



# September 2007



# ACAP:

towards greater  
compliance on the  
internet

Four left-pointing chevrons (less-than signs) in a gradient of colors from light cyan to dark blue, arranged in a horizontal line.

WIPO: 17 September 2007

Emerging fields of application for RMI  
Search engines and users

*Mark Bide  
Rightscom  
ACAP Project Coordinator*

*September 2007*



## agenda

- where did ACAP come from?
- who is involved in ACAP?
- what is the ACAP pilot project?
- where is ACAP technically?
- why does ACAP matter?
- where is ACAP going next?



# developing the Automated Content Access Protocol

where did ACAP come from?



**TERMS AND CONDITIONS FOR THE WEBSITES**

Last updated on 30 November, 2004

Please read carefully the following Terms and Conditions. They apply to the Websites (as defined below), which are owned and operated by members of the News International Group and, by accessing any of the Websites, you are agreeing to abide and be bound by such Terms and Conditions.

No charge is made for your use of the Websites (unless otherwise stated), although you should be aware that telephone call charges, at rates determined by your telephone operator, may apply (including WAP over GPRS or other telephony charges). Click on the links below to access sections of this document:

**Definitions**

In these Terms and Conditions the following terms shall have the meanings set out below: "Electronic Device" means a computer, mobile phone, WAP phone, personal digital assistant, or other electronic device capable of accessing the Websites; "Microsite" means any page on a Website; "News International Group" means News International Ltd (whose registered office is at 1 Virginia Street, London, E9 1XY), the holding company of News International Ltd and any subsidiary in time to time, News International Ltd or its wholly owned subsidiary also referred to as "we", "us" and "our". Holding company and subsidiary have the meanings given in sections 736 and 736A of the Companies Act 1985. If you have any queries about whether a particular company is a part of the News International Group, please contact us at any of the addresses listed in clause 19. "Newspapers" means The Times, The Sunday Times, The Sun and The News of the World, the Internet editions of which appear on the Website for each Newspaper.

"Trade Marks" means any of the registered or unregistered trade marks owned by the News International Group, which includes "The Times", "The Sunday Times", "The Sun", "The News of the World", "Page Three", "Page 13", "Page 13.com", and any associated word or device marks and combinations of the same, and any other trade marks as maybe added to this list from time to time. "Websites" means the websites and Web sites including their subpages with their home pages as set out below (and "Website" means any one of them):

**2. Acceptable Use Policy**

-You agree to abide by all applicable laws, regulations and codes of conduct and ensure that any content uploaded or distributed or stored by you does not infringe the rights of others.

-All material on the Websites and any material sent to you by e-mail or any other form from the Websites by "content" of any kind relating to the Websites belongs to our licensors or us. You may remove and display content from the Websites on the Electronic Device on which you first accessed it or downloaded it, print a single copy of individual pages, snippets and screenshots for caching purposes only for your personal and non-commercial use alone.

-We, or our licensors, own the copyright and all other intellectual property rights associated with the content, save where otherwise stated.

-You may not do any of the following without prior written permission from us:

- reproduce the content (other than allowed under this Acceptable Use Policy), or modify or in any way comment on, all or part of the content;
- redistribute any of the content (including using it as part of any library, archive or similar service);
- obtain the copyright or other intellectual property rights in any copies of content made in accordance with these Terms and Conditions;

-We reserve the right to make any changes to the content of the Websites by automatically downloading and storing all and any of the content. Requests to republish, redistribute or syndicate content should be addressed to: enquiries@nsvindication.com

-You acknowledge that we own the Trade Marks and that you may not use a trademark without our prior written permission. Other product and company names and logos mentioned or displayed in the Websites may be the trade marks, service marks or trading names of third parties. To check whether any product or service is a registered trade mark of ours please contact: trademarks@newsint.co.uk.

