

WHITE PAPER

How do you feed relevant external content into your Intelligence system?

Finding, validating & optimizing use of external content in Intelligence2day[®]

Contents

1	Introduction.....	3
2	Right content, focused result	4
3	All content has a cost - understanding TCO	7
4	Maximize the mix of effort, cost and relevance.....	10
5	From source to intelligence	12
6	Use cases: When you have the right content.....	15
7	Next steps.....	19
8	About Comintelli	20

1 Introduction

Every manufacturer knows that you can't create a quality product if you haven't got the right raw materials to start with. The same is true for intelligence systems: If you want high-quality results and impact, it is essential that you put the right data into the system in the first place.

Finding information for your intelligence system might seem like an easy task. After all, we're overwhelmed by a constant flood of data - that's one of the reasons we value intelligence systems.

However, getting the most out of a system - and achieving the business impacts you and your stakeholders are looking for - requires an understanding of:

- Sources that may have the kind of information that will lead to the result you need
- The variables to consider when selecting sources
- What it takes to manage the right mix of source materials for intelligence.

Your intelligence system might incorporate internal data (customer information, email server contents, reports, etc.) and external data (web-based content, externally produced newsletters, public filings, etc.)

This White Paper focuses entirely on **external sources of data** and provides tips on how to find and validate sources to get the most value from your intelligence system.



2 Right content, focused result

Many organizations get started with an intelligence system because they need the power of technology to make sense of the overwhelming amount of information they need to process in order to innovate, work efficiently, and boost meaningful collaboration. The notion of a platform that can ingest data from diverse sources, organize it, and present it in precise, useful formats will make any busy executive sigh with relief. Not so fast: It's technology, not magic.

The more you can feed the system with relevant data, the more you will get relevant results. Which begs the question: What is "relevant"? The answer to that question is different in every organization - and perhaps in different departments or business units of the same organization.

To determine relevance, consider the following variables:

- Are your topics customary to the business world? The consumer world? Or are they niche topics that have fewer sources of useful information?
- Do you need up-to-the-minute data, or are you more interested in long-term views based on older information?
- Are there particular authors, influencers or thought-leaders whose insights are particularly important? What about specific publication brands - will users trust the system if it doesn't include a particular source?
- Do you need to consider foreign language sources? Will translated versions of those sources be available and trustworthy?
- Is critical information available through multimedia sources, like video or audio files?

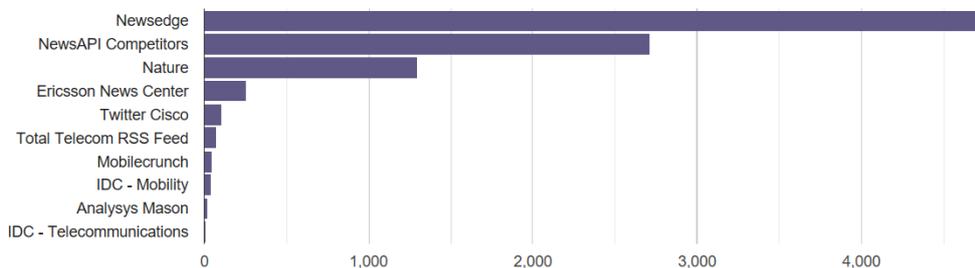
Work through these questions, with input from the people who will be (or already are) using the system. Make a list of sources already in use, regardless of where they come from. Ask questions about why these sources are valuable and trustworthy, and think about where you might find other sources like them.

Keep in mind that not every kind of information you want may be available at a price you are willing to pay. Many organizations would love to have detailed competitive pricing information... but that data may not be available. However, if you are creative about the types of information that might make reasonable proxies, you can often come up with insights that fill the bill.

In your intelligence platform, you should review statistics on source usage to look for any patterns on what seems to be rising to the top in terms of quality and usefulness. For example, the figures below show screenshots from the intelligence system Intelligence2day®, indicating how frequently information from different sources is used. Once you have this kind of insight, you can consider what qualities the most valuable sources have, which you can replicate with other sources.

