

# APE 2008 Academic Publishing in Europe, Quality and Publishing

*January 22–23, 2008, Berlin*

Jens Bammel

*International Publishers Association*

Good afternoon ladies and gentlemen. I assume I am the last thing that stands between all of you and lunch which is never a happy place to be. So I will try to be short and brief and talk to you about the Automated Content Access Protocol – what it is and what it isn't, because there are a lot of misunderstandings. Before I start I would like to have a quick raise of hands: how many of you have ever heard before today of ACAP or the Automated Content Access Protocol? Is there anybody who has never heard of it before? Okay. That is extremely useful. So I will have to give some basic information as well as more detailed information. I will be talking today about what it is actually, ACAP. More important is also, what it isn't because there are a lot of misunderstandings. Why we are creating this new kind of protocol and put it into the context of all the various standards, movements that we have from ONIX-PL, which we just heard about to International DLI Foundation etc. I will also talk about the next steps and then finally I will ask the publishers here to do some homework and I will give you some homework what I think what you should be doing after you have come home from this conference. Who you should speak to and what you should tell them. So let me start with what is ACAP? It is actually quite simple. It is an open business model neutral standard tool for the expression of permissions and other terms in machine-readable form. I cannot say these words slow enough because they all have a specific meaning and they also say something about what it isn't. When I speak to other people and ask them: "What have you heard about ACAP", they think: "Well, this is all about newspaper publishers, these old dinosaurs and their last attempt to lock away content and to prevent search engines from actually ruining their outdated 19th century business model. That is, of course, not true. First of all, and let me talk about these misconceptions, it is not just about search engine communication. I will be talking about the first use case. The most publicised use case that we have, which is, in fact, improving the communication between websites and search engines. But that is just a first use case. It is a first starting point. It is the place where we started because that is where we have go interest. That is where we got attention and funding to get this project moving. The second use case which is currently under way actually is about communicating between the British Library and UK websites where the British Library has the right to collect data for legal deposit but the law allows it only a very very restricted use of this harvested content and ACAP will make it much easier for British Library to get much broader and more generous use out of the website owners. So that is the second one that is already being worked on and that will be completed hopefully in the next few months. There are further use cases which I will be talking about as well. It is not just about newspapers nor is it just about publishers. At the moment the consortium which governs and runs the ACAP protocol consists of the International Publishers Association, that is me.

It consists of the European Publishers Council which are basically big European media companies and the World Association of Newspapers. But that is just because we had to start somewhere. The membership includes the World Blind Union; it includes the Motion Pictures Association. There are many organisations that are also participating plus which is about photographers. And, most important, it is not about locking away content. Yes, again, like ONIX-PL we are expressing terms. But at the moment ACAP cannot even claim that what we are expressing are simple licensing terms. They are claims. And the claims can refer to anything. I think that one of the important things, and I feel very passionately about that, is, we should stop focusing either on the one hand on copyright and licensing terms and on the other hand on useful information that people need to act on in their daily lives. I actually believe that these two areas are very very closely linked and, just to give you one example, last week I was speaking with colleagues from the International Federation of Library Associations. And we discovered how much joint interest we have in accurate author information. And also information is on the one hand about identifying who has written what, making sure you have complete records. It is also about copyright and how long copyrights subsist because only if you know when somebody was born and, even more importantly when he died, do you know whether there is something in copyright or not. So legal terms and practical terms are actually often quite closely intermingled. It is also not a technical protection measure. Again: nothing. It is all claims and what you can do with it. This is quite different. So, what are we doing with it? Or, why are we doing with it. Ultimately, we have heard a lot about the semantic web. I heard the word “web 3.0” for the first time today I must admit. I will now use the word “web 4.0”. Just so that I can claim I was the first one to use it. Let’s see if we can hear “web 5.0” this afternoon. But, in the future we have to communicate much better through machines about all kinds of things about content. When we hear that the music world has just disbanded DRM or the DRM dogma at least in some area of its market that does not mean that they do not need information about content. Quite the opposite and that includes information about authorship, about rights ownership, about access terms or access claims. We need these communications whether we are librarians or lawyers and anybody else – researchers. The reason why we are getting involved is because we cannot, if we know that communication in an expression of rights of descriptive terms is so important, we cannot afford that it is owned by one person. We all have paid much too much already to Bill Gates to allow another person to create a proprietary rights expression language and then all of us having to pay that person some kind of fee. We also cannot all afford to create new tools again and again and again as new people come up with new ideas, new little applications which again you find useful. And we have to be aware that we don’t get put into a situation where something is developed in another industry or another area and it simply doesn’t fit to our needs. So it is important that this is developed and it has to be developed at a global scale because standards . . . , there is no such thing in national standards really anymore. And the final reason why we are doing it is actually because the publishing industry knows a lot about how it should be done and at this point I would like to pay tribute to two people – two kinds of people. One is the actual people who are working on standards development in publishing and, like Brian Green here, you will have noticed over those who know him for some time that his hair may have greyed over time and I am sure that it is not just the passing of time it is also really the kind of effort that has been put into developing standards in this world. And the second group I have to pay tribute to are the librarians again. Simply, because standards development has all been about standards up and down the distribution and value chain of publishing and the librarians bring a unique perspective and that is let’s not just fix a short-term problem, let’s actually create something with a vision that will help us for decades, if not for centuries. So this long-term perspective has made it possible that models have emerged in the publishing world which are actually applicable and useful around and for other content industries as well. This

