

# Who's Making Money Off Your Content?

**How news organizations can level the playing field by tracking their stories and acting on reuse.**



# Introduction

Although people are consuming news at unprecedented levels—including through new means like websites, RSS readers, search engine aggregators, podcasts, and SMS headlines on mobile devices—news organizations currently face great financial pressure. The content they produce has the potential to reach ever greater audiences at ever lower costs, but the challenge of capturing the full value of stories produced in-house at great expense has never been more difficult.

Why is it so difficult? In part, because consumers now often navigate to individual stories rather than to a news organization's home page, through referrals from blogs and aggregators. And by replicating headlines, images, and excerpts, intermediaries like aggregators and blogs often take over the function of a news organization's home page—including garnering valuable advertising revenue. With these new distribution methods, licensing strategies have become more complex and harder to manage. Finally, there's the most painful kind of unauthorized reuse, where creators of spam blogs or "splogs" set up sites that automatically replicate news content without permission for financial gain.

Yet knowing where news content is being consumed, reproduced, remixed or rewritten, and reread is key to creating an effective content business in the digital age. In other words, protecting intellectual property begins with knowing where and how it is being used. Ken Doctor, industry analyst and president of media consulting company Content Bridges,<sup>1</sup> explains, "The world is moving towards a time where news is free—free to move in and around the lives of consumers. That means readers and consumers will be able to get at it whenever they want, wherever they want it, with whatever device. We now see that the way to monetize this is by associating certain kinds of advertising with specific content or users. Consequently, the most important thing for news publishers is to capture some benefit, largely through advertising, as that content moves freely around the world."

News organizations need ways to more effectively track and monetize all the content—articles, images, and video—they produce. Attributor's technology platform provides web-wide content visibility and automated content licensing tools, empowering publishers big and small to claim their content, understand how it is being used, and profit from its distribution. As part of its strategic initiative to enable new licensing and distribution models and to safeguard its intellectual property rights, the Associated Press is already implementing Attributor's technology.

---

<sup>1</sup> <http://www.contentbridges.com>

# New Realities in News Publishing

Nearly every aspect of the news industry today—from production and editorial, ad sales and circulation, to promotion and licensing—is in transition.

## News organizations remain trusted sources of information...

News sites remain among the most popular online destinations, and news content continues to drive high advertising revenues—largely because news organizations offer timely, credible content that is of interest to many consumers.

They are sources of information that are professionally produced and edited—using processes refined over decades to reduce error, ensure balance, assume little to no prior knowledge on the part of readers, and deliver on schedule. Sorting, sifting, and assessing different competing claims and sources, evaluating the significance of events and stories, editing for quality and clarity, and matching with imagery are deep competencies that many new digital content providers are still working on.

