and out-of-print works to facilitate rights clearance; it also stressed that "Europe's cultural material should be digitised, made available and preserved in full respect of copyright and related rights".[27]

---

[24] France, Germany, Hungary, Italy, Poland and Spain.

[25] Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions "i2010 – A European Information Society for growth and employment" - COM(2005)229, 1 June 2005 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0229:FIN:EN:PDF).

[26] Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions "i2010: Digital Libraries" - COM(2005)465, 30 September 2005 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0465:FIN:EN:PDF).

Member States reacted adopting a Council Conclusion in November 2006, which invited to reinforce national strategies and targets for digitisation with full respect for Community and international legislation in the field of intellectual property. Activities and goals included encouraging agreements between right holders and cultural institutions to make copyrighted material accessible online on contractual terms, and having mechanisms to facilitate digitisation and online access of orphan works and out of print works, while fully respecting content owners' interests and rights.[28]

In September 2007 the European Parliament adopted a Resolution recommending setting up a European digital library; the document suggested the inclusion of in-copyright works in the digital library (explicitly mentioning orphan and out-of-print works) whilst strictly respecting intellectual property laws.[29]

In August 2008 a new Communication by the Commission on progress on digitisation and online accessibility of cultural material announced the development of a European digital library called Europeana. The document highlighted the challenge of including in-copyright material in the project (to avoid the "20th century black hole") and called for national efforts to back European initiatives on orphan works (explicitly mentioning the Arrow project among those).[30]

The first prototype of Europeana was launched in November 2008 and provided access to some 2 million cultural items, including books. Its launch was accompanied by Council Conclusions supporting digitisation and online accessibility of cultural material. The Council invited Member States to take into consideration the objective of including protected works in Europeana while respecting national and Community systems of copyright and to establish mechanisms to facilitate digitisation and online access to orphan and to out-of-print works, while fully respecting right holders' rights and interests. It also invited the Commission to examine the feasibility of increasing financial support for the digitisation of cultural material.[31]

In August 2009, the Commission issued a new Communication to define the next steps for Europeana. The document set the goal of having 10 million objects accessible through the site in 2010 and reaffirmed the intention to include in-copyright material through collaboration with right holders and in full respect of copyright legislation. It also called for urgent action in the field of orphan works and pointed at Arrow as a possible part of the solution. It addressed as well the need for additional financing for Europeana after the development phase and made some suggestions (on public-private partnerships and public funding).[32]

In May 2010, the European Parliament adopted a new Resolution on "Europeana - next steps", which again called for stepping up digitisation efforts, increasing Europeana's content, including in-copyright works (particularly orphan and out-of-print), respecting copyright law and providing sustainable funding

[27] Commission Recommendation of 24 August 2006 on the digitisation and online accessibility of cultural material and digital preservation (2006/585/EC) (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:236:0028:0030:EN:PDF).

[28] Council Conclusions on the Digitisation and Online Accessibility of Cultural Material, and Digital Preservation (2006/C 297/01) (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:297:0001:0005:EN:PDF).

[29] European Parliament resolution of 27 September 2007 on "i2010: towards a European digital library" (2006/2040(INI)) (http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2007-0416+0+DOC+XML+V0//EN).

[30] Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions "Europe's cultural heritage at the click of a mouse - Progress on the digitisation and online accessibility of cultural material and digital preservation across the EU" - COM(2008)513, 11 August 2008 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0513:FIN:EN:PDF).

[31] Council conclusions of 20 November 2008 on the European digital library EUROPEANA (2008/C 319/07) (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:319:0018:0019:EN:PDF).

[32] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Europeana: next steps - COM(2009)440, 28 August 2009 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0440:FIN:EN:PDF).

to the digital library through sponsorships, public-private partnerships and public financing. The report welcomed the Arrow project as a tool to facilitate the retrieval of rights information and the clearance of rights.[33]

Finally, the same month, the Council adopted a new series of Conclusions on "Europeana: next steps", once more calling for enhancing the digitisation processes of the European cultural heritage in full respect of intellectual property rights, and for making rapid progress in finding solutions for digitising and making available online out-of-print works, and for addressing the orphan works issue. The Council also invited Member States to continue and strengthen their support to digitisation projects carried out by their cultural institutions and to provide until the end of 2013, on a voluntary basis and pending the development of a sustainable funding and governance model, the necessary level of support to Europeana as a complement to Community funding; and it invited the Commission to make proposals for the sustainable financing of Europeana post 2013.[34]

In short, the European institutions have unanimously embraced the idea that the European cultural heritage (including books) must be digitised in order to preserve it and make it available and have as well explicitly called for the inclusion of copyrighted works in the process, while fully respecting copyright; the issues of orphan and out-of-print works have been singled out as clear priorities.

It will be of course up to individual libraries, and in particular national libraries, to carry out good part of the digitisation activities (either by themselves - in house or outsourcing them - or through partnerships). Considering the basic requirements of the digitisation plans and the importance given to orphan and out-of-print works, Arrow's role as facilitating the identification of the rights status of books and the localisation of right holders, and as a consequence facilitating also the licensing of said works, makes of its services an ideal tool for digitisation initiatives, especially those linked to Europeana. Arrow's intended working in fact matches very well all the requirements set out by the institutions with regard to the steps to increase the content available on Europeana in the near future.

Europeana is poised then to be the main potential "customer" of Arrow (though indirectly, through its contributing libraries and cultural institutions in general). As said, the European Digital Library plans to reach 10 million digital objects by the end of 2010 (there are currently 6 million,[35] of which some 200,000 are books), and 25 million by 2014.[36] Whereas this is an ambitious plan, potentially implying a large number of mass digitisation initiatives across Europe, there is no indication of the proportion of those objects that should be represented by books. The same can be said of the many documents stating the political will to enhance the digitisation of the European cultural heritage. Nevertheless, statements that have been made by institutions with digitistsation plans and examination of current digitisation projects give reasons to believe that the intention would be to include a substantial number of books.

Some indications can also come from individual libraries' digitisation plans, be them related or not to Europeana, and from any studies based on cultural institutions' digitisation initiatives, current and planned. For example, in 2007-2009 the Commission funded a study, called Numeric, aimed at

---

[33] European Parliament Report of 5 May 2010 on "Europeana - the next steps" (2009/2158(INI)) by MEP Helga Trüpel (http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A7-2010-0028+0+DOC+PDF+V0//EN&language=EN).

[34] Council Conclusions of 10 May 2010 on Europeana: next steps (2010/C 137/07) (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:137:0019:0021:EN:PDF).