Top feeds, by articles July 2019

EXPORT TO EXCEL



Feeds, Usage 2018-12-01 – 2019-02-01

EXPORT TO EXCEL

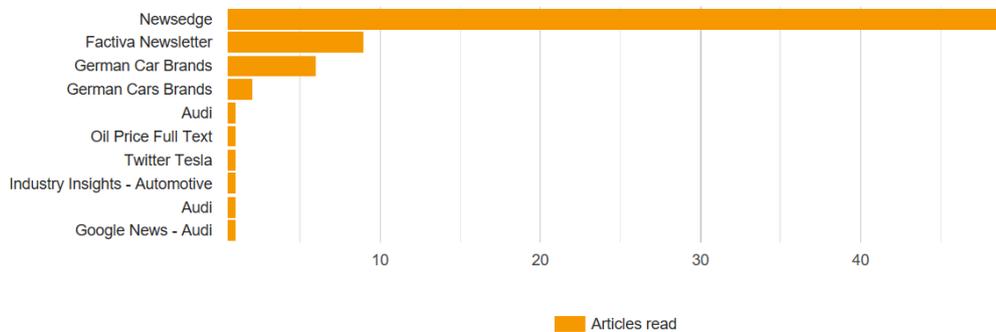


Figure 1 - Review usage statistics to determine most relevant sources

When you finish this process, you will have a long list of sources, which can include:

- Internet sites (corporate websites, blogs, news pages)
- Open Sources (Google, RSS feeds)
- Email newsletters
- Paid-for reports or newsletters (e.g., Automotive World, ENDS Europe)
- Premium Content Providers (e.g. full text databases like NewsEdge, Factiva, LexisNexis)

When you consider other types of sources you might want in your system, you might also run across:

- Government and regulatory data
- Patents
- Job listings
- Social media
- YouTube or other video environments

We'll talk in a bit about the pros and cons of each type of source further on, but for now, it's enough to know that a robust intelligence system can handle any of these.

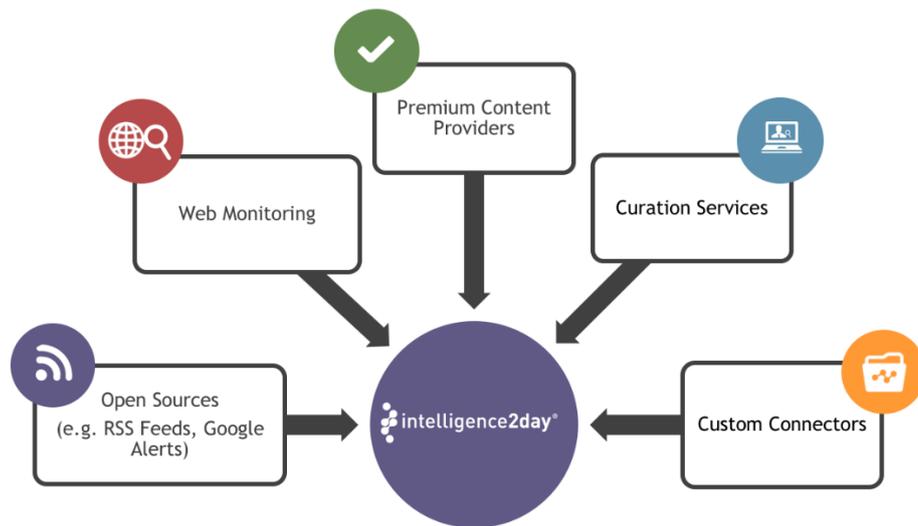


Figure 2 - Centralize and combine diverse sources

However, you should also know that just because you *can*, doesn't mean you *should*. It's tempting to simply add everything you have access to and hope for the best: Throw it all in and let the smart technology take it from there. But the extra time and effort you put in to analyze relevance will enable the smart technology to deliver more focused results - a true example of when "less is more".

3 All content has a cost - understanding TCO

When evaluating sources for your intelligence system, keep in mind the law of the universe: All content has a cost! There is no such thing as “free”. All sources cost you something.

