

"Thanks for a thoughtful reading of the ACAP proposals. If I may attempt to paraphrase your conclusion, it is that our proposals contain some interesting stuff, but in your view they are somewhat quirky (my interpretation of your word "strange") and are not in a form that you would consider ready for implementation. As the person primarily responsible for drafting these proposals, I'm obviously keen to understand how we can make them more ready for implementation, and your comments raise some important issues.

It's unfortunate that the fact that we are calling these proposals "ACAP Version 1.0" is in some quarters being interpreted as some kind of a "take it or leave it" ultimatum. Nothing could be further from the truth. We recognise that these are first public proposals for how publishers could more clearly communicate policies for access and use of what they choose to publish online, and no doubt there are many ways in which these proposals could be improved.

Regarding the overall "strangeness" of these proposals, I believe that this derives in large measure from the fact that the conversation that we have tried to have with search engines during their preparation has been lop-sided. Because the major search engines have not been able, for whatever reason, to engage formally in the ACAP project, much of the input to our work has been from publishers, and this is undoubtedly reflected in many ways - large and small - within the proposals. We would have liked more input from potential implementors on the receiving side, and this has obviously had an impact upon the style and substance of our proposals.

Our original intention was not to use REP at all. Some in the publishing community took - and perhaps still take - the view that REP is an inappropriate technology for communicating policies in terms that make sense to content owners. REP is essentially permissive by default, whereas the basis on which publishers generally license use of content reflects the law, which is that permitted uses are specified by the publisher. Our eventual decision to propose extensions to REP was based upon the blindingly obvious fact that REP is the established way for content owners to communicate routinely with crawler operators, and it will be far easier for crawler operators to implement extensions to what they are already able to interpret in REP than to propose an entirely new protocol.

Since we decided to propose to use the REP methodology to deliver access and usage permissions, most of the discussion has been with publishers about what they want to be able to communicate and, arguably, too little – but by no means none - of the discussion has been with implementors about how best to express this in technical terms. That is not ideal, and has resulted in a set of proposals that quite understandably make more sense to those involved in their development than to those who weren't involved. ACAP has always been an open project to implementors on the receiving side as well as publishers, and our invitation for them to participate remains open. We would greatly welcome their increased input.

Regarding the detailed points you make about the content of the specification, some of these are down to decisions that I personally took as to what should be in the public version of our proposals and what to leave out. I had to try to strike a balance between at one extreme issuing a set of proposals that only contain stuff that we are entirely confident about and at the other extreme giving a full picture of all the features that we'd like in there but which haven't been fully specified and tested yet. I obviously got the balance wrong from your point of view, so sorry about that. But here are some detailed responses:

Dates, times and durations - Yes, I'm sure that we have to be more precise about how these should be expressed and interpreted. I think just adopting standard ISO formats for date-times isn't a total solution. More input from prospective recipients would help us to ensure that what is expressed clearly and precisely also has a clear and precise interpretation in terms of the intended behaviour.

Interpretations by prior arrangement - I think this is a matter of opinion. I believe that we have to define how to express things that publishers wish to say but which search engines, for understandable reasons, won't in the general case be willing to act upon, but may be

prepared to in specific cases. There are several examples in our proposals of forms of expression that search engines, unless they make a special arrangement with the content owner, would be bound to treat as "cloaking". Maybe implementors on the receiving side would like us to divide our proposals more clearly into a core set of features that don't involve such issues and a supplementary set that might. From a publisher perspective a number of what you might see as "non-core" features are quite fundamental to what they need to be able to communicate, so at this stage I don't think it would be helpful to create such a separation.

"Work in progress" indications in the specification - I think this all boils down to a confusion as to whether or not our proposals are supposed to be a finished specification. They're not, and I freely admit that the "Version 1.0" label, which indicated that this is the output of our pilot phase, has probably not helped us to make that fact clear, especially among web developers who are used to "Version 1.0" being a much more complete and polished thing. I regret that confusion.

Lack of examples - Yes, you're right, we need to include more examples.