[35] July 2010.

[36] Outline Business Plan for Europeana as a service of the EDL Foundation, November 2008; available at: http://version1.europeana.eu/c/document_library/get_file?uuid=0c6c6078-8026-4297-9367-dd6d14b73c2e&groupId=10602.

measuring and reporting progress in digitisation initiatives of the European cultural heritage.[37] According to the Study Report, some half of the European national libraries had plans to digitise their collections, and so did about one third of other libraries (higher education, public, etc.); still, only 3.5% of the digitisation work had been carried out by national libraries, a little more by higher education libraries and almost one third by public libraries. The plans involved around two thirds of all the holdings of the national libraries, and a lower proportion for other libraries.

Again, while these figures are noteworthy, they do not provide for precise plans. However, the report estimated that in all kinds of institutions (national libraries, public libraries, higher education libraries and others) a total of over 220 million books and serials could be digitised, while only some 4 million had been already. Even taking into account that as we just mentioned probably about half of the collections are not included in digitisation plans, the potential is very high. The study estimated that some 90 million books (rare and not) should enter digitisation plans by libraries.[38]

Another research, a report on digitisation in European national libraries that was published as part of the EDL (European Digital Library) Project in February 2008,[39] provides some more precise information, although more limited in scope since it regards only national libraries. According to the report, the situation regarding digitisation of books by national libraries was critical, especially for 20th century material (mainly in-copyright works, therefore); the main reasons for this situation, according to the libraries, were difficulties with copyrighted works, but even more the costs of digitisation. The report provided an overall estimate of the targets of digitisation projects and programmes by national libraries, which envisaged passing from some 280,000 digitised books in 2006 to more than 1.5 million in 2012. The figure is fairly low, although it only takes into account national libraries.

Some more information can be found in the individual country reports by the Member States Expert Group (MSEG) set up to monitor progress and exchange information and good practices of Member States' policies and strategies on the digitisation and online accessibility of cultural material and digital preservation.[40] The reports provide in fact information on the quantitative and qualitative digitisation targets adopted by Member States: for example, the National Library of the Netherlands included in its strategic plan for 2010-2013 the intention to digitise 10% of all Dutch books, newspapers and periodicals;[41] Slovakia plans to contribute 0.5 million books to Europeana by 2015; the British Library plans to digitise some 80,000 books of 19th century literature; and so on.[42] Among the most active in this field, the French National Library in February 2010 digitised its millionth document, and it continues its efforts at a rate of 1,500 documents per day.[43] The German National Library, in turn, coordinates the German Digital Library, an ambitious programme that could digitise 5.5 million books in its first ten years (but needs to find the adequate funds); as the initiative is intended to be carried out in respect of copyright, a way will have to be devised to deal with rights information and clearance.

---

[37] Numeric Study Report, May 2009; available at: http://cordis.europa.eu/fp7/ict/telearn-digicult/publications_en.html.

[38] As we mentioned early, given the potential scale of the operations, it will be very important for libraries to do their best to avoid duplication of digitisation efforts; Arrow can help with that, too.

[39] EDL Report on Digitisation in European National Libraries 2006-2012, available at: http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/downloads/EDL-D3%201%20final.pdf.

[40] The Group, set up in 2007 in replacement of an informal intergovernmental group on digitisation, is made of representatives from the national ministries and/or national cultural institutions of all EU countries. Their periodical reports can be found at: http://ec.europa.eu/information_society/activities/digital_libraries/other_expert_groups/mseg/index_en.htm

[41] The long term digitisation programme of the Library actually aims to make every book and article printed in the Netherlands available online within the next few decades, where copyright legislation permits; part of its work consists in identifying and if possible resolve copyright issues.

[42] More examples: the Polish National Library estimated in 2008 that 1.6 million titles among its holdings should be digitised and made available online; the Czech National Library intends to digitise 17% of its book collections by 2015.

[43] Report Albanel (see note 20).

Undoubtedly, then, the next few years are likely to witness a number of digitisation programmes by libraries involving hundreds of thousands if not millions of books per year; and we already know that a large number of books in collections are likely to be out of print, and some probably will be found to be orphan.

Still, a certain degree of uncertainty remains with regards to such potential initiatives. First of all, there is the question whether *inter alia* governments will follow up their declarations and ensure that there will be sufficient funding to actually carry out the necessary activities, planned and not, in order to fulfil the political objectives set; in addition, it is unclear whether any new initiatives on a considerable scale will involve works under copyright and therefore need to look for information on rights and rightholders and possibly for clearance mechanisms for digitisation and making available.

On the financial side, the EDL report mentions that only a few libraries have an explicit and ongoing budget for digitisation, of between 100,000 and 200,000 € per year. The individual reports of the MSEG mostly confirm this information, as in general quite larger budgets are reported but for all kinds of digitisation activities. There are anyway cases of more substantial resource allocations, in a few countries where digitisation plans are rather advanced. The Norwegian National Library, for example, has a budget of 2.6 million € per year, and over 1.7 million € have been allocated to develop the Spanish Digital Library in 2010.[44] The Numeric study, in turn, estimated that all in all libraries had a budget of some 33 million € available for digitisation (25 of which allocated to national libraries). Still, put in perspective in the light of the costs of digitisation of books, these figures reveal that budget so far allocated are most likely insufficient for real mass digitisation projects, at least able to match the ambitions underlying.[45]

An interesting possible exception in this scenario is that of France; there the Government has launched a national loan, intended to raise 35 billion € of which 2.5 will be allocated for the digital economy. In turn, 750 million will be destined to the digitisation of the national cultural heritage and 142 million out of those would go to the French National Library for the digitisation of various kinds of materials, including books.[46]

In addition, the funds the European Union is going to devote to the digitisation of our cultural heritage must be taken into account. The same documents that state the political will to enhance the digitisation process, in fact, also give indications as to its financing. The Commission Communication on progress in digitisation and making available of cultural content[47] said the Commission would continue to support projects that enhance the online accessibility of cultural content and digital preservation and contribute to the development of Europeana.[48] The Communication on the future of Europeana[49] acknowledged the case for sustained public sector financing of the initiative, also based on the Council

---

[44] See country reports, note 39.

