



# Publishers and Search Engines



## FACING CHALLENGES & OPPORTUNITIES TOGETHER



### ‘Content Online’



**Angela Mills Wade**  
Executive Director  
European Publishers Council



### 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 them you are agreeing to 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:

#### 1.

##### 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; "Micro Site" means any page on a Website; "News International Group" means News International Ltd (whose registered office is at 1 Virginia Street, London, E3 8 1X) the holding company of News International Ltd and any subsidiary from time to time of News International Ltd or its holding company, 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 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", "Page3", "Page3.com", "Sun", 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 wap sites (including their constituent pages) with their home pages as set out below (and "Website" means any one of them):

#### 2.

##### 2. Acceptable Use Policy

#### 3.

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.

#### 4.

All material on the Websites and any material sent to you by e-mail or any other form from the Websites (the "content") or in any way relating to the Websites belongs to our licensors or us. You may retrieve 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 on paper and store such pages for caching purposes only, all for your personal and non-commercial use alone.

#### 5.

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 our prior written permission from us:

#### 7.

reproduce the content (other than allowed under this Acceptable Use Policy), or copy or in any way commercially exploit any of the content;

#### 8.

redistribute any of the content (including using it as part of any library, archive or similar service);

#### 9.

remove the copyright or trade mark notice(s) from any copies of content made in accordance with these Terms and Conditions;

#### 10.

create a database in electronic or structured manual form by systematically downloading and storing all and any of the content. Requests to republish, redistribute or syndicate content should be addressed to: enquiries@nisyndication.com.

#### 11.

You acknowledge that we own the Trade Marks and that you may not use any of them 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

● It's hard for anyone to make content available for access and use on the network 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...



**The solution is to make the rules easy to understand**



**The solution is ACAP**





# Why we need (and want) search engines...

- The **value** of search engines to users – and to publishers – is **incontrovertible**... but not universal
- **Search engines -**
  - Index billions of web pages
  - Point end-users to our content
  - Provide tools to sell advertising
  - Generally make life easier for end-users
- A multitude of **positive** business relationships exist between search engines and publishers



# The power of search engines

- The power of search technology and search companies has grown exponentially
- Search engines are **no longer** simple road signs on the information super-highway
- Most search engines have become publishers, portals, ad platforms and “digital libraries **operating for profit**”, relying on material created and paid-for by others:-
  - **They** collect, cache and index our content indefinitely
  - **They** decide what to display; how, when and where
  - **They** don't seek positive consent for their activities



# The power of search engines

- As Publishers, we don't always know what they are doing with our content or what they are planning to do with it
- They do not feel accountable to anybody for this
- For many commercial uses, they do not accept licensing agreements
- Some "*don't do evil*", but increasingly they act as both judge and jury on what constitutes evil



# Is this best for everyone?

- Publishers do **not** want to lock away our content
  - as large audiences help to drive our businesses
  - and some (larger) publishers are attracted to some of their ideas and business proposals
- However, smaller content owners currently have only one choice:
  - Allow no searches (and become invisible);  
*or*
  - Allow searches – but only on **THEIR** terms



# Should we allow them to do this?

However:

- This is not about **good** or **evil**...
- This is about the **management** and **promotion** of **quality content** within the framework of our current and future digital strategies

**So...**

**Who should decide how our content should be accessed and used?**



# Industry task force created

- In January 2006, the global associations of newspaper, magazine and book industries were brought together by W.A.N. to:



1. examine the options open to publishers to assert their intellectual property rights recognition;
2. improve the relationships between content creators/producers and news aggregators/ search engines;
3. ultimately improve universal access to content;

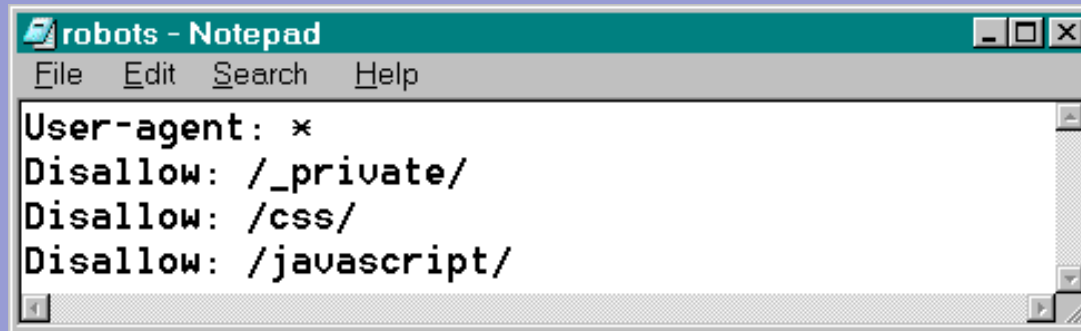


# Our goal... create a level playing field

- Search engines are the **pioneers** of the next generation of the internet: *fully automated transactions without human interference*
- Publishers want (and need) to engage with search engines and other intermediaries on the internet
- Automated searching and crawling requires **automated** information on rights and permissions

# Our goal... create a level playing field

- However, the **current** level of sophistication for expression of content access and use permissions is too low
  - Robots.txt **does not provide** an adequate solution for publishers



```
robots - Notepad
File Edit Search Help
User-agent: *
Disallow: /_private/
Disallow: /css/
Disallow: /javascript/
```



