The Arrow System

Rights Management In Mass Digitization: A European Approach

presented at the 2011 Global Market Forum (featuring Publishing in Italy)
Book Expo America 2011
a conversation with

Piero Attanasio, Head of International Projects, the Italian Publishers Association (AIE)
and the coordinator of the European Projects Arrow and Arrow Plus

and

Michael Healy, Executive Director, The Book Rights Registry

interviewed by

Christopher Kenneally of Copyright Clearance Center

KENNEALLY: My name is Christopher Kenneally. I'm Director of Author Relations
for the non-profit Copyright Clearance Center based in Boston. In a previous
lifetime, I was a journalist, and so this is the part of the program where the dumb
American journalist gets to ask questions of the very distinguished European
gentlemen, all right? And I say that with good humor, because both of these men
are good friends who, I believe, are tremendous experts in their fields. You're
going to learn a great deal from them. I'm going to learn a great deal from them,
and I'm looking forward to asking them some questions that I don't know the
answers to at all.

So first, I want to introduce Piero Attanasio. In 1995, he joined the Italian
Publishers Association, AIE, where he is currently Head of International Projects.
In this capacity, he is the coordinator of the European Projects Arrow and Arrow
Plus, and since 2004, he has served as CEO of MEDRA, which is the Multilingual
European DOI Registration Agency. Piero, welcome.

ATTANASIO: Thank you.

KENNEALLY: Bienvenuto, as I was going to say.

ATTANASIO: Grazie. (laughter)

KENNEALLY: And then also with him, to his right, Michael Healy. He is currently
Executive Director of the Books Right Registry, and previously Executive Director
of the Book Industry Study Group. He's been closely involved in developments of

standards for metadata, product information, and electronic commerce in all aspects of the book publishing industry. And Michael Healy, nice to see you.

HEALY: Thank you, Chris.

KENNEALLY: We'll get started here, and I want to sort of preface the questions by pointing out that more than ever, the publishing business is a technology business. And what are the requirements of a technology business? It is that it must be building something new. Technology is about the new, the innovative, the fresh – things that have not been done before. Technology is also very much data-centered, and I think that's a critical point that will be coming out throughout our discussion today.

And one of the faults, perhaps, of technology, is that it doesn't think about anything but itself. It doesn't consult the law, sometimes. It doesn't really consult the people who may be affected by the innovation.

But I think what you'll find from the discussion today is that the approach that Arrow is taking, and that other groups like it, is a strategic approach, solutions that hopefully will allow for an evolutionary change, one of benefit to publishers, authors, and their readers around the world.

So I wonder if we could start, Piero, because there may be some here familiar with Arrow, and there may be some who wonder just what it's all about. Before you describe what it does, what is the challenge that you're trying to address? What's the problem that you face right now?

ATTANASIO: Thank you, Chris. May just say you – before starting describing Arrow, the – that possibly, also publishing is something that always still with innovation, since publishing in Italy issue 60,000 new titles every eight years. So it's something that is very related to innovation.

But anyway, which was the problem for that stimulated us in creating Arrow. The problem was related to rights management, and the awareness that rights management is something requires management of rights information. Whatever business you have to do, you need some information about your business. And if the business is related to rights, you should have a more precise and sophisticated infrastructure for managing information about rights.

This is the general concept that we were aware of since long time, and probably we had some conversation many years ago with Michael. I remember in a pub in Oxford in particular, speaking about these things when he were still European. And he worked for Nielsen, and we were thinking about a good infrastructure that exists for book data while we have not for rights about book.

And the – then, after this general concept, the use case that allowed us to develop this project is the problem that there is in mass digitization programs carried out by the libraries, or by the individual companies, that requires to looking for rights holders and for rights on massive scale. And in this case, the need for better information is very, very clear.

KENNEALLY: Right, and that is otherwise known as the challenge of orphan works, which we will get into at some stage. So, can I ask you then to describe what Arrow aspires to be? What are some of the basic value propositions that you're hoping to achieve?

ATTANASIO: Speaking about the metaphor, starting from the metaphor of orphan works, I would say that Arrow was considered since the beginning to save the parents, not to identify the orphans. That means to reconnect the works to the parents, that are, the authors and the publishers, and possibly sometimes you don't find this information, and so you have a genuine orphan, and you have to take care about orphans.