In the case of paid-for sources - often referred to as “premium sources,” the cost includes the license or purchase fee, plus the time it takes a member of your staff to manage the relationship with the vendor, troubleshoot, and renew the contract. There may also be costs associated with your implementation, depending on any customization you need to do. And some suppliers (though not all) will have a separate fee for using their content in an intelligence platform.

But what about so-called free sources? Website updates, RSS feeds, free newsletters, blogs, even Google Alerts... yes, they have costs - sometimes significant costs - associated with them. It is essential to keep in mind these factors which contribute to the Total Cost of Ownership (TCO):

3.1 Cost of your time

The most significant cost associated with any content is the cost of your time - or of any other member of staff who needs to invest in:

- **Validating content:** Is the information you are collecting this way trustworthy? How do you know? How would you know if quality degrades over time?
- **Managing and troubleshooting:** Getting the content into the system may take some effort the first time, and then it needs to be managed. What happens if a link break? If a site starts blocking you?
- **Risk management:** If you have a contract for content, you have the peace of mind knowing that the supplier has a legal commitment to the supply. When you get content from free sources, you have no guarantees. Would your system - or the results you get from it - suffer if sources were suddenly unavailable? And what happens if a free source moves to a paid model - are you at risk for losing it?

3.2 Relevance cost

Another often misunderstood cost is that of relevance. The use of Google Alerts to support intelligence is an excellent example of using a free source that results in a significant



relevance cost. As handy as Google Alerts is, relying on it for intelligence means that you are absolutely missing relevant material... even when that material is in fact indexed by Google!

The company Contify [published an excellent blog item summarizing](#) the shortcomings of using Google Alerts to support intelligence efforts. There are a host of reasons that make Google Alerts a poor choice for intelligence, but the main things to keep in mind are these:

- Google's business model depends on users actively using search and clicking on search results
- Google Alerts circumvents that behavior.

To test the service, Contify staff set up Google Alerts on 240 companies and monitored the results. They found that only 10 percent of the results were business-relevant, and the alerts only picked up 46 percent of the potentially relevant information found through other sources. [Read the full results here.](#)

Google Alerts is only one example of how free sources can result in relevance cost, but we include it here because it is such a popular - and misunderstood - choice.

3.3 Balancing cost and impact

With an understanding of the true TCO, you can conduct a more thorough evaluation of content cost compared with its impact on your results. A matrix like the one in Figure 3 is a useful way of visualizing TCO against impact.