# Why **Robots.txt** doesn't work...

- Our web sites are visited by hundreds of web crawlers a day – and that number increases daily
- Many crawlers are new; many are unidentified -- but probably provide bona fide services (others simply **ignore** Robots.txt)
- Robots.txt **only** allows full blocking. Robots.txt **cannot** express policies
- Robots.txt is **inconsistently** applied by different crawlers



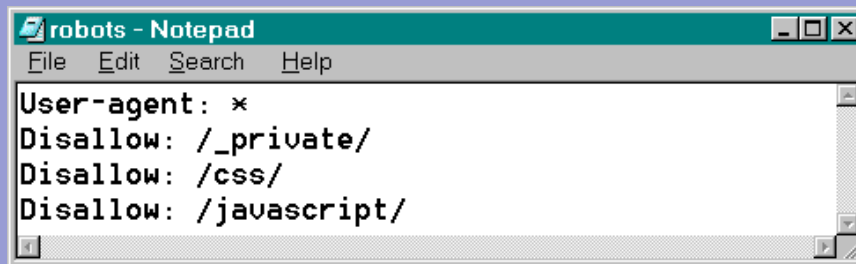
# Why **Robots.txt** doesn't work...

- Robots.txt was designed to be extended – but has **not** been significantly updated
- While Robots.txt may work for a few publishers, it does **not** work for the **majority** of publishers and their digital publishing strategies



# The Future

- The global industry needs a more expressive, automated and enabling solution for digital publishing in the 21<sup>st</sup> Century
- It needs to **build on** Robots.txt and to better reflect the realities of modern day digital publishing
- **From** this...



```
robots - Notepad
File Edit Search Help
User-agent: *
Disallow: /_private/
Disallow: /css/
Disallow: /javascript/
```



... **to** this





# Access and Use Expression Language

- We need an access and use expression language that search engines crawlers can pick up and that machines can interpret and act upon
  - **The book publishing industry has a great deal of experience in developing standards (ONIX)**

**Q:** Can these be expanded/ adapted for newspapers, magazines, periodicals, book searches, search engines and other e-commerce?



# Access and Use Expression Language

## A: We believe they can...

- So, we commissioned a detailed “Feasibility study” by Rightscom, which has resulted in the development of ACAP...



That feasibility study confirmed...

- **YES** -- all the building blocks are there
- **Industry reaction:** enormous interest and support from all different industry sectors



# Step One: a pilot project

- W.A.N., EPC and IPA have formed a consortium to develop a Pilot application for ACAP
- The ACAP Pilot will be looking at **2 scenarios**:
  - Machine-readable information on access and use for web crawlers
  - Licensing information on digitised books
- Pilot participants:
  - Newspaper publishers
  - Book publishers
  - Magazine publishers
- Pilot Project launches in November 2006

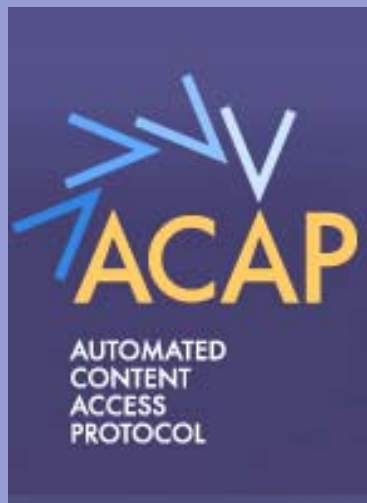


# Core principles of ACAP

- **Enabling**, not obstructive
- Based on **existing** technologies and **existing** infrastructure
- **Flexible**
- **Scalable**
- **Extensible**
- As **fully-automated** as possible
- **Open standards**-based
- **Inclusive**



## AUTOMATED CONTENT ACCESS PROTOCOL



The *right solution* at the *right time*,  
providing a more **expressive**,  
**automated** and **enabling technical**  
**solution** for the demands of digital  
content publishing for the 21st Century.



# Myths and Misunderstandings

- **“The Web already allows for this”**

***NO.** The current protocol (Robots.txt) only allows a choice between “allowing” and “disallowing” access. There is no standard mechanism for expressing conditional access.*

- **“Publishers just want to lock away content”**

***NO.** Licensing terms always exist, even with freeware/ shareware/ creative commons. We want everybody, including machines, to be able to find and understand them. ACAP is not a DRM or enforcement technology and does not technically prevent access to content.*

- **“This will prevent fair use”**

***NO.** The ability to express existing licensing terms in a machine-readable form is good for all parties of a licensing agreement. Use and access expression language presents no barriers to any legitimate use of content – and ACAP is not DRM!*



# The Basic Conundrum... answered

... with ACAP

- ✓ It will **now** be **easy** for anyone to make content available for access and use on the network with proper rules...
- ✓ It will **now** be **easy** to follow rules, now that you know what they are...
- ✓ It will **now** be **easy** for machines to automatically learn how to read, interpret and understand rules...



# ACAP -- Benefits for all

A shared access and use expression language will benefit all, for:-

## CONSUMERS

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

## PUBLISHERS

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

## LIBRARIES

**more** access, **extra** information, **new** tools for authentication and verification of digital content. This will not interfere with archival storage, but enhance it.



# ACAP -- Benefits for all

A shared access and use expression language will benefit all, for:-

**SEARCH  
ENGINES**

**more** quality branded searchable 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



# 21<sup>st</sup> Century Solution for 21<sup>st</sup> Century Publishers

“ Opening the Door for  
Digital Publishing for the  
21<sup>st</sup> Century ”



AUTOMATED CONTENT ACCESS PROTOCOL

FOR MORE INFORMATION:

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

E: [info@the-acap.org](mailto:info@the-acap.org)