But the key point is to find the information that allow you to identify who is the right person where you can ask permission for any use. That's the key approach for us.

KENNEALLY: It's a simple question, but to get there takes a great deal of work, and you've put together quite an assembly of partners to help you with all of this.

ATTANASIO: Yes, yes.

KENNEALLY: Tell us who they are, which particular aspects of the European publishing landscape you've drawn on for this.

ATTANASIO: Yes, the Arrow project involves, I will say, all the stakeholders community in the pub – in the book world in Europe, since we are a consortium that includes libraries, other associations, publisher association like us, and a Collective Management Organization and a reproduction rights organization in particular. So we have the whole chain, between all those publishers, right management organization and libraries that are directly involved.

KENNEALLY: Right. And because, of course, it is a pan-European project, what you are working with is a variety of communities across an entire continent, and trying to make sure that someone in Germany can get the information they need for work that may well have been published in France or elsewhere.

ATTANASIO: Yes. In principle, yes, of course, step-by-step, in the sense that we started with some countries, we piloted – the system is now in place, we'll be piloted in just four countries. Italy is not included within those four countries in Italy. However, in Italy, we developed all the technologies for this. And we started

with the countries where there was more demand for that, because there are digitization plan, including copyright works that is not the case in Italy.

And the four countries are Germany, France, UK and Spain. What happens now is that if you – in any country you ask information about a book that is published in one of those four countries, you can get the information we are able to find.

And in the Arrow Plus project, this was just started one month ago, we plan to include other ten countries in this project. So in Europe, we have 27 different countries with 27 different legislation, 27 different book data infrastructures, and so it's quite challenging, but we start step by step, hoping to –

KENNEALLY: Right. Well, I mentioned at the beginning that innovation is about doing something new, and you correctly pointed out that every time a publisher publishes a book, it's new. And certainly for you, this is a lifetime career. You aren't doing something entirely new, you've been familiar with the establishment of standards and the need for standards for some time now.

In the book business, at some point, it was about the commercial questions that were involved. Can you talk about what are some of the questions and what are some of the advantages of, once this system is in place for the book business?

ATTANASIO: For the book business, you mean beyond this particular use case of the association?

KENNEALLY: Right, but it is not – it does not have commercial intentions at the beginning –

ATTANASIO: Yes.

KENNEALLY: But it has the potential for commercial benefit later. That's –

ATTANASIO: Yes, later, in the sense that coming back to the original concept where we started from, we think that the digital – I think, in particular, that the digital market is much to do with the licensing, because when you sell an e-book, you actually don't sell just the file. You sell certain particular authorization for using that file, that is an explicit, or more often, implicit way, very similar to sell a license.

And so, you need the – in the long term, the same type of infrastructure that we have for the book trade. And so we – for the book trade, we have a wonderful infrastructure in many countries, where you can get from databases all the information you need that's a retailer, for instance, for a book. So you know the price, you know the availability, you know the weight, the size. You know that you have a description and a lot of other things.

For the e-book trade, I think that will be shortly a need for having also some addition to this, not in – and additional information that this requires. You need also some information about rights to complete your dealing.

KENNEALLY: Right, and that gets back to your original point, that rights information gathering is going to shape the kinds of rights management solutions we'll see in the future, and Arrow is right at the center of all of that.

ATTANASIO: I have this – I will not – I'm always very skeptical to speak about too much the future. It's – we are tried to do our job as well as possible for the particular purpose we have developed. We tried to be scalable, but I don't speak – don't ask me to speak too much about the future.

KENNEALLY: Fair enough. But I won't doubt that we'll come back and we'll hear more from you.

ATTANASIO: They just started future selling is the second oldest job in the world. (laughter)

KENNEALLY: Well, the weathermen know all about. (laughter) But one of the points that you have, maybe, because you are thinking about the future, nevertheless. I won't ask you to predict the future, but I think we can be confident that you're thinking about the future. That would be a good thing to do.

And you were telling me, when we were chatting about the program, that what you would like to see is something that you called not a static solution, not a flat solution, but one that allows and provides an incentive to create more sophisticated solutions. Tell us about that principle.

