Does copyright have a future?
The time has come to tame the frontier

Mark Bide, Project Director, ACAP Project

(to be published in Business Information Review in December'09)

Biography

Mark Bide is the Project Director of the ACAP (Automated Content Access Protocol) Project. He is also the Executive Director of EDItEUR, the global trade standards organization for the book and journal publishing industries; and a Director of Rightscom, the specialist media consultancy. Mark has worked in and around the publishing industry for nearly 40 years, having been a Director of the European subsidiaries of both CBS Publishing and John Wiley & Sons. He is a Visiting Professor of the University of the Arts London.

Key Words

ACAP; Copyright; Internet; Publishing; Search

Abstract

The internet has been a frontier, and like many frontiers, a fundamentally lawless place. One of the early casualties was copyright. As a society, we have to decide whether we are willing to allow lawlessness on the global network to continue, or whether the internet should be brought within the structures of civil society. One element is the renewal of proper respect for copyright. In creating the ACAP project, the publishing media have recognised that it is essential to work with technology, not against it, developing mechanisms to manage copyright in business relationships at machine to machine level. The time has come to tame the frontier, and allow the media to develop appropriate business models for the 21st Century.

Introduction – the internet as frontier

Imagine life without the internet. It is very hard to think yourself back to the world of business before the internet, email and the World Wide Web. But the internet did not always welcome business. There will be readers of this article who are not old enough to remember the controversies of the early 1990s about whether the internet should ever be used for "commercial purposes" – where anyone who was thought to be advertising was anathematised and "flamed". These arguments must seem vaguely comical to a generation that has known the global network only as a commerce-driven machine, a place where fortunes are made – and probably more often lost. It no longer provides the rather academic, dare we say geeky, environment to which its original inhabitants once aspired.

In the intervening period, the global network has become a vast engine of value creation, by bringing extraordinary communication capabilities within the reach of a global community. "Web 2.0" revolutionised and democratised the capability of individuals to create and publish their own content, with a potential audience of billions – something that would have been almost unimaginable less than a quarter of a century ago.

But few would argue with the contention that the internet has also proved to have its downside. It is hard to imagine that Sir Tim Berners Lee, when he devised the World Wide Web 20 years ago, could have foreseen and far less would have welcomed the darker side of the internet – the widespread availability of content in less positive genres, from child pornography to do-it-yourself guides to the making of improvised explosive devices. We are constantly warned by security experts of how close we are to losing the battle to keep our private information (notably our bank and credit card accounts) out of the hands of fraudsters. Like many another advance in human civilisation, the internet is capable of being used for good and for ill.

The internet prides itself in being "a frontier". Frontiers are by their nature lawless places, places where individualism is prized, where Darwinism is at its most obvious, where the naïve and the careless have little protection, where the rule of law and the tenets of civil society do not run. Frontiers have historically been places of great economic and social dynamism, but this freedom does not last forever. Eventually, the greater good – of the majority at least – is seen to lie in the bringing of a more ordered society.

Perhaps the time has now come when our attention needs to be turned to the question of mechanisms for civilising the internet, and also to understanding better the potential implications for the world if the lawlessness of the internet cannot be tamed. This paper looks at only one aspect of this challenge – although there may be lessons to be learned more widely.

Copyright and the internet

One early casualty on the frontier was copyright. The digital revolution had made the creation of copies of bits of content ridiculously easy – without any degradation of quality. The internet made it possible to disseminate those copies widely, to share them with friends – and indeed with people we do not know. This has become so commonplace that we have even seen the arrival in the European Parliament of a representative of a political party claiming this way of sharing content, of disregarding the rights of creators, as some sort of inalienable human right (http://news.bbc.co.uk/1/hi/technology/8089102.stm).

The introduction of a law of copyright was a response to the arrival of another technological revolution – the widespread deployment of the printing press. The printing press, rather like the internet, facilitated the creation of copies of the work of creators – and, without the protection afforded by copyright, it was not authors but printers who were in a position to exploit the "new technology" – to their considerable advantage. The law of copyright, now one year short of its 300th birthday – the first legislation of this type being the Statute of Anne of 1710 – was introduced to ensure that those who create content can make choices over its subsequent exploitation. This essential principle – subject to limited exceptions – is the cornerstone on which our diverse and highly successful media and knowledge economy have been built. Although there are subtle differences between the more mercantile view of copyright that dominates in the

Anglo-Saxon legal tradition, and the "natural right" of the creator to control his or her creation that lies at the heart of the European *droit d'auteur*, there is common ground over this core principle.