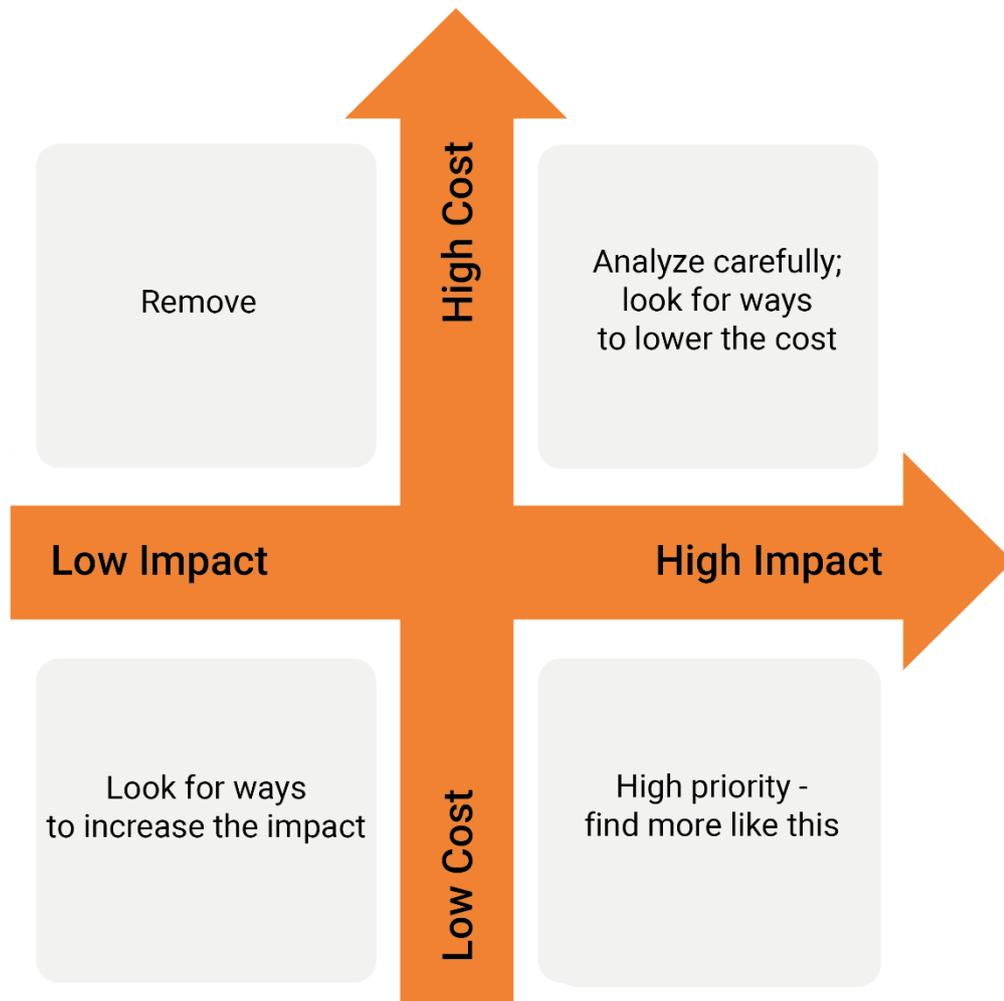


Figure 3 - Content Cost/Impact Matrix

Note on the grid where content sources belong: Move them up the vertical axis as the cost goes up and move them to the right on the horizontal axis as their impact improves.

4 Maximize the mix of effort, cost and relevance

It's here that the understanding of the pros and cons of different types of content becomes important. Some of the cons will significantly add to costs associated with your time and with the relevance of the content for your intelligence process. At the same time, pros will add to the value, so you want to balance these out, along with actual contract costs to come to the right mix of effort, cost and relevance.



The table below walks through the pros and cons, and then summarizes the key considerations you need to take when using external content.

Source	Pro (+)	Con (-)	Considerations
Open sources (RSS, Google)	Easy; free	Rarely full-text; links break and require monitoring	Manage through proactive review to ensure links work and content remains valid
Web monitoring	Easy; free	High-value content is rarely published on free websites; links break, even more frequently than RSS feeds	Manage through proactive review to ensure links work and content remains valid
Social media	Easy; free; high volume of data	Most social media information from companies is promotional; rarely high-quality for most B2B needs, except through purpose-built monitoring systems	Proactive review needed, but not as frequently as for websites and RSS feeds
Premium content - single title	High quality; contractual agreement to supply content; often not available on internet; available in full-text; often includes metadata and other features that enable it to deliver greater value through an intelligence platform; support from supplier to enable use; you have greatest control over content selection	Costs may go up at renewals; use within a platform may not be included in standard T&Cs; full-text or other content features (e.g. figures, tables) may not be available	Vendor relationship can be as important as content itself; contract management will require investment of time
Premium content-databases	Broad range of quality sources selected by professionals in content validation and licensing; content is normalized by supplier and enhanced with metadata, indexing, etc.; full-text usually available, for at least some content	Costs may go up at renewals; not all curated content from a supplier may be available for use within a platform, or with other restrictions (e.g., no full text, no figures, etc.); individual titles may be removed from the service without warning	Depending on your needs, may require monitoring of sources in the collection, to ensure you don't lose access to something you need; vendor relationship can be as important as content itself; contract management will require investment of time

Figure 4 - Pros and Cons of using different types of external content

5 From source to intelligence

Once you have a portfolio of sources, they need to be ingested into the platform in order to start producing intelligence. Intelligence2day® includes a number of different ways of adding feeds and connecting content, depending on your need.