ATTANASIO: Yeah, I think that it's probably better to tell you – the audience some words about how the system works, that is the basis for this. The system works in a way that libraries submit to us a file with multiple records of books, and we go around and distribute the system in a network of databases, where we can find the best information available about rights of that book.

So we – what we find is, first, we identify the book, it's less trivial than you can imagine, in particular, for books that has been published before the emergence of the ISBN. And then you need – we need to identify the work that is included in that book. And then we go find information about the author and the publisher. And when – then we go in other sources for understanding if that book is still in commerce or not. And then we go to the Collective Management Organization looking for information about right ownership.

So we do a very long round querying in segments different databases, and then we provide information. And the long-term value of this is precisely in the concept of

making all those sources interoperable. So that's the basis for scalability of the system.

So that when you have different sources, that allowed – are not able to dialogue between themselves, essentially, the – our job is to take value from creating all of those, and getting the different information from the different – and the – in some way, this good infrastructure also address another problem, that this is maybe a more political problem.

When we – the discussion about the orphan works, and the discussion in particular about the needs for a simplified system over for dealing with rights in digitization programs, in my view, had a wrong basis. When people thought that this job is too difficult, so we have a – we need a legal solution that simplified the landscape, and we just ignore some basic rules of copyright.

And the problem is quite complex, but to a complex problem, you should provide a complex answer, instead of a simplified answer that is static, per se.

KENNEALLY: Right. Well, there's –

ATTANASIO: And then, what is an evolution, is always answering the complexity through complex solutions.

KENNEALLY: Right, a real challenge, I would imagine. And to the point of orphan works, you know, there's – in the shopping malls in America, there's usually someplace where if a child gets lost, it can sit down, and someone will go out and look. They don't immediately be declared orphans and sent off to an orphanage. Someone makes the real effort to locate them, and that this is what you're enabling.

So ultimately – and again, maybe this is a future looking question, and you can decline to answer it, but the result will be a diminishing of the number of orphan works over time. And that's going to be of value to absolutely everybody, whether they're librarians, researchers, publishers, and all.

ATTANASIO: Yes, yes, yes. That's absolutely the objective. When I use the joke with the saving the parents, that means reducing the orphans, (inaudible).

KENNEALLY: Right. We want to see as few orphans as we possibly can, that's right.

ATTANASIO: Yes, yes, absolutely.

KENNEALLY: Well, I'd like to turn to Michael Healy now. Michael, welcome again. And I was reminded of your background as a librarian, and although, as you told me one time, you were a librarian for a very brief period of time, but it has informed everything else you've done since.

The goal of a librarian is to make information available, I think, if I can put it on a bumper sticker. What is your reaction to where we are right now with Arrow, speaking from a librarian's perspective? Why is this important, do you think?

HEALY: As you say, the purpose of a librarian, or the purpose of a library, is to encourage the widest possible access to the works that they collect and distribute. However, libraries, like everybody else, have to operate within a particular legal framework, and in this context, within a particular copyright legal framework.

When libraries became motivated to engage in mass digitization projects, whether the motivation was pure preservation of materials that were going to decline in quality over a period of time, or whether the motivation was simply to make wider and wider access over a network environment, you immediately encounter the constraints of the existing legislative and legal regime. And those two things are not working in harmony anywhere right now.

Now, the interesting thing about Arrow, and I think Piero is characteristically modest about this, is that it's – you can begin to see the influence of Arrow. It's – in terms of its thinking, as governments struggle to keep up with technology, and to keep up with the products and services that technology offers.

If you want to see how influential Arrow has been, I'd encourage you to go and read a UK government report that was issued last week called the *Hargreaves Report*. The British government sponsored a report on the future of intellectual property, because there was a recognition that the legal and legislative framework in the UK was, in some way, frustrating or constraining commercial innovation. And this was bad for the British economy.

So the UK government sponsored a report which became known as the *Hargreaves Report*, and there, right in the center of the *Hargreaves Report* is a statement of the orphan works problem. And along with a statement of that problem, there is a proposed solution, which is that the UK government must introduce legislation to allow the collective licensing of orphan works on a mass basis, but also on an individual rights clearance basis. And using your analogy of the lost child in the shopping mall or wherever, it recommends the creation of something called the digital copyright exchange, not just for textual works, but also for design works and pictorial works and video works and so on.