The law has adapted and grown over the three intervening centuries, dealing with many different copying and dissemination technologies that have arisen – the phonograph, broadcasting, photocopying. More recently, the law has been updated around the world to encompass the digital age.

But this time, changes to the law have not proved to be sufficient to alter the way in which copyright is ignored – and often pilloried – on the network. Those of us who believe that copyright exists for the good of society, to "*promote the progress of science and useful arts*" (in the words of Article 1, Section 8 of the US Constitution) must accept the realities of the new environment in which we find ourselves. If copyright is to survive beyond the early part of the 21st Century, it needs to be better adapted to this environment.

This does not imply the need for more law, or indeed for substantially different law. But nor does it imply some pressing reason for there to be *less* law. In the words of the late Charles Clark, the UK's leading copyright thinker of the latter part of the 20th Century, "*the answer to the machine is in the machine*". We are faced with a challenge that is a direct result of a technological revolution, and our response must not be a Luddite rejection of that technology. Rather we need to embrace technology as our core response to the challenge.

In the past, processes for licensing the use and reuse of content have been mediated by people. The internet operates at a scale where human mediation of transactions is no longer viable. We need to seek solutions through which machines can negotiate transactions, where the permissions that are granted can be clearly understood and codified by those machines. In order to achieve this, a first step is the development of an unambiguous "permissions" language that machines can be taught to read, interpret and act on.

This is the objective of the ACAP project.

ACAP – origins

ACAP (Automated Content Access Protocol) can trace its origins to a meeting of publishers and publishing trade associations held in Paris in 2006. By this time, it was already becoming increasingly apparent that (at least in some sectors of publishing) the profitability of publishers' online efforts was being fatally undermined by third parties taking their content and republishing it without explicit permission to do so.

The concern here was not primarily about individual consumers, but about the growth of a whole layer of intermediary businesses, whose entire business model depends on them being able to acquire content free and then monetise it – typically through advertising, but in some cases by essentially reselling that content (in filtered or customised form) to third party clients.

The greatest concern – particularly for the news sector – was the growth of specialist news search services. Although the major search engines offered some sort of *quid pro quo* in terms of the delivery of visitor traffic to the news sites, this was proving to be

entirely inadequate. For the first time, newspapers were experiencing a phenomenon to which they had no response: more and more people were seeing their content, using their content, but the close relationship that had always existed between reach and revenue had been severed. They were persuaded to optimise their websites for search (at considerable cost) to draw more customers in, with the expectation that more customers would bring more revenue. But the customers drawn in from search turned out typically to be low value customers with low "dwell time" on the site and therefore of low value to advertisers. An entire online news sector had been established which on any current economic projection is unsustainable.

But what was remarkable about ACAP was that it wasn't simply the news publishers. From the beginning, it drew together an unprecedented coalition of publishing sectors, seeking a solution to the destruction of value that the casual copying of their content represents.

ACAP – the pilot project

In 2007, ACAP undertook a pilot study to demonstrate whether it would be possible for publishers to communicate permissions for the use of their content more effectively machine to machine. A group of publishers from different sectors came together with a search engine to test the possibility of communicating permissions in this way.

At the urging of the search engines, whose representatives were involved in the project from its inception, this pilot made use of a 15 year old protocol – the Robots Exclusion Protocol – a communication protocol devised at a time when the online world was very different from the way it is today.

The Robots Exclusion Protocol is a very widely adopted *de facto* standard, which was initially proposed in 1993 as a solution to what was then a very pressing problem – the shortage of network bandwidth. It was designed to control the activities of "crawlers" used by search engines to copy the internet – an activity that was seen even then as having considerable downside (in terms of soaking up bandwidth capacity that the website owners preferred to retain for individual human users). While the problem of bandwidth has somewhat faded for most online publishers as a day-to-day concern (the obvious exception are those who have to manage the distribution of large quantities of audiovisual content), the Robots Exclusion Protocol (REP) morphed into something the search engines believed (and apparently in some cases continue to believe) is the complete solution to the management of copyright on the network.

