



ACAP: A briefing paper

Improving universal access to content published on the World Wide Web:

- *It's hard for anyone to make content available for access and use on the internet without any rules....*
- *It's hard to follow rules if you don't know what they are...*
- *It's hard to learn how to read and understand rules if you are a machine....*

1 The ACAP Pilot Project – a summary

ACAP (Automated Content Access Protocol) is being developed as an industry standard to enable the providers of all types of content published on the World Wide Web to communicate information relating to permission for access and use of that content in a form that can be readily recognised and interpreted by a search engine so that the search engine operator is enabled systematically to comply with the publisher's policies. ACAP will provide a technical framework that will allow publishers worldwide to express access and use policies in a language that the search engine's robot "spiders" can be taught to understand.

As a result, it will be possible for publishers to make more content available to users through the search engines, and to continue to innovate in the development of business models for network publishing.

In the past decade, online models of publishing and search have become increasingly sophisticated. ACAP will allow these models to be properly reflected in the permissions that are granted.

The current protocol (robots.txt) allows only a simple choice between allowing and disallowing access. There is no standard mechanism for expressing conditional access.

ACAP will officially launch its pilot phase in November 2006.

2 Genesis – a solution to the “search engine” dilemma?

At the beginning of 2006, the major Europe-based publishing trade associations – including the World Association of Newspapers (WAN); the European Publishers Council (EPC); the European Newspaper Publishers Association (ENPA); the International Publishers Association (IPA); the European Federation of Magazine Publishers FAEP); the Federation of European Publishers (FEP); the World Editors Forum (WEF); the International Federation of the Periodical Press (FIPP) and Agence France Presse – established a Working Party to consider the issues that are posed by search engines for publishers, and to look at ways in which mutually beneficial relationships can be established between publishers and search engine operators, in which the interests of both parties can be properly balanced.

All sectors of publishing have a “search engine dilemma”. The search engines are an unavoidable and valued port of call for anyone seeking an audience on the internet. Search engines sit between internet users and the content they are seeking out and have found brilliantly simple and effective ways to make money from the audience that they attract. As a result, they have become so dominant that even the largest website owners are not large enough to have any serious impact on their commercial fortunes.

The benefits of powerful search technology to both users and providers of content are well recognised by publishers. At the same time, publishers are aware that search engines are gradually but inevitably moving into a publisher-like role, initially merely pointing, then caching and, finally, aggregating and, in effect, “publishing” content. Indeed they may even create content themselves while re-using publishers’ content without reference to the actual creators or owners.

In the current state of technology, there can be none of the differentiation of terms of access and use which characterises content-trading relationships in traditional publishing environments, whether electronic or physical. The search engines can and do reasonably argue that, since their systems are completely automated they cannot possibly enter into and manage individual and different agreements with every website they encounter. There is no practical alternative to their current *modus operandi*.

The search engines are able to make their own rules and decide for themselves whose interests are worth considering. This means that publishers are reluctant to make available for search much content which might otherwise be available to users.

If publishers are to take the initiative in establishing orderly business relationships with the search engine operators, the response must be to help those operators to address the problem, both to fill the technical gap and to ensure its political implementation. To paraphrase the former copyright adviser to the UK Publishers Association Charles Clark’s famous claim that “the answer to the machine is in the machine”, the challenges that are created by technology are best resolved by technology. Since search engine operators rely on robot “spiders” to manage their automated processes, publishers’ web sites need to start speaking a language which the operators can teach their spiders to understand. What is required is a standardised way of describing the permissions which apply to a website or webpage so that it can be decoded by a clever if uncritical machine without the help of an expensive lawyer.

In this way, one of the search engines’ most reliable rationalisations of their “our way or no way” approach will have been removed, and a structure

which embraces and supports the diverse present and future internet strategies and business models of online publishers will have been created.

As a result of the efforts of the Working Party, a proposal was made to develop a permissions-based framework for online content. This would be a technical specification which would allow the publisher of a website or any piece of content to attach extra data which would specify what use by search engines was allowable for that piece of content or website. The aim will be for this to become a widely implemented standard, ultimately embedded into website and content creation software.