KENNEALLY: Potentially a one-stop shop.

HEALY: A one-stop shop. And when I talk about the influence of Arrow, what's interesting to me is that the *Hargreaves Report* talks explicitly about – and bear in mind, this is a government document, not a document issued by – it's written by a copyright lawyer. It's not written by a librarian or a publisher. It talks explicitly about establishing a network of interoperable databases that would provide a common platform for licensing transactions.

Now, what's interesting to me is you can actually – if you know about Arrow –

KENNEALLY: You could lift it from one of Piero's presentations.

HEALY: You could have lifted it from one of Piero's presentations. And the other thing is, the digital – I was struck by the digital copyright exchange, which looks awfully similar to me to the Book Rights Registry. It has been lifted out, because that's what's needed. In other words, the orphan works problem, or – it's not really a problem about orphan works. It's about, as Piero was saying, the identification of parents, left untouched, it makes mass digitization programs in libraries much more legally precarious than they need to be. But it also constrains the operation of business.

KENNEALLY: The need for innovation that we talked about at the beginning. And so we've heard about the European approach through Arrow, and you mentioned, Michael, the Hargreaves Report, and that still remains to be resolved in Parliament. We should, because we have an American audience here among the Italians and others, provide people with an updated – I'm sure the world knows by now, but very briefly, there was, as part of the proposed Google Books settlement, there was a solution to the orphan works problem proposed. Judge Chin ruled about a month and a half ago now, sending the settlement back to the plaintiffs and to the defendant to review and to see where they may go.

There were some points that he made, and of course, it's all in private negotiations at this point. Right now in the US, we don't have such a solution at all.

HEALY: We don't. We don't have a legal framework that allows truly orphaned works to be exploited without risk of litigation. Now, you talk about the Google Books settlement, and I was talking about the influences of Arrow and where those influences can be felt. I think wherever you stand on the Google Book settlement, whether you were a supporter or a critic, everybody recognizes that, to use an American sporting metaphor, it's kicked the can down field. In other words, the conversation about orphan works is now very prominent.

And whether you believe the right way is a Congressional solution, or as those who supported the settlement believed, a commercial solution will get you there. Whichever side you come down on, there's no avoiding now the centrality of the orphan works issue.

And yes, you know, the settlement had a solution in view for works that were out of copyright – sorry, out of print, but still in copyright – out of print but still in copyright, where the orphans were – where the parents were difficult to find, which was a sort of opt out solution. And the irony is, if, out of all this controversy, Congress enacts a – or indeed, any other legislative body in Europe or elsewhere,

enacts copyright legislation that allows the exploitation of orphan works, it is also necessarily going to be an opt out solution.

And the settlement itself, when it came to in-print works – in other words, commercially available works, it was always an opt in solution.

So the problems that the settlement was trying to solve and the ways the settlement was trying to solve those problems, I suspect they're going to look very similar to how legislatures, governments are going to solve them anyway. We will come around to solutions which look awfully like the Google Books settlement, but they won't be commercial conclusions between authors, publishers and Google. They'll come out of a government-led solution. But they'll look very similar, I suspect, because they have to.

KENNEALLY: Right. Well, the solution that you and Piero talked about in that pub about five, six years ago – a lot of revolutions have started in pubs. I believe the American Revolution in Boston, for example, got started at – I think it was called the Bell in Hand. But that's where they got the crazy idea of dressing up as Indians and raiding tea ships.

But at that discussion, did you – how quickly did you think the solution would come into place? How quickly did you think the problem would be addressed? Here we are now well into the second decade of the new millennium, the e-book marketplace is in place.

You are both realistic men. You've dealt with these problems before in previous challenges around books and ISBN and so forth. Is the solution coming as quickly as you anticipated, or not? Michael, you shake your head.

HEALY: Well, the most successful evenings in bars and clubs and pubs are the ones you don't remember. (laughter)

KENNEALLY: But you remember this one, so I think that's a comment, right?

HEALY: Well, my memory is perhaps not as precise as Piero's would be. (laughter) But I don't – I think we were both very realistic, because we got to know each other and became friends in an environment of developing standards for the book industry, and that's a painfully slow process, isn't it? It is.