[45] As seen already in the First Edition of this Report, the costs of digitising a book, especially starting from a printed copy, are very variable and sometimes quite high. The EDL study mentions from a minimum of 0.08 to 0.15 € per page (16 to 30 € for an average 200-page book) to as much as 1.30 € per page (260 € per 200 pages) considering all costs related - adding metadata, OCR (Optical character recognition), selection and preparation of materials, quality assurance, hard and software, etc. The Numeric study reported a median cost among libraries ranging from 0.10 € per page (higher education libraries) to 0.70 € (public libraries), with national libraries in between at 0.48 € (respectively 20, 160 and 96 € per 200 pages).

[46] The French Minister of Culture, Frédéric Mitterrand, recently put forward the idea to digitise a large number of out of print works, and orphan works as well, including some 500,000 to one million books. The financing, however, as we will see below, will have to be matched by private funding, as it will be a loan.

[47] See note 30.

[48] The document mentioned a budget of € 69 million for the period 2009-2010, for the specific objective of digital libraries and digital preservation included within the 7th Framework Programme for Research and Development. It also foresaw an amount close to € 25 million to be allocated for digital libraries in 2009 and 2010 within the Competitiveness and Innovation programme.

[49] See note 32.

Conclusions on Europeana[50] that recommended active sustained support for digitisation and to examine the feasibility of increasing financial support[51] for the digitisation of cultural material and contributing to Europeana. The Council Conclusions on the next steps for Europeana[52] invited Member States to continue and strengthen their support to digitisation projects carried out by their cultural institutions. Finally, the Commission allocated several million € for funding projects aimed at digitising content for Europeana in the next few years.[53]

The other element of uncertainty, even more relevant, is the possible inclusion of works in-copyright in the digitisation plans of libraries. While this is certainly not the only way to make copyrighted works available to the public in electronic format,[54] it is the way that most likely would require a rights information service such as Arrow's. Upon enquiring with the national libraries of some of the largest countries in the EU, the first finding is that where there are mass digitisation programmes, these do not include copyrighted works. As we have already mentioned, public domain works are often given the priority, for various reasons, ranging from the urgency of preserving the oldest works to the possible difficulties in handling copyright related aspects. For instance, a survey of Dutch cultural institutions found that expectations were that by 2012 digitisation would be achieved of 1% of all pre-1850 Dutch books; Slovak national projects for the time being exclude orphan works from digitisation; the British Library plan involves 19th century literature[55]; the Spanish National Library mass digitisation plans involve works in the public domain; also in Germany, current concrete plans for mass digitisation only exist for works in the public domain; and so on. Similarly, when the Italian Ministry of Culture signed an agreement with Google for digitising collections owned by a number of Italian libraries, the agreement explicitly excluded copyrighted books.[56]

As we said before, there is the intention to include protected works in the German Digital Library project, and the subsequent necessity to respect copyright has prompted the search for solutions, still to be devised. On the other hand, in France, access to the national loan funds for digitisation projects could be linked to finding commercially viable solutions to repay grants; this might increase the incentive to find agreements with rightholders and enter into public-private partnerships for the exploitation of works, especially those in copyright.[57] In Italy, both parties of the mentioned agreement declared that the inclusion of copyrighted works will be possible only with the agreement of the rightholders community and following models set with them. Any such model would require a system of information management, such as Arrow's.

Furthermore, at least in terms of planning and discussions, there are constructive dialogues between libraries and rightholders and RROs to further develop practical solutions to defined digital libraries

---

[50] See note 31.

[51] Through the existing Community programmes and within the framework of the current financial perspective.

[52] See note 33.

[53] As part of the digital libraries area of the ICT Policy Support Programme (ICT PSP), a Commission programme included in the Competitiveness and Innovation Framework Programme. See:
http://ec.europa.eu/information_society/activities/ict_psp/documents/ict_psp_wp2010_agreed_at_committee_191109.pdf.

[54] Publishers themselves, for example, as they enter the e-book market by digitising their catalogues, fulfil this task. Rightholders in general are the first who have - and should have - the opportunity to digitise and make available their works.

[55] Likely to be out of copyright, although for the latter part of the century there is a good probability to find many works still protected.

[56] Likewise, the Polish plans include only books in the public domain, and the Portuguese plans involve printed books up to 1850.

[57] It is worth mentioning a declaration of December 2009 by the French Interassociation of Archives-Libraries-Documentation (IADB, Interassociation Archives-Bibliothèques-Documentation) on the digitisation and making available of orphan works in full respect of the law. The declaration in fact basically accepts the essential notion of the need for a diligent search prior to the determination of the status of a work and calls for an easy way of conducting such search. Again, Arrow is meant to be the kind of tool to make diligent search easier. Text of the declaration: http://www.iabd.fr/spip.php?article91.

challenges. In Germany, negotiations between rightholders and the National Library are under way on digitisation plans for out of print works published up to 1965; rights clearance would be managed collectively by the RRO, while diligent search to determine the status of works should be conducted via Arrow. Similar plans, for out of print works in the scientific and academic fields, are at an earlier stage of discussion in Denmark. In France, a proposal is under consideration to introduce compulsory collective management of orphan works, based on prior diligent search, and there is an intention to find an agreement also for out of print works (Arrow could play a role in these cases as well).[58]

Still, even when a digitisation plan is supposed to only include works in the public domain, at some point it might be appropriate to check whether the works presumed to be in the public domain really are, and here again a role can be envisaged for services like Arrow's.

Of course, not all digitisation projects by libraries will be linked to Europeana; still, they can have a need to retrieve rights information and clear rights, and therefore potentially require a service like the one Arrow will offer. Such initiatives are relatively common, often limited in scale, mostly driven by thematic considerations (focusing on a specific topic, a territory, an event, etc.); sometimes they are carried out to meet a known demand and are and supported by a sustainable business case (a revenue model - like with digitisation-on-demand, which we will see later on- or a judgement of social and/or cultural impact and value). University and research libraries with special collections are particularly likely to have similar plans.

A typical example is the pilot digitisation programme by the Wellcome Library built around the theme "modern Genetics and its Foundations". The initiative, funded by the Wellcome Trust, in addition to procuring the necessary infrastructure to develop and deliver the "Wellcome Digital Library", will allow the Library to digitise relevant content, including a number of books (around 1,500 in total). Those books are all relevant to the study of genetics and were published between 1850 and 1990; those in charge of the project aim at identifying the right status of the works involved (whether in print, out of print or orphan) and seeking permission from rightholders to digitise and make the works available for free through the web. For works identified as orphan, they will make these items freely available on the web while providing a take-down notice: in the event that a rightholder turned out, they will remove the item form the web, and then negotiate to make the content freely available. This kind of project is a good example of an initiative that could do with a service such as the one offered by Arrow.[59] Those responsible for it will actually cooperate with Arrow for the testing of diligent search and also have expectations on the use of the system once operational.