-In accessing the Websites, or any one of them, you agree not to:

- impersonate another person or use a false name or a name you are unauthorised to use or create a false identity or e-mail address or try to mislead others as to the identity or origin of any communications;
- modify, access or make available data stored on a computer device which you have accessed through our network (unless otherwise permitted by the Website or these Terms and Conditions);
- make available or upload files that contain software or other material, data or information not owned or licensed to you or collect information about others (eg names/addresses) without their prior consent;
- damage, interfere with or disrupt access to the Websites or do anything which may interrupt or impair their functionality;
- make any commercial or business use of the Websites or resell or commercially benefit from any part or aspect of the Websites;
- publish, post, distribute, disseminate or otherwise transmit defamatory, offensive, infringing, obscene, indecent or other unlawful or objectionable material or information;
- threaten, harass, stalk, abuse, disrupt or otherwise violate the rights (including rights of privacy and publicity) of others;
- make available, upload or distribute by any means any material or files that contain any viruses, bugs, corrupt data, "trojan horses", "worms" or any other harmful software;
- falsify the true ownership of software or other material or information contained in files made available via the Websites;

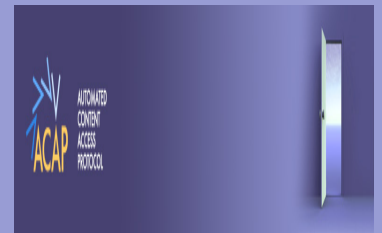
# the basic conundrum...

✓ it's hard for anyone to make content available for access and use on the network without any rules regarding access and use...

✓ but it's hard to comply with those rules if you don't know what they are...

✓ it's particularly hard to learn how to read and understand rules if you are a machine...

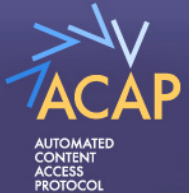
❖ **the solution? Make the rules accessible and easy to understand automatically without human intervention**





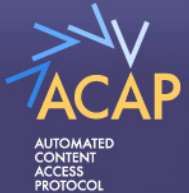
# ACAP began with the search engines

- the value of search engines to users – and to publishers – is incontrovertible...
- a multitude of positive business relationships exist between search engines and publishers
  - index billions of web pages
  - point end-users to our content
  - provide tools to sell advertising
- the power of search technology and search companies has grown exponentially
  - **they** collect, cache and index content indefinitely
  - **they** decide what to display; how, when and where
  - **they** don't seek positive consent for their activities



## the goal...to redress the balance

- task force created January 2006
- publishers want (and need) to engage with search engines and other intermediaries on the internet
  - but in business-to-business relationships like any other
- search engines are the pioneers of the next generation of the internet
  - the future lies in collaboration not in opposition
- automated crawling and indexing requires automated information on rights and permissions
  - all parties need fully automated transactions without the need for human intervention



# building on what has gone before

- the current level of sophistication for expression of content access and use permissions is too low
- robots.txt as currently specified does not provide an adequate solution for publishers as it cannot express conditional policies
- robots.txt is inconsistently applied by different crawlers

A screenshot of a Notepad window titled 'robots - Notepad'. The window has a standard menu bar with 'File', 'Edit', 'Search', and 'Help'. The text area contains the following lines of code:

```
User-agent: *  
Disallow: /_private/  
Disallow: /css/  
Disallow: /javascript/
```



# ACAP – benefits for everyone in the chain

## Consumers

**more** access, **more** content, in **more** places, with **less** content inaccessible behind firewalls

## Publishers

**new** opportunities to deliver **more** content to consumers online in a way that satisfies their legitimate commercial interests

## Intermediaries

**more** quality branded content, made accessible under mutually agreed and clearly expressed terms of access and usage

## Combined industry

**NO** standards wars;  
**NO** proprietary monopoly;  
**Open** collaboration;  
**Win:Win** for all stakeholders



# developing the Automated Content Access Protocol

who is involved in ACAP?