Following the commissioning of a brief feasibility study, WAN and EPC have taken the initiative to establish a project to develop and pilot this framework to express publishers' access and use policies. This project, publicly announced at Frankfurt Book Fair in October, will be formally launched in November 2006.

3 ACAP – the vision

ACAP will develop and pilot a system by which the owners of content published on the World Wide Web can provide permissions information (relating to access and use of their content) in a form in which it can be recognised and where necessary interpreted by a search engine "spider", so that the search engine operator (and perhaps, ultimately, any other user) is enabled¹ systematically to comply with such a policy or licence. Permissions may be in the form of

- **policy statements** which require no formal agreement on the part of a user
- **formal licences** agreed between the content owner and the search engine operator.

There are two distinct levels of permissions which need to be managed within this framework:

- The **permission** given to the search engine operators for their own operations (access, copy and download, cache, index, make available for display)²
- The **delegation of rights** given to the search engine operators to grant permissions of access and use to search engine users (search, access, view, copy, download, etc)³

Although these can be managed within the same framework, it is important that the differences between them are recognised.

4 Business requirements

Although it will be an integral part of the ACAP project to further develop and confirm the business requirements for the operation of the framework, significant progress has already been made in identifying the high level

¹ Note that **there is no question of enforcing compliance on the search engine**. There is confidence that creating a **capability** of compliance is sufficient in business to business relationships.

² There is a further distinction to be made here between modifying the behaviour of the spider (in terms of what it crawls) and modifying the behaviour of other elements of search engine operation (in terms, for example, of what is cached).

³ In which case, of course, we are talking of the search engine at least in part requiring compliance with the permissions granted by the content owner (simply in terms of what is presented to the user and under what conditions). In terms of compliance, there is a considerable difference between business-to-business and business-to-consumer relationships

business requirements against which any technical solution must be measured. In summary, the solution must be:

- **enabling not obstructive:** facilitating normal business relationships, not interfering with them, while providing content owners with appropriate mechanisms for managing permissions for access and use of their content
- **flexible and extensible:** the technical approach should not impose limitations on individual business relationships which might be agreed between content owners and search engine operators; and it should be compatible with different search technologies, so that it does not become rapidly obsolete
- **able to manage permissions associated with arbitrary levels of granularity of content:** from a single digital object to a complete website, to many websites managed by the same content owner
- **universally applicable:** the technical approach should initially be suitable for implementation by all text-based content industries, and so far as possible should be extensible to (or at the very least interoperable with) solutions adopted in other media
- **able to manage both generic and specific:** able to express default terms which a content owner might choose to apply to any search engine operator and equally able to express the terms of a specific licence between an individual search engine operator and an individual content owner
- **as fully automated as possible:** requiring human intervention only where this is essential in order to make decisions which cannot be made by machines
- **efficient:** inexpensive to implement, by enabling seamless integration with electronic production processes and simple maintenance tools
- **open standards based:** a pro-competitive development open to all, with the lowest possible barriers to entry for both content owners and search engine operators
- **based on existing technologies and existing infrastructure:** wherever suitable solutions exist, we should adopt and (where necessary) extend them – not reinvent the wheel
- **consumer friendly:** normal day to day access by consumers to website content will be unaffected by ACAP

The approach taken should also be capable of staged implementation – it should be possible for initial applications to be relatively simple, while providing the basis for seamless extension into more sophisticated permissions management.

Although the scope of the project is initially limited to the relationship between publishers and search engine operators, a framework which meets these requirements should be readily extensible to other business relationships (although details of implementation would not be the same in every case).

5 The Pilot Project

The ACAP pilot project is expected to last for 12 months. In outline, the project will:

- confirm and prioritise the business and technical requirements with the widest possible constituency: agreement with stakeholders is essential if the project is to succeed in the long term
- agree which specific Use Cases should be implemented in the pilot phase of the project, starting with a relatively simple approach
- develop the elements of the technical solution: it is anticipated that this will primarily involve the development of standards for policy expression, although it will also be necessary to develop the tools for the implementation of those standards
- identify a suitable group of organisations willing and able to participate in the pilot project; it is currently anticipated that this could involve half-a-dozen publishers, selected to represent different sectors and both sides of the Atlantic, and at least one of the major search engines; participants will need to be in a position to dedicate technical and time resources to the project to enable it to succeed
- pilot the standards and the tools, to prove the underlying concepts

In parallel with the development of the technical solution, a significant stream of project work will involve the development of a *sustainable governance structure* to manage and extend the standards (and any related technical services) which will be needed after the project phase of ACAP is complete.