The National Library of Wales is looking towards creating a mass digitisation project for all holdings of Welsh and Welsh-language books and it is the intention of those in charge to explore models which would allow works in copyright to be part of that project. In Spain, the documentation centre Emakunde- Instituto Vasco de la Mujer, specialised in gender equality, is planning to implement an online consultation and loan service, and is therefore preparing to digitise part of its collection (including a small number of books). The institute is negotiating with a Spanish collecting society to clear rights for the 124 works of publishers that are its members. Again, it is the kind of initiatives that would greatly benefit from a tool like Arrow.

In all such cases, Arrow is set to provide help both by reducing the cost of rights search and clearance and by facilitating the inclusion of in-copyright works in digitisation initiatives. Therefore, as long as there are going to be funds for projects to put into practice the clear political will to digitise and make

---

[58] For more information, see the Arrow deliverable D3.5. Report on legal framework - Edition 2, available on www.arrow-net.eu.

[59] The project managers will actually cooperate with Arrow by providing evidence of diligent search cases, and are counting on the Arrow infrastructure to facilitate the realisation of their initiative.

available the European cultural heritage, libraries, insofar as they are involved in the process, will have an interest in using a tool like Arrow.

Moreover, as we mentioned already, Arrow is actually likely to increase and accelerate digitisation plans, considering its potential role in removing obstacles linked to the search of rights information and enhancing the possibilities for rights clearance, as well as avoiding the costly duplication of digitisation efforts. There is indeed a certain degree of expectation by the library community regarding the Arrow system, with some institutions seemingly waiting for it to be operational before making their decisions about digitisation.[60] This highlights even further the potential for involvement of a tool such as Arrow in any plan for mass digitisation by cultural institutions.

4.2 Private players

We have already briefly examined the emergence of a number of new players in the digital book value chain: search engines offering tools to search books online, online shops developing their own bookstores, device manufacturers entering the content market, mobile carriers and internet service providers. In many cases, such actors have taken advantage of their major (sometimes dominant) roles in connection to consumer in other fields: Google in the search business, Apple in consumer electronics, Amazon in online retail sales, etc.

Their interest in entering the e-book market, as we have seen, can be linked to the will to offer content commercially (to complement their service offers, feed their devices, expand their sale areas) or to provide access in the form of search results (in models usually based on other sources of revenue, mainly advertisement). In any case, these players have a particular interest in the availability of the largest possible collections; beyond the value of individual items, for those seeking to exploit the "long tail" on a large scale, a certain degree of comprehensiveness is essential.

This also depends from the fact that in the digital economy, for a number of players the primary currency is not content but attention and authority. For some industries it is difficult to build viable business models based on primary content transactions, particularly as the proliferation of new content tends to devalue the monetary potential of individual assets. Even in the book sector, where the intrinsic value of content is still paramount and has not been diminished by the switch to digital, some of the new players might find content less valuable in its own right.

Hence, some private actors seeking a return on their investment in digital activities are likely to focus more on the role of content in helping to achieve an economically valuable audience, in building communities and platforms which aggregate a significant number of users. Again, it is the value of comprehensiveness these actors are interested into. When this is pursued in order to be able to present customers with the largest possible catalogue - it is the case for device manufacturers and online retailers, and all those more sales-oriented - the aim is to attract and also retain those customers (hence the interest in including also public domain works, available for free, and possibly orphan and out of print ones); when the aim is to expand to the maximum the basis for searches performed by users - clearly, in the case of search engines - the ultimate goal is to outperform competitors in terms of completeness of results.

It is clear that for these actors, no matter the sector they come from, an important goal is to gain a large market share; this is because, as we anticipated already, in markets where comprehensiveness is relevant there can be a natural tendency to forming monopolies, as competition is based mainly on size. Offering a catalogue larger and more complete than that of the competition or granting a larger

---

[60] As Arrow is a neutral instrument, digitisation plans by libraries will also depend on the legal framework.

number of results for searches can actually confer to some players a dominant position relatively easily. In addition, these actors are obviously interested in being able to put a price as low as possible on the content they provide (zero in the case of search engines delivering search results), and therefore to secure content for their operations at a low cost as well.[61]

All of this is related to another important aspect of this part of the market: the players we are talking about also have in common the fact that they are not the ones that have invested in the creation of the content they are dealing with in the first place.[62] This therefore leads them to either buying digitised content (actually, licensing the right to sell it) or to digitising it themselves, again upon licences or, where possible, for free.[63] Therefore the use of Arrow, be it to identify works that can be digitised for free, to locate rightholders for negotiating the use of protected works, to ensure the respect of copyright in some digitisation project, or in general to improve the legal certainty of those willing to digitise some kinds of works (but then, only if coupled with an appropriate legal framework), can fit very well the needs of this kind of private actors.

As we mentioned earlier, Google was the first to engage in this kind of activity, starting the mass digitisation of the content of a number of libraries; the main problem of its strategy was the lack of respect for copyright, which provoked the understandable reactions of rightholders all over the world and of course seriously hampered the company's plans. Amazon followed soon; the giant online retailer, already very strong in the sale of physical books via internet, entered the e-book market as well with its own reading device, the Kindle. Then Apple came, the device manufacturer, which decided to extend its strategy of selling content for its devices also to the e-book sector.

In this section we will have a look to the activities of these and other private players in the field of digitisation and making available of books, all of which might have an interest in using the services that a tool like Arrow can offer: easier identification of the rights status of works and easier localisation of rightholders, leading to better chances to negotiate licences for the reuse of works (thus catering for those interested in specific categories of works - public domain, orphan, etc. - and for those willing to negotiate the largest number of licences). Arrow, as we already said, can also mitigate the monopolistic tendencies of this kind of market.

4.2.1 Search engines

The emergence of digital libraries received a strong impulse from the development of usage practices related to making searches on the internet, in particular the universal use of search engines.[64] The possibility they offer to access a virtually unlimited range of resources is a valuable asset for users, all the more so if these resources include items such as books and their content. This in turn constitutes an incentive to digitise content and make it available online, in particular for search engines, which have an interest in having available the largest possible base of documents in order to improve the wealth and relevance of their search results (and thus, ultimately, increase their advertisement revenues).

The bigger the search engine, the stronger interest they have in having their systems harvest the largest possible amount of content. It is not surprising therefore that Google was the first to launch a

---

[61] This element, coupled with the tendency to acquire dominant positions in some field, poses serious risks of abusive behaviours, in terms of price-setting policies or of violations of rights of rightholders; several cases have already been reported.