And so we were used to being ahead of the game. And we were approaching it, I am sure, in an environment dominated by a library perspective. And indeed, Arrow still reflects that perspective very strongly.

What we couldn't have been aware of, I suspect, was how – because this was, what, 2004, maybe?

ATTANASIO: Yes, 2005 probably.

HEALY: 2004, 2005. I don't think we could have realized how commercially and politically sensitive copyright would become.

KENNEALLY: And the reality of mass digitization wasn't quite there yet.

HEALY: Absolutely.

KENNEALLY: It was, perhaps, about to come above the horizon, but it had not yet – that day had not yet dawned. And so, what has changed, and what has come so fast – Piero, you're nodding your head – is that mass digitization is here. We are well – it's noontime on that day of mass digitization.

ATTANASIO: Yes, that's also – this is here, it's not so fast, also, in this case. If you compare this speed of digitization program by Google with any other product Google launched, it is much slower.

HEALY: Yes. I also think the definition of mass has changed, because we – when I – I remember back in, early in my career in digital publishing, we were digitizing the full text of the archives of the Soviet Communist party, and we were converting every poem in the English language into digital, searchable text. And those were measured in hundreds of thousands of works.

Bear in mind, since Google started the Google Library program, they're at – what, 15 million books, and they're continuing. So our sense of scale has changed, and we could not have seen just how many separate and different digitization projects would mushroom so quickly.

KENNEALLY: Well, that takes us back to a point that, again, is part of your professional concern, which is standardization. There are so many different projects going on right now, Arrow being very much a leader. But do you worry about other solutions perhaps imposing conditions on what Arrow is trying to do? Is there a potential for that?

ATTANASIO: Well, we – I want approaches to use and to be very connected and to use the results of the other initiative, when they are generally initiative related to standards. Standards –

KENNEALLY: So it's really a global vision, then.

ATTANASIO: Yeah, it's a global vision for sure. Standard is a strange word, because there's a lot of people that develop, prepare to be solution of every kind and code standards, because it's for publicity reason. When we speak about standards, since we had quite a long experience on that, we speak about standards. So something that is recognized as such from the stakeholders community, this is what counts.

And the – our approach is to use everything that has been already developed, and to stay strictly in connection with the standard setting organization, in order to be sure that every single step we do is in collaboration with those organizations, and incorporated in the development standards.

For instance, for dialogue in between all of those databases, we exchange messages. Those messages are – has been created in strict collaboration with Editor, in order to be sure that they are completely Onyx (sp?) for licensing terms compliant. Now, within the Onyx for licensing term standard, and so, it's really all the messages that we exchange between the library and us, between us and the other library databases, between us and the books it print, and between us and the Collective Management Organization, are fully standardized in the proper sense, in the sense that are approved by the relevant standard at the organization.

And to make another example, we developed those routines that directly from the results from (inaudible), when we put together all the books that are in the – say, that contain the same work, we have a routine that automatic way is able to register an ISDC that is the standard for the work, to the relevant registration agents. And the same for other aspects of the project.

KENNEALLY: And for me, and I would imagine, many in the audience, it's fascinating to hear about the layers that are involved here. They're just like sedimentary layers in some stone somewhere. I'm so glad you're both doing it – I couldn't handle it.

I wonder if I could ask Michael one last question about a term you've used when we've chatted in the past, about rights expression and so forth. Can you help us understand what's important about that to the Book Rights Registry, but also to this whole problem at large?

HEALY: Well, as Piero has suggested, nobody should be under any illusion that there's ever going to be a single repository of data about all the works that people are going to be interested in digitizing and licensing. So in other words, the future, as everybody recognizes, is a distributed network of databases, in dozens, if not scores, if not hundreds of countries, ultimately.

And if you're talking about a distributed network, and communication around a distributed network, then that – those databases in the network have to speak to each other seamlessly. And that's where a standards infrastructure is unavoidable. It's irreplaceable. It's – you know, we talk about the orphan works problem. The orphan works problem is a data problem, in big measure.