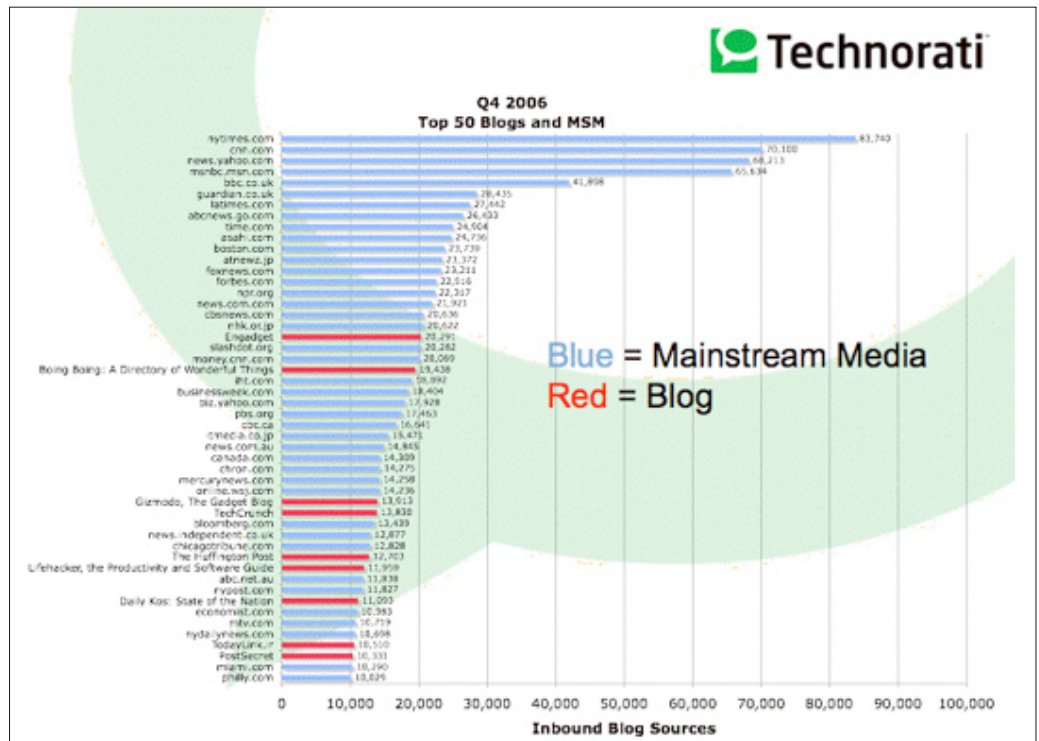


Figure 1. Popularity of Blogs vs. the Mainstream Media.

Blogs may be overtaking some news sites in popularity, but not all of them create original materials, and many if not most rely on newspaper content. Blogs thus represent an opportunity for redistribution, low-cost marketing, story ideas, and reader feedback, as well as a new form of competition, and perhaps less savory forms of reuse.

Source: Technorati's State of the Live Web, April 2007. <http://www.sifry.com/alerts/archives/000493.html>

"... the most important thing for news publishers is to capture some benefit, largely through advertising, as [news] content moves freely through the world."

— Ken Doctor, *Content Bridges*

### But readership and consumer habits are changing...

More and more people depend on computers, mobile phones, PDAs, and other digital devices for many parts of their daily life—including how they get news. In the U.S., one in three people regularly went online to get news in 2006,<sup>2</sup> and the balance between traditional print and online news consumption continues to shift towards online access. People are accessing news content via Yahoo News and Google News, MSNBC.com, individual sites such as NYTimes.com and WSJ.com, or the millions of blogs that cover and aggregate news about specific topics.

### ...Leading to new concerns around access, advertising, and distribution...

While viewing news stories online is convenient for consumers, the rapid shift in online behavior has surfaced strategic concerns around access to content. Online consumption of news has also led to changes in advertiser expectations—and means that news organizations must cope with changing ratios in revenues from subscriptions, print or on-air advertising, and online ads. Attempting to offset shrinking print ad revenues by selling more online advertising is no longer working to close the gap, as online ad revenue growth for newspapers recently slowed.<sup>3</sup> In fact, in 2006, newspapers lost 14.3 percent of their ad dollars from major advertisers, while those advertisers spent 17.8 percent more on Internet advertising.<sup>4</sup>

Search engines and aggregators have had an enormous impact: They have effectively become the gatekeepers for online news. Like other online content, online news is now fueled by distributed advertising like the AdSense and ValueClick networks. Meanwhile, social networks like MySpace (MySpace News), Digg, MyBlogLog, and Topix are becoming the new promotional platform for news stories.

The result? These trends towards disintermediation make it harder for publishers to capture online ad revenue. They also make it more difficult for papers to get a complete picture of where their content goes and how sticky it really is—and without those insights, publishers can't create effective strategies to increase online revenue.

### Because rank has its privileges

Newspaper publishers need to track where on the web their content is going, where it's generating the most revenue, and for whom. But their options for doing so are arduous or non-existent. Meanwhile, disintermediated distribution means that search engine optimization (SEO) techniques alone no longer suffice to capture viewers.

To be viewed and produce revenue, news organizations' stories and websites need high rankings in web search results, as most viewers never look past the first page of search results, according to a study by iProspect that found that 62 percent of users do not look past the first page of search results.<sup>5</sup> Rank on Google News and other news aggregators is just as important. And for news publishers, up-to-date knowledge about where their site ranks on the Google News page—versus where a licensee, or someone who's reposting without permission, ranks with the same story—is crucial to getting more traffic.