[62] This is another relevant incentive to adopt abusive practices (see note 61).

[63] When works are in the public domain or when rightholders agree to such conditions.

[64] The issue is treated more in detail in the *Rapport sur la numérisation du patrimoine écrit* (Report on the digitisation of the written heritage, January 2010), by Marc Tessier; available at: http://www.culture.gouv.fr/mcc/Actualites/A-la-une/Mission-sur-la-numerisation-du-patrimoine-ecrit/Rapport-Tessier.

platform to this end, in October 2004, called Google Print (it became Google Book Search a year later, and it is now called Google Books). Google planned to digitise 15 million works in 10 years, relying as a start on the collections of a number of US and UK libraries.[65]

As Google's platform did not allow other search engines to index the content it hosted, other players of the sector engaged in similar initiatives. In December 2006, Microsoft launched its own digitisation programme; the e-book platform Live Book Search became one of the services linked to Microsoft's search engine Live Search, and made agreements with a number of libraries to obtain content.[66] However, the project did not take off as expected and was finally abandoned in May 2008. Also Yahoo! tried to develop its own digitisation programme: in partnership with the Internet Archive[67], it established the Open Content Alliance[68]; their website hosts 1.6 million books, either in public domain or licensed by rightholders. Again, these and similar initiatives would benefit from and could be potential users of Arrow.

Google Books has eventually become the largest of these projects. Its platform hosts a database and an internal search engine and stores and indexes the content of books digitised by Google. The platform's content is also accessible through the general Google search engine, in order to enrich its results. Users can perform full text searches of the books, and then, according to the works' status and to whether Google has made an agreement with the rightholders, they can access and download the whole book, view a few pages and be redirected to the publishers' sites (for partner publishers), or see just a few "snippets".[69]

The main sources of material for the website are the libraries that have signed agreements for the digitisation of their collection, generally of public domain works, and the partner publishers, which provide metadata of their works to Google and have their websites linked. Nonetheless - and this is by far the main shortcoming of Google's initiative - the project also involved (and still does) the digitisation of works in copyright from the catalogues of large US libraries without asking prior permission to rightholders. Indeed, at the beginning of 2010 Google announced that the Google Books platform allowed searching in full text more than 10 million books, of which 2 million had been digitised through publisher partnerships, 1.5 million were in the public domain and the rest were works in copyright digitised without the consent of right holders.

The digitisation, display online (though partial) and potential for commercial exploitation on behalf of Google of works for which it did not have permission form the legitimate rightholders obviously provoked widespread commotion in the publishing world. Longstanding litigation led the publishers' and authors' representatives in the US to sign a settlement agreement with Google; this underwent harsh criticism from a number of rightholders not represented and from authorities in the US, so that an amended version was submitted before the judge in charge could decide upon the original version. One of the main amendments restricted the scope to books published originally in the US, the UK, Canada and Australia, as well as those registered in the US Copyright Office; this means a large number of works by third party rightholder (many European) are still included in the agreement, which

---

[65] The largest partners of this project were the New York Public Library, the University Libraries of Harvard, Stanford and Michigan and the Bodleian Library in Oxford.

[66] The main libraries involved were the British Library, the New York Public Library and the University Libraries of Cornell, Toronto and California.

[67] A US non-profit founded in 1996 to create an internet library. More on it in paragraph 4.2.2.

[68] Other participants included the University Libraries of California and Toronto and the British National Archives.

[69] Books in the public domain can be viewed entirely and downloaded for free. For books in copyright, if Google has made an agreement with the publisher users can view a certain number of pages and are redirected to the publisher's website for further access, according to the conditions set by the publisher. If there has been no agreement, users can see information

would allow Google to keep using their works against their will. In addition, another potential consequence of the approval of the amended settlement (the judge's decision is due soon) would be that Google would end up having a monopoly on the exploitation of works "unclaimed" by rightholders, i.e. in a way to somehow "orphan" works.[70] Further developments will likely occur once the judgement on the settlement is passed, but its approval is far from certain.

Google is continuing its digitisation activities in the US, and it is expanding into Europe through agreements with libraries for digitising public domain works;[71] as the most recent data talk of 12 million books digitised, 3 million more will have to be digitised to reach the 15-million goal, and the process could go on for long afterwards. Google's Books project could therefore be the paradigm of the kind of potential private user for the information infrastructure that is planned by Arrow. This of course this depends on the direction Google takes in pursuing its objectives: if the company accepts to abide by the rules of copyright and ask for prior consent before digitising books in the US, it will have to make some sort of diligent search for rightholders. The same applies in case there were doubts on the nature of works presumed to be in the public domain that it is digitising in Europe, or if it were to include copyrighted works in its European activities.

Interestingly, however, even an initiative like Google's needed to establish a rights information management system, the Book Rights Registry (BRR). The BRR contains information about the books digitised and allows rightholders to claim their works; its purposes and functioning bear some resemblance with some elements of the Arrow system (in particular some RRO functions). If the settlement is approved, the BRR would end up handling works from the countries included in the agreement, with all the related limitations; for European works in general, the use of a system like Arrow's would allow for the conclusion of a wider variety of agreements.

It is also worth noting that later in 2010 Google is expected to launch its new e-bookstore, Google Edition, where it will sell e-books in agreement with their publishers. This is not strictly relevant for Arrow, but it completes the general picture; Google is in fact clearly aiming at Amazon's market share (and Apple's), and to do this it will make its digital editions accessible via a wide range of devices (to be more attractive than closed, proprietary systems) and aims at having a greater selection than its rivals in the field (thanks to the 12 million books already digitised). And the potential monopolistic position that the settlement, if approved, could provide Google, adds one more comparative advantage to their arsenal.

4.2.2   Other commercial and non-commercial players

As anticipated, various other players have entered or might enter the field of digitisation and making available of electronic books, and therefore might have the need for a rights and rightholders information service. Some of them entered the e-book market with a view to taking advantage of a consolidated position in other markets; besides search engines, we have those who decided to start providing content either because they were already doing it in a related field or because they had already a strong position in the consumer electronics market and wanted to couple that with content distribution.

---

about the book and possibly some "snippets" (short extracts). Links to bookstore and library websites are also always provided.

[70] This is because the settlement only shields Google from possible legal actions, and not any other player that wanted to use the same works. This has raised serious concerns among competition authorities in the US and abroad.

[71] More details below, in paragraph 4.3.