And it is multi-layered, as you say so rightly, because we in the book industry have been very good at developing product information standards, product identification standards, standards for describing products. We have a very rich foundation layer, which these distributed databases can use to communicate to one another. ISBN,

and Onyx, and all the ones Piero has been referring to, they have been there, available for use – in some cases for 30 years or more, and many more recently than that.

But the least mature stratum is the communication of rights. The terminology, the vocabulary, the taxonomies that make the communication of rights possible in this distributed environment. And that's where rights expression languages come in, and we're very fortunate, I think, as Piero has suggested, that in Editor, we have a body that has on hand the expertise to develop these rights expression languages, and a lot of work is being done, but there's a lot more to do, too.

But as we know, and as Piero and I know very painfully and personally, having the standards is one thing. Adopting them, getting widespread adoption, and then getting affordable implementation is very difficult, which is why it's necessary and important that governments have stepped up to fund programs like Arrow, because without that, perhaps some of these things would never have got off the ground.

KENNEALLY: You're shaking your head yes, yes.

ATTANASIO: Yes, absolutely. Yes.

KENNEALLY: OK. Very briefly, then, before we turn to some questions, your own thoughts, though, about – again, taking you back to your original point, that the info, the rights information that we gather, will determine the way that we will manage those rights in the future.

ATTANASIO: Yes, yes, I think so. I think so, and I think that this can – as I said before, this can influence also some political decisions, and that, if I can comment on this, what Michael was saying about the possible solutions on that may be evolving directly the – voluntary agreements between the players, commercial, non-commercial players that are involved, or legislative solutions, or probably a combination between the two.

These solutions can be much more sophisticated and much more – and grant the much more the right holders in – the actual right holders protection, if there is more sophisticated infrastructure for data management. I mean, that's a – for instance, in the Google settlement, one of the recent (inaudible) from Europe criticized a lot was, it's not only because it was an opt out system, and because it was just an opt out system. If an opt out system is related just to the real (inaudible), so if you start looking for the right holders, and try to elect to speak to the right holders first, and when you don't find the guys, you then establish an opt out, it's completely different than just saying there is an opt out for everyone.

KENNEALLY: Right. And if I may add, I mean, for the Europeans in the audience, Copyright Clearance Center, which is an ROO (sp?) like many of the ones that you

work with in this country, is an opt in system, and many of the regimes in Europe, based on statutory licensing, would be an opt out.

HEALY: Right.

ATTANASIO: Yes, or a mandatory license, either without any opt out right, an opportunity that's – because it's mandatory rights in some cases. But it's very limited in use in that case.

KENNEALLY: Right, right. Well, a very complicated challenge, one ably handled in this discussion here by Piero Attanasio, who is from Italian Publishers Association, where he heads International Projects and is coordinator of the European projects Arrow and Arrow Plus, and by Michael Healy, Executive Director of the Book Rights Registry. Thank you, Michael, and Piero, grazie mille.

ATTANASIO: Thank you very much.

KENNEALLY: It's the Americans' responsibility to keep us on time. We have probably a minute for a question, if there is, from the audience. Your hand came up first. Yes, sir?

M: What's the (inaudible)?

ATTANASIO: Sorry?

KENNEALLY: Very briefly, Arrow and Arrow Plus.

HEALY: What's the difference?

KENNEALLY: What's the difference?

ATTANASIO: Arrow, simply, it's – Arrow Plus is the combination, the continuation of Arrow. It's the – since the name derives from the official name of the project that are funded by the European Commission, we were not allowed to use the same name. But – I am honest, (overlapping conversations; inaudible) (laughter).

KENNEALLY: They had a trademark on that, did they?

ATTANASIO: So for us, the system is the product, and the service is Arrow. Arrow and Arrow Plus are just the name of the projects that are funded by the European Commission.

KENNEALLY: Very briefly, one other question, if there was? All right, well – well, we have a follow up?

M: You mentioned something about one system (inaudible).

ATTANASIO: Yes – no, the system is the same system, and that we are now four countries fully connected, and we plan within two years to have another ten.

KENNEALLY: And those four were Germany, France, Spain and the UK.

ATTANASIO: Yes.

KENNEALLY: OK. All right. Again, thank you all very much for attending, and we'll move on. Thank you. And thank you to Ruediger (sp?) and the Italian Trade Commission, too.

END OF FILE