brings me to ONIX and ACAP and ONIX-PL. ACAP is in a way taking a step back and says: we want to develop use cases. We want to develop a language in many different cases, in many use cases. But let's not make these use cases insular. Let's make sure there is a shared dictionary. Let's make sure if somebody is called an agent for example, everybody understands exactly what the word agent means. Or when the word "display" is used or something like that, everybody understands what the word actually means. So using the same kind of words, agreeing what they mean, having a compatible data model is incredibly important and will save all of you a lot of money. And that is what ACAP is trying to do. ONIX-PL is about publishers and libraries but there are so many other relationships which require similar information and where interoperability is a key issue. So from a publishing perspective ACAP enables those people who are now using ONIX-PL and other ONIX based systems to be assured that the investment they are doing there will actually be useful at a later time when other applications come up, when other use cases come up. They do not have to re-invest and re-invent everything. Where is ACAP now? We have talked about the search engines very briefly. In the past 12 months we have worked hardest and most successfully on the use case for search engines. So we have created a protocol which you can put into a specific place on your website called robots.txt which every search engine goes to first to find out more about what it can and cannot do on that website. And we have created a language which allows publishers to express in much more detail exactly what they believe and what they think these search engines should do. We all think of only three search engines. But we should not forget there are hundreds of them out there. And behind each crawler there is somebody who is really interested in your content and wants to use it and display it and show it to somebody else. So there is a real business opportunity there. The current structure is going to be extended for another six months. We are just working in a way on a string to get the use case for the British Library finished, to get the use case for the search engines actually fully finished that includes tools for example for authenticated search engines making sure that you know, when somebody is knocking on your door, that he is actually the one you want to let in. But there are new use cases which we are about to embark on. One will certainly involve syndication. Another one will involve compliance. Yes, the newspaper publishers are very keen on finding all the places where their content is being displayed and finding out whether these people are licensed or not. But non-publishing and public interest projects are also raising their head. I am particularly delighted of the beginning dialogue with creative commons. And it was interesting to note how much we actually have in terms of our whole approach. How much we have in common with creative commons. There are few technical things and one or two political issues, where we disagree. But overall the idea of expressing it and making it simple for people to know what they can and cannot do. And if they cannot do something straight away, to make it easy for them to find out where and how they can get the rights they need, is a very useful tool and everybody wants to work on that. When we get the new stakeholders on board we will have to change our governments. IPA, EPC, World Association of Newspapers, we all agree that we are just one pawn among hundreds of pawns and that we have to get all of them on board and that we will take a role and have to give away the hold on the governments that we currently by sheer necessity have. What about search engine support? This is the one question we always get while we are doing all these things. What happens to the top three search engines. Well, it is a fact; none of the three are formally participating in ACAP. It is also a fact that all are aware of the technical work. We have sent all our stuff to them. We had people speak about how they operate so we know that what we have done is technically feasible. We are also quite aware that in all the major search engines there were clearly signs of heavy internal debate. And individuals do can see that, while it is comfortable for them at the moment to simply be able to do what they please more or less, that, in the future, there is a value in more sophisticated communication tools. And the publishers who are participating in the pilot,

namely Elsevier and Wiley, and some of the newspaper publishers have actually found that it would be a very useful tool and it would be a good thing to be able to control your content while making the more valuable stuff behind your firewalls more visible to the public through search engines. The important thing about search engines is of course, at this point, that they have no real incentive to participate. They would have to make a big investment. They would have to change the way they behave in the way their crawlers behave. And why would they want to do that. There are only two reasons why they would do it. Either because there is a stick or because there is a carrot. Either because they have to or because they will have some kind of advantage. And at the moment that is really what we would like to work on with you. So what is the question? This is all very interesting but why am I telling you this? Why should you be interested in this issue? What does this mean actually for you? And here is where I come to my homework for you. ACAP will become the flexible universal communications tool for all content users if you support it now. Critical mass is the important thing. And what does support now really means? What it means is that where ACAP or ACAP-compatible applications are available, start using them. It does not mean you have to change your policy. To be honest, most publishers I speak to don't even know that they have a website policy or that they could have a website policy. Tell your administrators about it. Actually, implementing ACAP in the so-called robots.txt file takes a medium-skilled website administrator about five minutes. It actually has no impact on the way search engines behave at the moment. But by adding the ACAP translated information that you have already on your website in the robots.txt file, by adding that in an ACAP format, you are actually making the point. And at some point you might want to change your policy. And that takes me to the next point. Actually, talk to your business strategists and ask them about website policy and start thinking about it. As I was saying, there are hundreds of search engines and crawlers coming. Some of them are from search engines, some of quite different people. Why are they coming to your site? Why are they crawling these things? How can you intensify that relationship? How can you make more of it? There is a lot of work that can be done there. And the short tail, or the long tail, sorry, starts after three. Because everybody says you know, we have deals with the big three search engines. But that is quite a long tail if it starts with the fourth crawler that knocks onto your door, your website door. So, the real question that you have to take home and ask yourself is: "Where do you need better content communications?". The publishers, the librarians, the blind, the people in creative comments. Where do we all need to speak better and let's develop use cases around that which actually fit the needs of that particular community. Thank you and "Guten Appetit".