For the time being, the most important players of this kind are Amazon and Apple.[72] Amazon, a leader in online retail (notably of books), developed an integrated offer of electronic books and created its own reading device - the Kindle - to this purpose. With a large catalogue of e-books (see above) and a reputation among book buyers, Amazon has a solid position in the sector. Apple, for its part, is a leader in device manufacturing, dominating the consumer electronic market with products such as the iPod and the iPhone. Its recently launched iPad further consolidates Apple's interest in having a presence in the content market as well, including that of e-books.[73] Other players have entered the market this way, and more could in the future; Sony, for example, again a consumer electronics manufacturer, produces its own reader device and has an e-bookstore as well.

Google's planned digital bookstore may eventually encourage others, especially Amazon, to reduce the price of their devices. It will also put pressure on other e-book retailers of sort to further expand their catalogues. Having the most comprehensive collection of books on offer will become a great competitive advantage for any of these players and make it easier for them to gain and retain customers. As we said before, intermediaries can benefit significantly from small sales of a wide variety of titles,[74] therefore they will go for the largest possible stock, including the most very cheap if not free titles.[75] Even if little or no gain is made on a number of individual titles, either the total sale figures that can be achieved through a wide customer base or the increased sales of devices (both Amazon[76] and Apple sell them) will eventually compensate.

Hence an interest for such actors to both license content from as many rightholders as possible and to digitise the content that for whatever reason is not made available by rightholders (orphan works, out of print works for which rightholders grant permission, and of course the public domain). For any digitisation plans that might involve works in copyright, Arrow will offer a way to facilitate rights information searches and licensing, thus enhancing the possibilities to widen catalogues and lowering the barriers to entry in the market.

There are also a number of non-commercial entities or projects that carry out digitisation activities on a large scale, in order to build digital libraries and make them available online. These are mainly voluntary projects, often funded by grants or sponsorships, that most of the times aim at displaying the world's cultural heritage and focus on public domain works.

The Internet Archive[77] (mentioned above) is one of those; it has been scanning books and making them available for researchers, historians, scholars, people with disabilities, and the general public for free on its website since 2005. It partners with the University of Toronto and over 150 libraries and universities around the world to create a freely accessible archive of texts and, through its network,

---

[72] The case could be extended to other large players in e-book retailing, such as the US chain Barnes & Noble, which also sells its own reader device, the Nook.

[73] Although the iPad itself seems more adapt to reading newspapers and documents, its book reading capacity has raised much interest. 5 days after the tablet's launch, Steve Jobs, CEO of Apple, declared that 600,000 e-books had been already downloaded by the first iPad owners. A month later, downloads had passed 1.5 million.

[74] If it is true that for many publishers backlists are an important and especially steady source of income, it is all the more important for online retailers to be able to offer the backlists of as many publishers as possible.

[75] It is telling that in May 2010, the top 10 Kindle bestsellers were all free. Amazon, moreover, has a strong interest in being able to continue claiming to be the world's best-stocked bookstore; it is a business built - at least to some extent - on comprehensiveness.

[76] One possible interpretation of the fact that Amazon has been selling e-books at a loss is that it would be a way to sell more Kindles.

[77] http://www.archive.org/details/texts.

currently offers access to more than 2 million free items. Another one, linked to the previous, is the Project Gutenberg,[78] which offers over 33,000 books for free download.

These initiatives, too, need to make sure the books they are digitising are really in the public domain (unless they set a remote time threshold and renounce to digitising any more recent works). In addition, various factors - including different copyright regimes (in particular between the US and Europe) - may create some uncertainties in terms of usage of the works made available.[79] The Arrow system could help with all these difficulties and improve the quality of the service offered by these and other similar entities, especially for access from within the EU.

4.3 Public-private partnerships

Public-private partnerships (PPP) in the field of digitisation are usually agreements between public institutions (the cultural institutions responsible for a digitisation programme: libraries, archives, etc.) and some private players, according to which the private actors contribute - with funds, expertise, or both - to the digitisation of the collections of the institutions.

Such partnership offer a solution to libraries, whose budgets are usually quite insufficient for undertaking mass digitisation programmes; the advantages for the private contributors can be simply in terms of reputation gains as culture patrons (and possibly subsequent tax benefits), or entail more direct returns. The latter is the case of the players we described in the previous section; those were all indeed somehow kinds of PPP, since private parties willing to digitise books have to address libraries, the entities holding the books.

This should not make us forget, however, that the primary way of digitising and making available in-copyright works is through the rightholders themselves. In this case as well PPP are possible: public authorities can contribute financially to the digitisation efforts by publishers, for example, to stimulate or accelerate such processes, especially when it comes to backlists. This is what is happening in France, in connection with the Gallica project, and to a lesser extent in Spain, for the Enclave project.[80] Nonetheless, as in these cases there is likely no need for a rights information search tool, we will focus on PPP between libraries and private players.

Recourse to similar solutions has been advocated (or at least suggested) also by the European institutions in most of the recent documents on digitisation and on Europeana, especially in connection with the future financing and sustainability of the European Digital Library.[81]

There have been some examples of cultural institutions embracing this model to support digitisation already in the past. In Spain, the telecom company Telefónica has recently begun a mass digitalisation project of the collections of the Spanish National Library, through technological cooperation which will

---

[78] http://www.gutenberg.org/wiki/Main_Page.

[79] For example, the Internet Archive's copyright policy states: *"The Internet Archive respects the intellectual property rights and other proprietary rights of others. The Internet Archive may, in appropriate circumstances and at its discretion, remove certain content or disable access to content that appears to infringe the copyright or other intellectual property rights of others. If you believe that your copyright has been violated by material available through the Internet Archive, please provide the Internet Archive Copyright Agent with the following information…"*. The Project Gutenberg website suggests: *"If you don't live in the U.S. you have to check the copyright laws of your country before downloading an ebook! PG does not know the copyright status of any of its ebooks in any country except the U.S. You may download a copyrighted ebook for your personal use but have to contact the copyright owner if you want to redistribute it."*

[80] The two cases have been described extensively in the First Edition of this Report.

[81] All those documents have been listed and reviewed in paragraph 4.1.

be ongoing until 2012.[82] The aim of the project is to disseminate the cultural heritage on the internet and increase the content in Spanish on the web by contributing more than 25 million pages of different types of works (books, engravings, maps, manuscripts, press and photographs) which can be consulted freely and free of charge on the National Library website.