Used resource types - You're right that these combine a number of things, but in the end what the publisher wants to say is whether or not you can "present" (you may believe this a bad choice, but oh! what interminable debate we've already had about the right term to use!) something, and these are an attempt to list the things that get presented in search results: snippets, thumbnails, links, current and past copies that have been stored ("preserved") in the search engine cache, and sometimes the original page retrieved in real-time from the content owner's site. By the way, we've dropped "extract" from the current list in the corrected set of proposals that are now available from the ACAP website (the second "extract" should have been "oldextract" - a typo - but that's gone too now). If you exclude "extract", do you still find the list confusing? Yes, we fully understand that these things get produced/retrieved in a variety of different ways for presentation to the end-user, and in anything other than REP I'd propose a very different way of specifying them. We've tried to keep the syntax simple. You can of course generalize and just use "present" without the specialization for a particular used resource type, in which case the permission/prohibition applies to all forms in which a resource might be presented.

Max-length, max-size - One of the items on our immediate "to do" list is to review what needs to be added to support use cases involving images and other non-text resources. We have so far deliberately focussed on permissions relating primarily to text resources.

Version 1.0 - Yes, as I've already agreed, this means different things in different worlds, and I regret any confusion that this has caused about the status of the proposals that we have now made public.

Other - This clearly depends upon your point of view. If you're a publisher and you have a contractual obligation to your authors and other contributors not to allow their content to be used for purposes other than those that are agreed to be permitted, you need to be able to communicate a blanket prohibition. That's the only sensible application of this usage type that we've identified. You don't need to express a blanket permission in REP, because REP is fundamentally permissive by default, so anything you don't explicitly prohibit can be assumed to be permitted. On that basis, if you say nothing, that will convey exactly the same as "ACAP-allow-other", so it would be a redundant expression. By the way, I fundamentally disagree with you about the advisability of letting large corporations do what they like with your data, but that's another matter."

Fc response to jim grimmelman

<http://arstechnica.com/articles/paedia/skeptical-look-at-acap.ars>

"Internet publishers have devised a complicated new system called the Automated Content Access Protocol (ACAP) that aims to supplement the aging Robots Exclusion Protocol by providing a means for site operators to provide more detailed instructions to web spiders and indexing programs. Although there are contexts in which a richer and more expressive mechanism for establishing indexing parameters would be advantageous for site operators, ACAP goes far beyond indexing and attempts to dictate the terms under which content can be used and displayed. ACAP, which is being devised by a cabal of publishing companies with very little regard for implementation challenges, would likely represent a significant burden for major search engines-assuming that any decide to support the protocol."

ACAP reply:

Thank you for writing about ACAP. I would like to address some of the criticisms you make about us, and also tackle a couple of misconceptions.

When you write that ACAP has been devised by "Internet publishers" this rather belies the strength of industry support behind this project - the initiative of the European Publishers Council, the World Association of Newspapers and the International Publishers Association representing worldwide interests of newspaper, magazine, book and journal publishers. In addition, ACAP's members represent more than 18,000 publications and 4,100 companies and organisations with a turnover of over \$260 billion.

This "complicated" new system is easy for anyone publishing online, big or small, to implement. In its simplest form it between 15 minutes to half an hour to implement. There is a page available on the ACAP website which includes step-by-step guidance on basic implementation (<http://www.the-acap.org/implement-acap.php>). Of course as the protocol becomes more sophisticated then the ways in which it can be implemented and the business models it supports will become more complex but this will a sign of success and demand by the market rather than an indication that ACAP has complexity built in.

In terms of the "burden" on major search engines, the search engine, Exalead, that implemented and tested ACAP during our pilot managed the work from scratch with just one person during a four-week period and did not report any significant impact on their servers overall.

We are in close contact with the major search engines who have all collaborated with ACAP on an informal basis. Indeed their input underpinned many key decisions including the initial implementation of ACAP as a development of robots.txt which is easier for them to implement. We are confident that they will come on board in the

future to enhance their relationships with content providers and to gain access to yet more high value content on the net.

ACAP has been devised as an enabling communications tool, to stimulate more availability of content online. In its present form it updates the Robots.txt protocol which, after all, was created back in 1993 in the early days of the Internet - and then indeed to help manage scarce bandwidth. It was then subsequently used as a permissions protocol but is now outdated and fails to meet the needs and sophistication of today's digital environment. Anyone putting content on the net will benefit from ACAP - from individuals to large corporations. It is open, global, and freely available to anyone with an interest in having a say in how their content is used. ACAP will give confidence to all content providers to put their content, especially high-value content, on the Internet, thus benefiting every user of the Internet.