To avoid duplication of effort, ACAP will also establish liaisons with relevant standards developments elsewhere. In particular, the project is already in contact with EDItEUR⁴ with respect to its development of *ONIX for Licensing Terms*; and, in view of the significance of identification issues, with the International DOI Foundation.⁵

6 Next steps

It is anticipated that the project will be launched publicly in November 2006; at launch it will be possible to be more explicit about plans and expectations.

The feasibility study commissioned by WAN, EPC and ENPA concluded that this project is technically feasible – and indeed requires little in the way of genuinely new technology. Rather, it requires the integration and implementation of identification and metadata technologies that are already well understood. It is also possible to chart a developmental path which does not demand that every element of the framework must be in place before any of it can be usefully implemented.

However, this is not to suggest that everything will be simple, nor that it can be achieved without cost. A significant part of the project cost will have to be borne by those organisations that agree to participate in the pilot, in the development of their own systems; however, there will also be central costs, to which it is hoped that other publishers will be prepared to contribute.

⁴ www.editeur.org

⁵ www.doi.org

7 Frequently Asked Questions

What is ACAP?

ACAP (Automated Content Access Protocol) is being developed as an industry standard by the publishing industry, working with search engines and other technical and commercial partners.

ACAP will enable the providers of all types of content published on the World Wide Web to communicate permissions information (relating to access and use of that content) in a form that can be automatically recognised and interpreted, so that business partners can systematically comply with the publishers' policies. In the first instance, ACAP will provide a framework that will allow any publisher, large or small, to express access and use policies in a language that search engines' robot "spiders" can be taught to understand. It is anticipated that, in future, the scope of ACAP will be extended to other business relationships and other media types.

What will the ACAP Pilot Project do?

The ACAP Pilot Project is designed to develop the v1.0 of the Automated Content Access Protocol.

A group of publishers drawn from different sectors of publishing, in Europe and the United States, will work together with technical partners (including a major search engine) to develop and demonstrate initial implementations of ACAP.

Which partners are involved in the pilot project?

The identities of the organisations which will participate in the pilot will be revealed when the project is launched.

Discussions are actively in progress with a significant number of interested organisations. The pilot project is currently being steered by a group representing the World Association of Newspapers (WAN), the European Publishers Council (EPC) and the International Publishers Association (IPA).

Where is the driving force behind ACAP?

The primary drivers of ACAP are the World Association of Newspapers (WAN), the European Publishers Council (EPC) and the International Publishers Association (IPA).

Together with the European Newspaper Association (ENPA) these three organisations have steered and financed the initial stages of the project. The **European Publishers Council** is a high level group of Chairmen and CEOs of European media corporations actively involved in multimedia markets spanning newspaper, magazine and online database publishers. Many EPC members also have significant interests in commercial television and radio. The **World Association of Newspapers** groups 72 national newspaper associations, individual newspaper executives in 100 nations, 13 news agencies, and nine regional press organisations, representing more than 18,000 publications in all international discussions on media issues, to defend both press freedom and the professional and business interests of the press. The **International Publishers Association** is a Non Governmental Organisation with consultative relations with the United Nations. Its constituency is book and journal publishers worldwide, assembled into 78 publishers associations at national, regional and specialized level. The **European Newspaper Publishers' Association** is a non-profit association currently representing 5,100 national, regional and local newspapers. These daily, weekly and Sunday titles are published in 24

European countries where ENPA's members are operating in their national markets.

What is the timetable for the ACAP Pilot Project?

It is anticipated that the project will be launched in November 2006 and will run for 12 months.

Part of the defined task of the pilot project will be to develop a transition plan to move from pilot to sustainable implementation.

How important is ACAP to the publishing industry?

For the first time ever, the newspaper, magazine and book publishing industry are working together on a joint standard.