In the UK, the British Library has achieved significant advances in the digitisation and online accessibility of its collections through a number of medium to large-scale partnerships: the first one, with Microsoft, announced in 2005, aimed to digitise 25m pages across more than 100,000 out-of-copyright books; it came to an end in 2008, having digitised some 60,000 books, and the resulting digital material continues to be available as part of the BL's offer.[83]

Another significant example is the Bodleian Library at Oxford, which is currently engaged in two successful PPP initiatives to digitise its collections. The first of these is with ProQuest, to digitise some 65,000 items from the John Johnston archive of printed ephemera. The second is a partnership with Google to digitise the majority of its out-of-copyright works (i.e. those published before 1885). The digitisation will be carried out onsite by Google staff.

We have already clarified the reasons for Google to enter in this kind of partnerships. It is here worth mentioning that, counting Europe only, they have made similar agreements with a remarkable number of libraries: the Austrian National Library,[84] the Bavarian State Library,[85] the Ghent University Library,[86] the Lyon Municipal Library, the National Library of Catalonia,[87] the University Complutense of Madrid and the University Library of Lausanne. The latest of these agreements, made with the Royal Library of the Netherlands, was announced in mid-July 2010.[88]

In addition, in March 2010 Google reached its first publishing partnership with a national government: it made a deal with the Italian Government to digitise up to one million books in the public domain (up to 1870) held in the National Libraries in Rome and Florence.[89]

As for ProQuest, it is a company that creates specialised information resources and technologies, serving all kinds of libraries. It also engages in digitisation initiatives, such as the Early English Books microfilming project.[90] The project dates back to the mid-20th century and aims to create surrogates for all known editions of books in the English language or printed in Britain and its former dominions before 1701. The advent of new technologies meant that it became possible to generate digital images

---

[82] See http://www.telefonica.com/es/sponsorship/html/cultura/literatura.shtml and http://www.bne.es/es/Catalogos/BibliotecaDigital/presentacion/.

[83] This initial partnership had a follow-up with the recent announcement of a joint British Library/Microsoft platform - the Research Information Centre (RIC). The RIC enables teams of researchers to collaborate via the Internet using a common fund of digital research resources.

[84] The most recent addition; the agreement regards some 400,000 works out of copyright (collections from 16th to 19th century) and will become operational in 2011 and last 6 years. Google will sustain the costs of digitisation, while the Library will prepare the books, store the data and manage access to it. The digitised works will be available on the Library's website, on Google Books and on Europeana.

[85] Started in 2007, the initiative has led so far to digitising around 1 million books. Digitisation proceeds at a rate of some 5,000 works per week. The Library will be able to provide free access to the digitised works.

[86] Since 2007, involving 300,000 works in the public domain.

[87] The plan regards some 100,000 works in the public domain.

[88] More than 160,000 books in public domain from the 18th and 19th century will be digitised and made available on the Library's website, on Google Books and on Europeana.

[89] Here as well the conditions are quite interesting: the contract does not provide any exclusivity, so that Italian libraries joining the programme are free to give the same books to Google competitors for further scanning; Italian libraries will receive files usable by any Italian Public administration without limitation, except giving them to Google competitors; all books will be thus published in Italian digital library collections and fully indexed within Europeana.

[90] See http://eebo.chadwyck.com/marketing/about.htm and http://www.proquest.com/en-US/default.shtml.

from the microfilms - leading to the creation of Early English Books Online (EEBO). The project, taken over by ProQuest, is well under way in the UK and continues to expand.[91]

ProQuest actually plans to extend the scope of the project to other European countries, always including books printed until 1700.One of these is Italy, where ProQuest has made an agreement with the National Library in Florence: the company will digitise the collections, and will try to exploit them commercially,[92] paying a royalty to the Library, which will be able to display the works digitised for free to Italian users. Another similar partnership involves ProQuest and the Royal Library of Denmark, for works of the period 1482-1600.

These plans show once more the potential interest of various kinds of players for digitisation initiatives and also further ways of monetising such activities. PPPs are being considered in France, in fact, where plans for digitisation are among the most advanced and potentially well funded. Various experts[93] have recommended to associate private partners to the digitisation effort; one idea is to set up an "economic interest group" gathering public and private subjects, in particular the players of the book sector. The group could manage the exploitation of the works digitised: through operations based on quality and prestige or comprehensiveness for the public domain, upon an adequate legal basis and agreements with rightholders for orphan and some out of print works. Any such initiative could use the services of Arrow to facilitate the identification of rights statuses and the localisation of right holders.

In general, if libraries want to conclude public-private partnership with commercial companies, the question is whether these commercial companies would be ready to digitise and make available books for which the copyright status is somehow uncertain (especially in terms of being orphan). If they indeed digitised books and make them online available, they would face a bigger risk than a library and would then refrain from supporting libraries in making available their collections without a proper diligent search. This is actually why these initiatives almost exclusively deal with works safely in the public domain. The use of Arrow not only would provide more clarity and certainty, it would allow expanding the scope of these initiatives and possibly improve the chances of engaging in some commercial reuse. Arrow will be available as a tool regardless of the kind of partnership enacted and of the subsequent funding model for the digitisation, as well as of the goals of the individual players involved.

4.4 RROs

Reproduction Rights Organisations (RROs) can have an important role in the field of digitisation initiatives such as those we have described so far. They might administer rights in relation to orphan works, collecting fees from users and holding the money in escrow to compensate possible reappearing rightholders. They can also assist authors and publishers in the licensing of out of print works, if these so wish, for digitisation projects.

This, however, would be a consequence for RROs of the operation of the Arrow system, rather than a direct usage by those organisations. A use of Arrow by RROs might nonetheless be envisaged if users addressed them not only for licensing purposes, when rightholders are already identified and located (or when a work has already been identified as an orphan, according to the relevant legal framework),

---

[91] For example, the Edinburgh University Library's large is participating in a project with ProQuest: some 140 books from its collections should be microfilmed at the National Library of Scotland's premises, so that microfilm surrogates can be produced for libraries around the world and digital versions for EEBO. In return, the Library will get microfilm master copies and digital files for their own archive, as well as a payment per item from ProQuest.

[92] The leverage for making the offer of public domain works attractive for a paying public will have to do, among others, with its comprehensiveness.

[93] See reprots Albanel and Tessier, notes 20 and 64.

but also to actually perform the diligent search in their place. Arrow would then represent a cost saving factor for RROs that have to undertake diligent searches. In case of licensing schemes, RROs would use a part of Arrow's revenue stream to cover the cost of administering such schemes and, as said before, hold money in escrow to be paid out to potential rightholders to orphan works.