<sup>2</sup>"Online Papers Modestly Boost Newspaper Readership," Pew Research Center for the People and the Press, July 30, 2006, <http://people-press.org/reports/display.php?ReportID=282>

<sup>3</sup>"Newspaper Online Ad Growth Slows," AdWeek, May 30, 2007, [http://www.adweek.com/aw/search/article\\_display.jsp?vnu\\_content\\_id=1003591653](http://www.adweek.com/aw/search/article_display.jsp?vnu_content_id=1003591653)

<sup>4</sup>"Advertising Industry Share Shifts," Wachovia Equity Research, July 9, 2007.

<sup>5</sup>"iProspect Search Engine User Behavior Study," April 2006. [http://www.iprospect.com/about/whitepaper\\_seuserbehavior\\_apr06.htm](http://www.iprospect.com/about/whitepaper_seuserbehavior_apr06.htm)



**Figure 2: Example of a Google News Ranking Report**

This custom chart shows the number of webpages reusing a piece of original content that, on Google News, outrank the site that originally published the content.

Many sites will repost large chunks of content without any licensing agreement or even attribution.

### Difficulties Tracking Content Online Make the Web a Plagiarist’s Paradise

Many sites will repost large chunks of content without any licensing agreement or even attribution. The ability to instantly monetize webpages has spawned a new generation of plagiarists, with millions of “splogs” (spam blogs) using RSS feeds to automatically scrape content for financial gain. But manually finding sites that repost content without permission is difficult, manually tracking and managing them is not realistic, and the options to resolve these situations are limited.

In summary, to thrive in the digital age, news organizations require more visibility into what happens to their content across the web, and how it ranks. Without that information, publishers cannot pursue the most promising strategies for content licensing and syndication. “News publishers are having to make the transition from thinking about paid circulation to what is largely a free environment on the web, fueled by ad monetization. This means the news and information they create must be tracked in a different way,” says Ken Doctor. “What that will take is a simple, universal tagging system that shows not only the source of the content—its publisher or owner—but also what kind of content it is, by geography and topic. The system that wins will be one that is simple and universal, and that allows content to freely move and get monetized easily.”

# Requirements for an Effective Content-Monitoring Solution

Which stories (and reporters) are garnering the most attention and inbound links?

Given the many changes the news industry is going through, it makes sense to take a step back and think through a comprehensive approach to monetizing content, improving search engine rankings, monitoring reuse, and protecting intellectual property and valuable brands. At the same time, it should ensure that managers and publishers get the strategic insights they need about viewers, advertisers, and licensees.

To get more insight into what happens to their content, and to be able to exert better control over how their content is used online, news organizations need solutions that simplify the monitoring and tracking of their content.

To be effective, a content-monitoring solution must be fast, robust, and scalable. Crawling the entire web at speed to find both authorized and unauthorized uses must also be done at a granular level. Searching with a keyword-based tool alone is insufficient and incomplete: tracking must be done continually, completely, and must go beyond keywords to find all instances of reuse of text, video, and images.

The licensing portion of the system must offer a more graduated response to unauthorized reuse than simply a) doing nothing, or b) writing a DMCA take-down notice. After all, the point is not to restrict readership—instead, it is to monetize an expanded audience.

In addition, the licensing part of the system should answer questions such as:

Which licensees are using my content most actively?

- Who is going beyond the license scope?
- Which licensees are not using my content at all?
- How can I package licenses more attractively?
- Where should the licensing, legal, and content-producing departments follow up?

The system must also be easy to use, and provide the information that senior management needs to make strategic decisions related to:

- What content is most worthwhile to make available for free, to expand reach and readership.
- Which stories (and reporters) are garnering the most attention and inbound links?
- What the best candidates are for new ad sales campaigns and reprint offers?
- Whether any emerging patterns of behavior online need to be addressed.

And for day-to-day monitoring, the system should provide fast insight into which headline and which type of related content—text, photo or video—spreads furthest. It should also rapidly show what the impact is of adding an image or other rich media to a story across licensees and the rest of the web.