There are several reasons why publishers dispute this reading of REP.

- It isn't a standard, and has no governance process; there is no mechanism for publishers to propose changes to REP.
- The basic specification is implemented differently by different search engines (even among the major players) – and although the three largest have now published full descriptions of their implementations, these serve more to highlight these differences than to resolve them.
- The "big three" search engines have each implemented their own proprietary extensions to REP – and this has led to even greater inconsistency.
- While REP might deal with some Use Cases, it does not allow publishers to express all the permissions that they believe are important to them.
- There are myriad smaller businesses crawling the web – not hundreds but thousands of crawlers; and there are other businesses using other techniques for acquiring

their content without permission. Many of these businesses totally ignore Robots Exclusion Protocol.

Nevertheless, the initial implementation of ACAP was specified as an extension to REP, and showed that the protocol, however primitive, could be used in the way proposed. The search engine which was involved in the pilot – the Paris-based Exalead – demonstrated that changing ACAP instructions on a website could easily be reflected in changes in search results. The technical effort involved was not great – a few man-weeks of engineering effort; but we continue to run into objections from the search engines.

ACAP – the major search engines

At least one of the big three[1] – Microsoft – has expressed real support for the ACAP concept. Thomas Rubin, the Chief Counsel for Intellectual Property Strategy in Microsoft said in a speech in London in November 2008, "*To the extent ACAP can develop into an enabler of content flow…and not become an inhibitor like some failed experiments with digital rights management, it has the potential to be an important element of more vibrant business models for publishers in the future.*"

Nevertheless, Microsoft has not implemented ACAP in its recently launched Bing search service. This is not really surprising. Microsoft is a distant third in search, even in the US, and everyone holds their breath waiting for number one – Google. It would be very hard for Microsoft – or indeed Yahoo! – to implement an approach to the management of copyright on the network if Google continues to argue that it is unnecessary and that it will not implement.

The arguments that Google uses to argue against the implementation of ACAP have changed slightly over time, but in essence:

1. ACAP is unnecessary, because everything that publishers want to achieve can be achieved using Robots Exclusion Protocol – and most particularly that publishers can simply exclude all those search engines which honour REP
2. ACAP is dangerous, because it encourages spamming
3. ACAP only addresses the needs of a small minority of content owners

The first of these arguments holds limited water. Most publishers do not want to exclude their content in this way – search is a valuable tool – but they do want more control over how their content is used in particular contexts, by the thousands of unidentified crawlers which access it as well as by Google.

The second argument relates to some specific features of ACAP that are designed to manage permissions in a trusted environment, particularly where search engines are allowed access to content "behind the firewall". There is a general problem of trust and identity about elements of the ACAP specification which is widely acknowledged – and which goes well beyond ACAP.

[1] For years we have had a "big three" in search – Google, Yahoo! and Microsoft. Microsoft and Yahoo! have always trailed by a long way – even in the US, where some vestige of real competition has remained. The very recent announcement of a collaboration in search between Yahoo! and Microsoft may change these dynamics, but it is much too early to judge what the impact will be.

The final argument is one which perhaps goes to the heart of the argument. For those content owners who publish for reasons that are not directly commercial, control over copyright may appear to be an unnecessary luxury. For those whose business is in content creation and management, the issues are completely different. Professional publishers may make up a small proportion of the content available on the web, but that does not mean that they should not have the rights to control their content if they choose to do so.

There may well be flaws in ACAP – it would be rare indeed for v1.0 of anything to be flawless – but the solution to refining and perfecting the specification lies in a collaborative development effort. Why do the search engines not choose that way forward?

This is a question which we should perhaps allow them to answer for themselves. Anything else can only be speculative; but is it unfair to speculate that there is of course a natural resistance to the idea that copyright owners might now regain control over their content after a decade and a half when such control has not been exercised? I am sure the 18th Century printing industry felt much the same. However things looked then, publishing and printing have successfully coexisted for three centuries in an amicable partnership, and we need to create the context for the same thing to happen between publishing and search and other content supply chain intermediaries.