This collaboration not only maximizes and concentrates available resources, skills and knowledge; it also increases the likelihood of success. The project will also strengthen the publisher-search engine relationship. It will enable new uses of works, and increase their visibility, without any compromise to publishers' commercial freedoms.

Isn't this simply an attempt by publishers to "lock up" their content?

No – precisely the opposite is true.

When it is implemented, ACAP will provide publishers with the confidence to make content available to search engines much more widely than is currently the case.

Why not use existing protocols for managing this?

The current protocol (robots.txt) allows only a simple choice between allowing and disallowing access. There is no standard mechanism for expressing conditional access.

Robots.txt will continue to have an important role to play, and ACAP is not intended to supersede it as a protocol for everyone. However, it has proved inadequate for the task which is now required. This is unsurprising, since the protocol is, in internet terms, a very old one and has never been updated; it was designed for a much simpler age. In the past decade, online models of publishing and search have become increasingly sophisticated. ACAP will allow these models to be properly reflected in the permissions that are granted.

Is ACAP restricted to text media?

No: it is fully intended that ACAP will be designed to be extensible to all types of content published online, including audio and video.

In the short term, the pilot project will focus on the requirements of the text-based media; like any project of this type, it is essential to maintain a manageable scope. However, initial conversations have already taken place with experts from other media sectors, and provision has been made for more formal liaisons to be established during the pilot project. This will allow for the future requirements for extensibility and interoperability to be taken into account in the design of ACAP.

Isn't this all about money?

No: but no one would deny that it is partly about money.

Publishers are not ashamed about making money out of publishing – that is their business. They make substantial investments in the creation and distribution of content, and believe that they should be able to make a fair return on those investments. Business models are changing, and publishers

need a protocol to express permissions of access and use that is flexible and extensible as new business models arise. ACAP will be entirely agnostic with respect to business models, but will ensure that revenues can be distributed appropriately.

What's in all this for search engine users?

More content will be searchable: ACAP will give content owners the confidence to allow search engines to index their content under clear terms of use.

Publishers are in the business of providing content to users; on the internet, search engines have proved to be an invaluable intermediary in the process. Other innovative models of intermediation are also appearing. The development of standards for expressing online access and use permissions will enable the development of an orderly information value chain which will provide users with access to the largest possible aggregations of authoritative content.

Isn't ACAP focused entirely on publishers' recent disagreements with Google?

No: ACAP is a long-term strategic project, not a search for a short-term tactical remedy.

The various court cases that have arisen between publishers and Google are a symptom of the problem that ACAP seeks to solve, not the problem itself. Just like Google, publishers would like to find solutions outside the courtroom. This is ACAP's objective. The relationship between publishers and aggregators should not be adversarial but co-operative – for the benefit of their common customers.

Aren't Google offering commercial deals to some publishers?

Yes: but Google is not the only search engine with which publishers have relationships – and search engines have to have relationships with a very large number of publishers.

Business relationships on the internet should not simply be about deals done between very large corporations. It will not be possible to manage the very large number of business relationships in the absence of much greater automation. ACAP aims to enable the majority of smaller publishers, smaller search engines and other innovative intermediaries to enter the growing market for online content with confidence.

Have you discussed this with the search engines?

Informal discussions have already taken place, and it is anticipated that at least one search engine will participate in the pilot project.

The identities of the organisations which will participate in the pilot will be revealed when the project is launched.

How will ACAP work?

The full technical specification of ACAP has yet to be defined – this is first objective of the pilot project.

A feasibility study was commissioned earlier in 2006, which revealed that there was little or nothing which has to be invented or built "from scratch" in order to implement ACAP. However, different approaches might be taken and the specific details will be worked out during the pilot project itself.

8 Contacts

ACAP will be an open standard and we welcome comments and suggestions. If you want to comment on what you have read, or you have any questions about this project please contact:

Project Co-ordinator:

Mark Bide
Rightscom Limited
Lincoln House
75 Westminster Bridge Road
London SE1 7HS, UK

Tel: +44 (0)20 7620 4433
info@the-acap.org

Media enquiries:

Heidi Lambert
Heidi Lambert Communications Ltd

Tel: +44 (0)1245 476 265
heidilambert@hlcltd.demon.co.uk

For more information and to stay up to date with progress, visit the project web site:

www.the-acap.org