# Under the Hood—Attributor’s Technology and Features

Attributor is designed to offer publishers greater visibility into content reuse and new tools for content monetization.

## The technology: Fast and cost-effective

To help publishers monitor and manage their content across the web, Attributor developed cost-effective “content fingerprinting” technology and proprietary algorithms that can detect the presence of matches down to a very granular level. Publishers benefit by receiving Attributor’s comprehensive, accurate, and timely reports on content duplication—all in context, so they can focus appropriate resources on the matches that matter.

Here’s how it works:

- 1. Claiming and fingerprinting.** When a publisher of any size claims a story or piece of content—whether text or image or, soon, a video file—this tells Attributor’s system to fingerprint it.
- 2. Continuous monitoring.** All content claimed through Attributor gets analyzed by Attributor’s system, so that during its constant crawling of the web, it knows what text to look for. To do so, Attributor developed algorithms that work in a very targeted way—the system can detect matches of even a few sentences, and in near real time. With the prevalence of revenue-stealing splogs, this is a real benefit, as it helps detect sites that copy only part of a story providing details on search engine placement—making it easier and faster to take action.
- 3. Reporting matches.** The system automatically reports every match, qualifying those matches by what percentage of the original text was copied. The results can be shown and sorted in the context of their (re)use—ad-supported or not, by type of licensee, etc.
- 4. Taking action.** Attributor makes it easy to follow up with users of content. For instance, it will track and flag reposters of content for follow-up by the licensing sales department, and simplify requests for link-backs and attribution. If needed, Attributor empowers publishers to send DCMA takedown notices and remove duplicate content from search engines.

Today, the technology is based around finding as many matches as possible of sentences and phrases used in content claimed through Attributor. Instead of searching for keywords and removing most of the results that look similar, as search engines do, Attributor was specifically designed to look for *any* chunks of claimed text that were copied and reused, whether with or without permission.

More importantly, Attributor was created to work with all types of digital content—later this year, Attributor will extend its services to include fingerprinting and monitoring of images, as well as video content.

### Features and benefits: Simple to use, comprehensive reports, and actionable results

With its easy-to-use interface and its extensive and near real-time reporting capabilities, Attributor offers unprecedented visibility into content use around the web. The system continuously monitors billions of webpages for relevant matches, eliminating the need to run time-consuming recurring manual searches, or create in-house tools to track and manage licensed content. To make it a turn-key solution, Attributor hosts and provides the technology as a web service. This means no costly installation or software upgrades on-site, and 24/7 secure access through the web for all members of a news organization who need it.

Through its dashboard and reports, Attributor shows the percentage copied of content, existence of attribution and linking, commercial use, and presence in search engines.

Attributor provides:

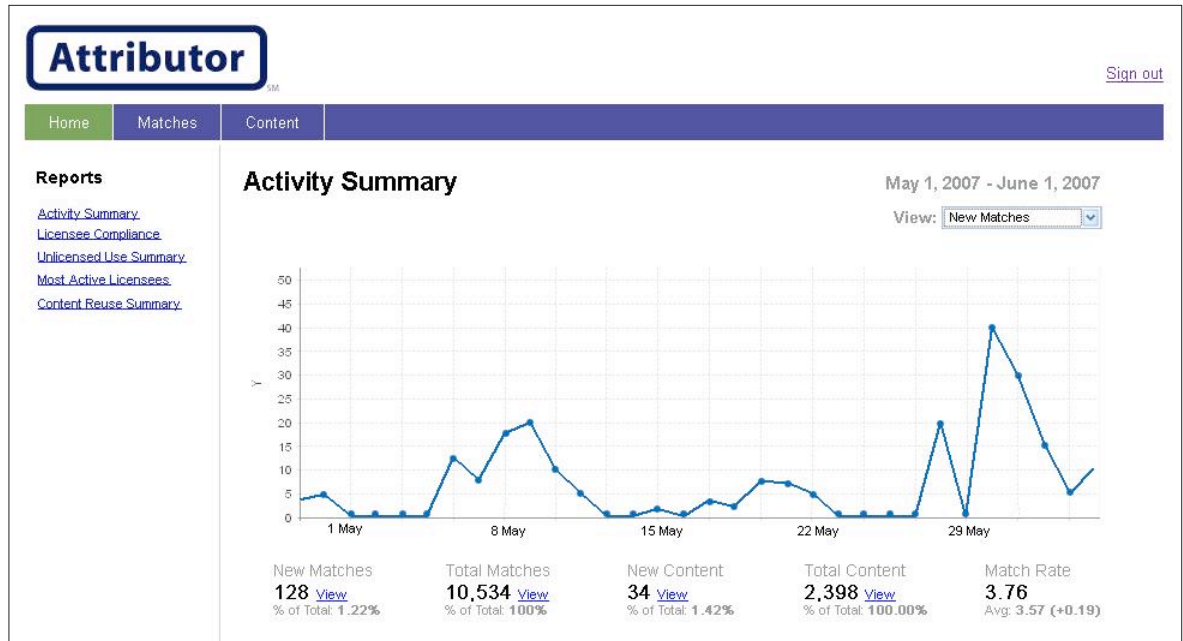
- **Context-sensitive analysis.** Publishers can see if:
  - o webpages that use their content are licensed or unauthorized
  - o the pages include ads (commercial reuse) or not
  - o how much of the content was reused
  - o attribution and links back to the publisher exist
  - o pages that reuse content are getting a higher search engine ranking than the original publisher
- **Accurate insights into the reach of content.** Attributor goes beyond link-tracking to show *all* matches where significant excerpts from articles are reused within html pages.
- **Trend analysis for all claimed content, not just the top tier.** From a single screen, publishers can monitor the reuse of all their content, even rank sections of content against each other, measuring the impact of editorial changes across the web.

The screenshot shows the Attributor Match Summary View. At the top, there's a navigation bar with 'Home', 'Matches', and 'Content'. Below that, a summary line states '151,296 matches on 4,434 match sites for All My Content'. There are filters for 'View' (All Matches, Flagged Matches, Licensed Matches, Unlicensed Matches, Matches by Status) and 'Group By' (None, Content, Match Site). The main table lists match sites with columns for Matches, Commercial, Not Attributed, and % Re-used, along with Licensee and Ignore checkboxes.

Match Sites	Matches	Commercial	Not Attributed	% Re-used	Licensee	Ignore
<a href="#">nysun.com</a>	8,464	0	8,464	59%		<input type="checkbox"/>
<a href="#">topix.net</a>	3,876	3,876	3,862	24%		<input type="checkbox"/>
<a href="#">mworld.com</a>	3,775	0	3,775	37%		<input type="checkbox"/>
<a href="#">nwsourc.com</a>	3,595	2,478	2,934	68%		<input type="checkbox"/>
<a href="#">hillsdale.net</a>	3,057	0	3,057	74%		<input type="checkbox"/>
<a href="#">examiner.com</a>	2,993	2,728	1,248	60%		<input type="checkbox"/>
<a href="#">chron.com</a>	2,848	2,848	2,037	73%		<input type="checkbox"/>
<a href="#">lycos.com</a>	2,773	0	682	74%		<input type="checkbox"/>
<a href="#">yahoo.com</a>	2,628	2,525	1,632	40%		<input type="checkbox"/>
<a href="#">cantonrep.com</a>	2,585	2,419	771	63%		<input type="checkbox"/>
<a href="#">hendersonvillenews.com</a>	2,176	0	2,176	72%		<input type="checkbox"/>
<a href="#">aol.com</a>	1,989	1,954	1,948	64%		<input type="checkbox"/>
<a href="#">cbsnews.com</a>	1,856	1,525	1,856	74%		<input type="checkbox"/>
<a href="#">wilmingtonstar.com</a>	1,820	1,820	1,820	67%	✓	<input type="checkbox"/>

**Figure 3: Sample Attributor Match Summary View.**

The Match Summary screen displays all matches to your content, grouped by domain. It also shows additional valuable—and actionable—information about the match, such as whether it gave attribution, was used commercially, and how much of the content was reused.