ACAP – what next?

It is now 18 months since ACAP published v1.0 of the specification; has anything happened in the interim? Well, perhaps the most significant development is that over 1250 publishers from all over the world have now implemented ACAP on their web sites. Very simple, symbolic implementations, true, but important for what they say. They announce very clearly that the publishing industry is no longer a passive observer of the destruction of value in the industry, but an active participant in the resolving the challenges that technology poses.

Since ACAP was launched, the overall position of the news industry globally has worsened quite significantly. No one needs reminding of the damage that the recession has done to newspapers' advertising revenues. And no one should assume for one minute that the industry blames its current state on the loss of copyright control. Life is much more complex than that. The printed news industry has for the most part depended on a mixture of cover price and advertising revenue – and has subsidised its online presence from its print revenues. Print advertising revenue is in retreat, while Google's revenue ($5.52bn in Q2 2009) and profits ($1.87bn) keep increasing. Online "cover prices" are, for the most part, non-existent. It is inevitable that publishers are questioning the cost – and lack of any foreseeable profitability – of their online presence.

Apologists for the status quo argue that publishers need to reinvent themselves, to "monetise the link economy" – but without a monetisation model (except for the vague possibility of supporting all content creation and dissemination through advertising). However, it takes only the most cursory examination of the economics to dispel this as myth. It is illusory to believe that a healthy online content economy could be entirely supported by online advertising; even if every dollar of advertising revenue were to flow direct to the media (a very significant proportion of current online global revenues flow, in one way or another, to Google, which is essentially an advertising funded business with

almost no content of its own), these revenues would be insufficient to support a healthy, plural, news media.

Unless we are to see a complete disappearance of professional news gathering from anyone other than government-supported agencies (such as the BBC), network content is going to have to be paid for. And paid-for content models need the full protection of copyright – not necessarily to make the content inaccessible to search, but to give the copyright owners the right to make decisions about how their content is to be used (so that it does not compete with paid-for models).

The most important work that ACAP has in hand is to continue to press the arguments in favour of the development of a machine-mediated copyright framework to operate on the internet; v1.0 of the ACAP specification provides a start on building that framework, but no one would claim it to be the complete answer.

However, we not only have to continue to make sure that voices in support of copyright are heard just as loudly as the seductive anti-copyright message. We also need to work on new use cases for ACAP and on new ways of expressing ACAP permissions in formats other than Robots Exclusion Protocol – ACAP, after all, is primarily there to facilitate future business models. We expect more progress on both these fronts during the second half on 2009.

Copyright – the future

No one who contemplates the future of publishing and the impact of the internet could possibly believe that the media do not need to change. That analysis includes everyone in the media.

New ways of creating value must be found, new business models will be essential to survival in the medium term. But it will prove hard to create value in a context where everyone seems to believe that it is legitimate to help themselves to other people's content, and to destroy the value of that content in the process. Why would anyone invest time or effort in creating content when there is no hope of getting a proper return on that investment?

The creative industries have been an engine for growth in the economy, something which we would be ill advised to ignore at a time when economic growth is at such a premium. The media are not asking for special treatment, they are simply asking that their rights should be properly protected.

"Information wants to be free." Well, perhaps, but couldn't we same thing about beer, to just as loud applause? And in fact this often quoted sentence is part of a much longer (and wiser) point made as long ago as 1984 by Stewart Brand:

> *On the one hand information wants to be expensive, because it's so valuable. The right information in the right place just changes your life. On the other hand, information wants to be free, because the cost of getting it out is getting lower and lower all the time. So you have these two fighting against each other."*

This is a battle that continues. In the absence of the mechanisms necessary to make copyright function properly on the Internet, it is a battle that will inevitably be won by

those who believe that information ought to be free – without copyright, there is no reward mechanism, no business model.

It has been suggested – in all seriousness – that "ACAP would destroy the internet". No doubt the same thing has been said about every attempt to tame a frontier. The internet needs to be tamed, needs to be brought within the constraints of civil society. ACAP is a modest step in that direction but a vital one if we are to maintain a vibrant and creative media.