4.5   Further scenarios

The previous list of possible business models and players engaged or likely to engage in digitisation projects and to need a service like Arrow's was quite complete but not exhaustive. Other possibilities can arise, even some that would be difficult to foresee at present.

One of the main potential activities that could benefit form a tool like Arrow is the digitisation-on-demand (DoD). It basically means that users request the digitisation of a specific work to the institution holding it, which performs the digitisation and delivers the file to the requester, possibly at a fee (usually to cover the digitisation costs).[94]

For example, since 2008 the Österreichische Nationalbibliothek additionally offers digitisation-on demand for almost all of its stocks; users may order various formats and resolutions for study or special purposes. Also the Royal Library of Denmark has a programme of DoD.

The main case, though, is the large-scale project "eBooks on Demand" (EOD): EOD is a service whereby millions of books are available on request in electronic form. The coordinator of the project is the Austrian University of Innsbruck. The project, in cooperation with 18 European libraries from 10 European countries, is co-funded by the EU Culture Programme. Users can order e-books via the common library catalogues; the libraries then digitise the requested item and send it to the user via the EOD service network, at a fee.[95] The books digitised in this way will simultaneously be incorporated into the digital libraries of the participating libraries and thus be accessible on the internet. Works involved are mainly books published from 1500 to 1900; some libraries nevertheless also offer the digitisation of books beyond that timeframe, namely for special user groups, e.g. researchers or people who are visually impaired or blind. For some out of print books EOD asks for a declaration of consent from the author or publisher. Due to the costs of digitisation, the income of EOD libraries covers only part of the direct expenses.

Anyway, at present there are only of a very few cases in which copyrighted works are digitised and made available by libraries. When a copyrighted out of print book is ordered via EOD, libraries try to contact rightholders and get permission for the described usage cases. Libraries involved report difficulties especially in the search for rightholders.

It is therefore clear how the use of a tool like Arrow would enhance a business such as the digitisation-on-demand, highly increasing the chances to successfully meet the demands of users.[96]

Other fields in which a system like Arrow could represent an enhancer are those of Extended collective licences (ECL), Compulsory Collective Management (CCL) and Special Collective Licences (SCL) for

---

[94] Once again, it is worth mentioning that many works can be found in electronic format because rightholders already put them in commerce; what we are illustrating here is the case in which the work is not commercially available in case and the user identified a work in a library collection and requests a digital copy. In that case the library can need a service like Arrow's to locate the rightholder and try to clear rights to meet the request.

[95] For example, the University and Regional Library of Tyrol, one of the partners, charges 50€ for an average 250 pages book.

[96] There is even the possibility that users consult Arrow directly before requesting the digital work to libraries (or to rightholders directly), but it is more unlikely, as a minimum knowledge of handling bibliographic files is needed.

orphan works.[97] In a SCL for orphan works there would be the need to determine the status of the work, normally through a diligent search. When recurring to an ECL or a CCL, there can be reduced requirements for prior search of rightholders potentially combined with limitations in the possible uses of the works involved (e.g. no printing or downloading,[98] display of the works of certain non-mandating rightholders limited to national boundaries, etc.).[99] The tendency to have important limitations can be seen as a consequence of the need to conclude agreements that are acceptable to the largest possible number of rightholders involved.

Also under the ECL and CCL it is relevant to licence the digitisation and/or making available of limited categories of works and rightholders (for example works from or before a defined period, out of print works only, out of print works published before a certain date only, works of a defined category of works or rightholders, inclusion or exclusion of works or rightholders, etc.) as well as mass usages. A tool like Arrow would support and could facilitate such agreements, including providing the necessary title by title search prior to the use in order to establish whether the licence would cover a particular book, thus improving the trust of rightholders and the legal certainty of users.

Arrow would basically help to determine if every book involved actually belongs to the category covered by the agreement, allowing stakeholders to negotiate more sophisticated agreements, as it would be easier to define categories of works of particular interest to libraries and users and for which rightholders would be ready to license rights for broader uses, lower fees, etc. Arrow would reduce the chance of having works withdrawn from a project ex-post and of wrongfully including works in digitisation programmes. A system like Arrow would also improve the management of transnational issues related to ECL or CCL, facilitating the identification of "national" works and enhancing contacts between collective management organisations in different countries. In sum, Arrow-like systems increase the degrees of freedom available to parties in the negotiation of an agreement such as an ECL or a CCL.

---

[97] ECL have been described both in the First Edition of this Report and on the Arrow deliverable D3.5. Report on legal framework - Edition 2, both available on www.arrow-net.eu.

[98] Conditions like these depend on the rightholders, regardless of the system.

[99] This is the case of the Bokhylla project in Norway, for example.

5 Conclusions

In conclusion, we have seen how the situation in terms of the business models that might relate to Arrow is at the moment still uncertain, while nonetheless having a great potential. For the time being not many digitisation initiatives deal with works in copyright, and therefore the need for searching rights information is still limited. On the other hand, there is an impressive political will to make also copyrighted books available online to the public. While for a large number of books in commerce this is already being done by the rightholders (publishers), ways forward are needed for the corpus of orphan and out of print works (when rightholders have no interest in their exploitation).

Libraries are thus ready to digitise hundreds of thousands, if not millions of works, for many of which a diligent search to identify the right status and thus the rightholders will be necessary. In addition, a number of commercial players have an interest in making available to their customers/users a large number of digitised books; whatever the reason - and the underlying business model - these players aim at having the most comprehensive offer they can.

Arrow will provide a tool to facilitate the search for rights information and the identification and location of right holders, thus also increasing the chances of clearing rights for the digitisation and making available of works. It will therefore most likely free resources and give impulse to digitisation initiatives, and in particular contribute to the solution of the orphan works issue.

Arrow will be neutral, both to the legal framework in which it might operate and to the various business models that might connect to it. It will provide a tool that can adapt to different legal settings, as it will deliver objective information upon which decisions can be made, but not make decisions itself. It will as well be open to use by all kinds of players, public and private, with all kinds of business models. This means the Arrow system will be flexible enough to run under a wide range of conditions, both legal and economic.

Furthermore, Arrow is likely to actually increase the number of digitisation programmes, by all kinds of subjects. By facilitating diligent search and - coupled with an adequate legal framework - providing legal certainty, Arrow can expand the scope of digitisation initiatives that various players can undertake. Orphan and out of print works will have a greater chance to be included in digitisation programmes. In addition, Arrow will lower the barriers to entry in this field by commercial players, enhancing various kinds of business models and contrasting the monopolistic tendencies of this particular market.