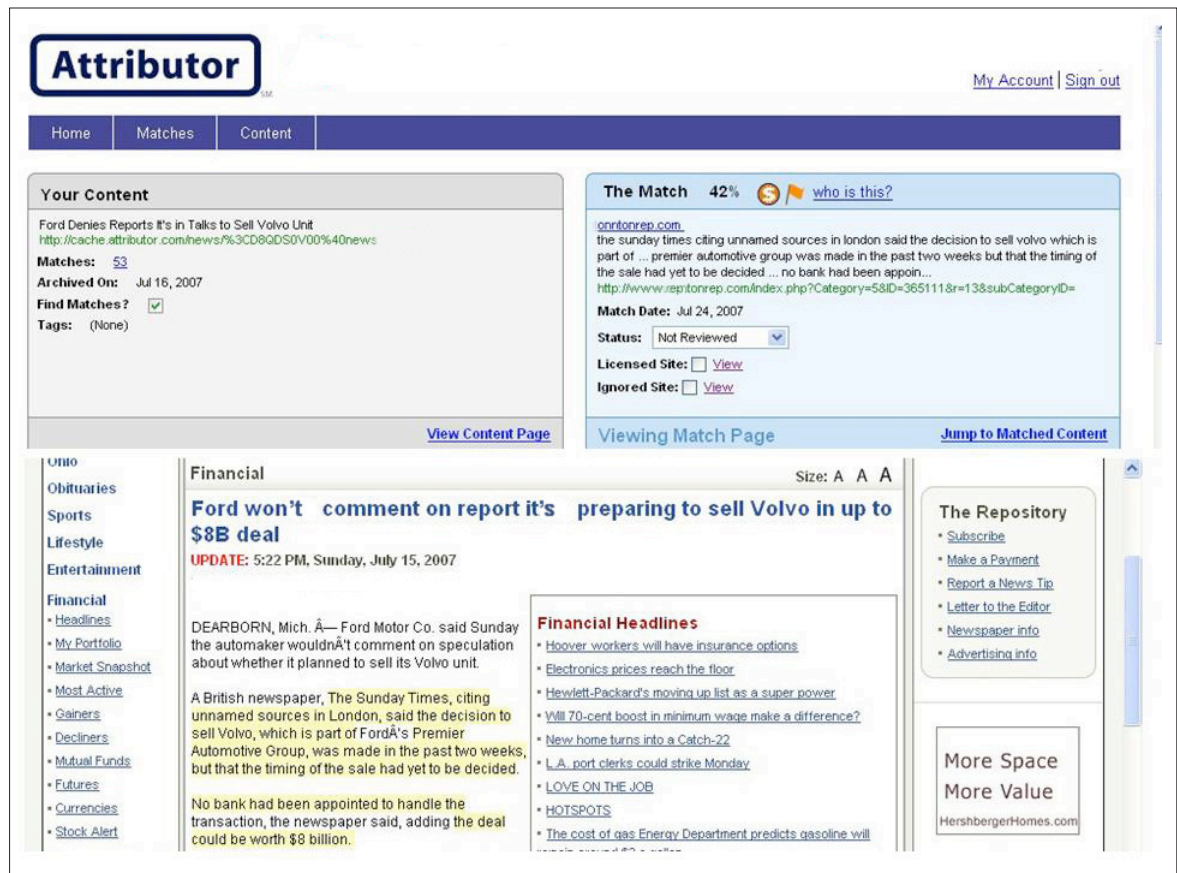


**Figure 4: Sample Attributor Activity Summary.**

The Activity Summary includes built-in reports, as well as export functionality to let you integrate Attributor data with other analytics. Together, these features enable a wide range of useful business insights.

With Attributor, users can easily manage different types of copyright and licensing agreements and act on reuse quickly, appropriately, and in an automated fashion. The system’s Remedy Platform automates requests for licensing, replacement, or removal of matching content, and allows users to track resolutions from a single screen.