# ACAP participants

## ■ leadership/funding

- World Association of Newspapers
- European Publishers Council
- International Publishers Association

## ■ publishers

- Agence France-Presse
- De Persgroep
- Holtzbrinck Macmillan
- Impresa
- Independent News & Media
- John Wiley & Sons
- Media 24
- Reed Elsevier
- Sanoma

## ■ technical partners

- Exalead
- British Library



# ACAP Members

- Access Copyright
- Associated Press
- Association of American Publishers
- Associazione Italiana Editori
- Australian Publishers Association
- Authors Licensing & Collecting Society
- Copyright Agency Limited
- Copyright Clearance Center
- Copyright Licensing Agency
- Dapper
- Deutsche Presse-Agentur
- EDItEUR
- European Alliance of News Agencies
- Express Newspapers
- European Newspaper Publishers Association
- Fairfax Business Media
- Federation of European Publishers
- Forlæggerforeningen
- Gazette Communications
- International Association of STM Publishers
- International DOI Foundation
- International Federation of Reproduction Rights Organisation
- International Press Telecommunications Council
- Mediargus
- Motion Picture Association
- Nederlands Uitgeversverbond
- News International
- News Limited Australia
- Newspaper Association of America
- Newspaper Licensing Agency
- Office of Public Sector Information (OPSI)
- Ovid Technologies
- PLUS Coalition
- Publishers Licensing Society
- Random House Group
- Recording Industry Association of America
- Reuters
- Scholastic
- Vlaamse Dagbladpers
- World Blind Union



# developing the Automated Content Access Protocol

what is the ACAP pilot project?



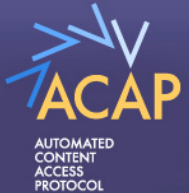
## objectives of the pilot project

- 12 month project – through 2007
- standardized framework for machine readable expression of permissions for access and use
  - elaborated for specific Use Cases
  - with a process for extending
  - proof of concept through pilot implementations by participants
- sustainable business plan for future management
  - transition plan
- communication and public affairs programme to influence opinion



# developing the Automated Content Access Protocol

where is ACAP technically?



# technical progress summary

- core set of Use Cases being piloted
  - representation from different parts of publishing
  - print based resources – both open and protected
- permission semantics [EDItEUR]
- framework for expressing ACAP semantics in robots exclusion protocol syntax
- XML expressions of ACAP semantics in development
- discussion papers on
  - crawler authentication
  - improving/automating take down procedures



# developing the Automated Content Access Protocol

why does ACAP matter?



## why ACAP?

- current relationships between publishers and [some] network partners are far from optimal
  - lawsuits all over the globe are a symptom, not the problem
- “the answer to the machine lies in the machine”
  - a technical solution to a technical challenge
- ACAP is a collaborative project to find such a solution
  - ...with support from all sectors of publishing
  - ...and our technology partners
  - ...and increasing interest from other media
  - ...and increasing political support
- no equivalent work is taking place elsewhere



## why ACAP?

- ACAP is committed to avoid reinventing wheels
  - being developed in close collaboration with EDItEUR
  - as a minimum, guaranteed interoperability with ONIX-LT
- our objective: a practical, pragmatic open standard
  - with a proof of concept implementation
  - the project is on schedule and on budget
- a standard
  - not a business model
  - not a set of services



# developing the Automated Content Access Protocol

ACAP – where next?



## next steps

- complete technical pilot  
– during October 2007
- complete work on design of future governance structure for taking ACAP forward  
– during November 2007
- present to final conference in New York on November 29 2007



[www.the-acap.com](http://www.the-acap.com)

>>>> thank you & <<<

[info@the-acap.org](mailto:info@the-acap.org)  
[mark.bide@rightscom.com](mailto:mark.bide@rightscom.com)

**Please contact me if you would like more information**

*Mark Bide  
Rightscom  
ACAP Project Coordinator*

*September 2007*