For example, when Attributor flags sites that are non-commercially reusing claimed content without attribution or link-backs, it offers a simple way to request attribution. For sites that are commercially reusing claimed content without authorization, Attributor makes it easy for licensing and sales departments to follow up, or if need be, the system empowers users to generate a DCMA takedown notice. To make this simple and efficient, the Attributor’s Remedy Platform screen displays the entire range of remedy options, tracks progress from one remedy to the next, and automatically triggers email alerts if a site that’s reposting content ignores a request. And by offering easy access to multiple remedy options, content publishers and those reusing their content can more easily reach satisfactory negotiated outcomes on reuse—potentially creating successful new distribution opportunities.



**Figure 5: The Attributor Remedy Platform turns reuse into opportunity.**

Review the details of each match and take the appropriate action using the Attributor Remedy Platform. You can track progress on all remedies from a single screen.

Millions of “splogs” (spam blogs) are using RSS feeds to automatically scrape content for financial gain.

Attributor also makes it possible to efficiently measure and test key online strategies, such as using full-text vs. excerpt-only RSS feeds. This lets news organizations understand the impact of either choice, and aim to optimize their distribution and reach with confidence.

# In Operation—Attributor at the Associated Press

## Put Attributor to Work for You

### 1. Claim your content.

Register at [www.attributor.com](http://www.attributor.com). Once you are logged in, you can claim your content—manually or by RSS feed—and set the rules for how others can use it online.

### 2. Understand its use.

See all content matches and view Attributor's reports to learn where and how—licensed or unlicensed, credited or not—your content appears on other websites and search engines.

### 3. Profit from its distribution.

Use Attributor's analytics to gain deeper insights and share distribution opportunities with licensing and editorial staff. In addition, Attributor's Remedy Platform can help protect your content's value by acting on unauthorized use and flagging previously untapped revenue opportunities.

Attributor's approach of claiming content, understanding it, and profiting from it is already hard at work for the Associated Press.

As one of the world's largest producers of news content, AP content is very broadly syndicated. And with Attributor's technology, the AP can track its content better than ever before. Attributor fingerprints each AP story, and analyzes its distribution across the web, providing AP with context-sensitive insight of how it's being used.

The benefits to the AP are clear: With greater visibility into where its content goes, the AP can more efficiently track and pursue unauthorized usage of its content. This also benefits AP's licensees, by protecting the content's value for these paying members. In addition, through better understanding of usage patterns, the AP can find new potential licensing and revenue-sharing opportunities.

As AP's content is so widely distributed, providing this kind of monitoring and analysis for AP's stories highlights the scalability of Attributor's platform—and shows off how comprehensive Attributor's web crawl coverage is for news and current information.

AP's member licensees (such as regional and local newspapers) can also benefit from Attributor's services. Greater visibility into reuse of their original content across the AP network and beyond—and better understanding of how it is being repackaged—lets them improve editorial focus and ad sales efforts. The Attributor system is easy to use and affordable, so regional and local news organizations can quickly learn what strategies are most effective online and increase much-needed revenues.

"The publishing and broadcasting world has learned that networks are necessary to take advantage of the scale of the web. As the technology to tag, track, and monetize content gets into place, it will not be implemented simply publisher by publisher, broadcaster by broadcaster, but by networks—the implementations that we're increasingly seeing in advertising and content distribution are all network-based," says Ken Doctor. "At this point, we see a series of networks doing deals with other networks. And underneath the business terms that evolve between these networks, this very simple tagging structure needs to exist—one that both gives the publisher the security of knowing how the content's being used, and how it is being monetized."

## Conclusion

News organizations and other original content creators must embrace new technologies to survive and thrive in the online world, where digital content moves about more and more freely. Fundamentally, understanding where their content is going, how it is being used, and by whom is crucial.

By automating and systematizing the monitoring of news content, Attributor provides publishers with the tools and insight they need to meet the challenges of the new online world. Attributor's technology helps news organizations gain visibility into the use of their content across the web, level the playing field, and capture full value from online distribution.

***For more information about Attributor and a free sample report that shows where your content is being reused around the web, visit [www.attributor.com](http://www.attributor.com) or send an email to [sales@attributor.com](mailto:sales@attributor.com).***