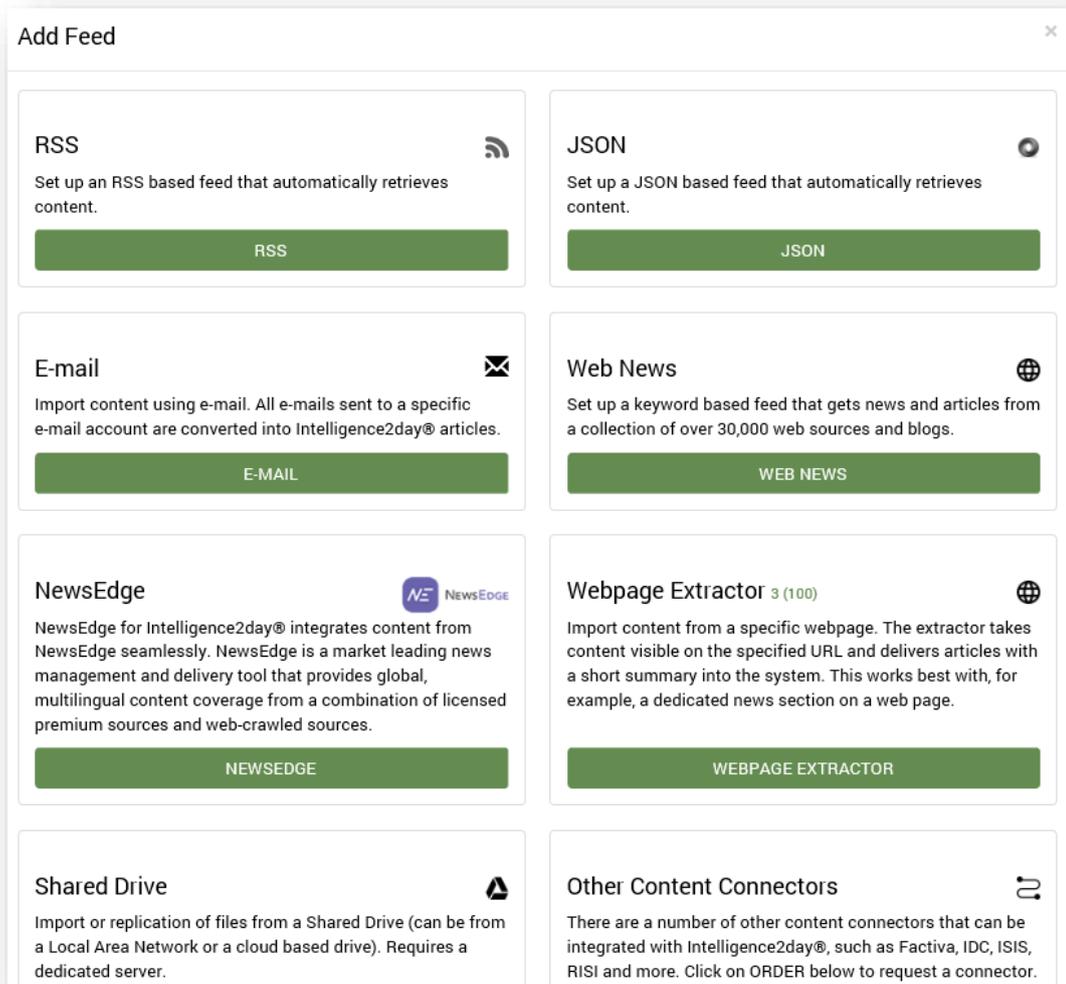


Figure 5 - Multiple methods of connecting content in Intelligence2day®

The main types of content that can be ingested are these:

- **RSS/JSON feeds:** Add the link to the feed to bring the results into Intelligence2day®. Note that feeds rarely incorporate full text unless specified by the publisher, and these formats will not support attachments.
- **E-mail:** If you subscribe to e-mail newsletters, you can set up an e-mail box in Intelligence2day® and direct them all to this one place where all users can share them.
- **Web News:** Given the importance of websites as sources of intelligence, Comintelli has integrated a web monitoring tool that crawls 30,000 news sources and blogs in over 50 countries to capture the latest articles and news headlines. You can define the keywords you are interested in, and title, summary and link for relevant content items will automatically be pulled into Intelligence2day®.
- **NewsEdge:** Comintelli's partnership with NewsEdge from Naviga enables us to provide easy access to this robust global database of premium and full text news. Note that adding NewsEdge to your Intelligence2day® is a premium feature for an additional fee.
- **Web Extractor:** Comintelli's Web Extractor helps you create custom feeds from any public webpage with frequently updated information like news, articles, discussion forums, reviews, jobs, events, products, blogs, press releases, social media (Facebook, Twitter, Instagram) etc. It utilizes powerful AI based machine learning capabilities to auto-detect relevant content for a feed from its source.

More than 90% of the time the Web Extractor AI picks out the right content and automatically creates a quality feed, but there are cases where it does not. There are a few reasons this could occur, and these cases will require manual intervention to solve:

- The structure of the source website is built in a way that "confuses" the machine, picking out the wrong content.
- The source website is built in a way that simply prevents content from being extracted.
- The source website only displays the "title" of the news on the page, no summary. It extracts the content that is displayed on the actual page, "what you see is what you get".

Intelligence2day® also includes Custom Connectors for many of the frequently requested premium sources and databases, such as:

- BMI Research
- Dow Jones Factiva
- FactSet
- Forrester
- GlobalData
- ICIS
- IDC
- IHS Markit
- Lexis Nexis
- Markets & Markets
- MergerMarkets
- M-Brain
- PatBase
- RISI
- Strategy Analytics
- and many more...

In addition to Custom Connectors, Intelligence2day® can work with just about any type of data - including images, audio and video files, presentations, PDFs... if it's digitized, we can find a way to pull it into your intelligence platform. Intelligence2day® has a number of **APIs** (Application Programming Interface) for integrating and connecting content from different sources. All API's uses RESTful webservices and can be used by any suitable programming language.

Finally, you can consider working with a **professional curation service**. These services add a human element to the gathering, filtering and sorting of external data from websites, social media, news and more, and can provide the results seamlessly into Intelligence2day®.

6 Use cases: When you have the right content.

To sum up, we have included two best practice cases that show how Intelligence2day® customers have enhanced their intelligence environments through careful selection, integration and management of the right set of sources.

6.1 CASE 1 - Global Technology Company: Efficiency and effectiveness of Market Intelligence

The information center at a global technology company with almost 100,000 employees was using an in-house solution to track market intelligence from hundreds of sources across sectors. In 2015 they found that their users wanted an improved way to search amongst these sources as well as the ability to receive e-mail alerts on content relevant to them. They also hoped to find an intelligence system that could accommodate social features such as the ability to like and comment on content. Intelligence2day® matched the information center's requirements.

Which sources?

The team incorporates three types of content into their system:

- Reports from licensed research vendors
- News from premium business sources
- Internally produced material

How does it run?

One employee, known as the information manager, is responsible for keeping the system running optimally. This includes:

- Evaluating sources
- Creating feeds
- Manually uploading content
- Troubleshooting
- Working with vendors to create FTP and API import mechanisms
- Negotiating content licenses.

The information manager speaks with approximately 25 of their 165 content vendors on a weekly basis and an additional 15 vendors on a monthly basis. Additionally, the

manager maintains a quarterly scorecard - a report of the type and volume of content that the vendor has delivered over the quarter and its usefulness to users.

The information manager also conducts annual surveys asking users to rate the types of content they use in their jobs. While these efforts are time-consuming, they produce these benefits:

- Knowing whether a content source is performing consistently. If not, they investigate whether they need to troubleshoot for a technical issue, or whether perhaps the content is no longer valuable to the business.
- Reinforcing confidence in the value of each vendor's content, which allows for data-driven budgeting decisions and advocacy as well as a position of strength in negotiations.

Content tip!

The information manager suggests paying close attention to negotiation and documentation of the **licensing terms for each piece of content**.

These terms are a critical element of copyright-related risk management.

Recommendations include paying for two or three sources with terms that enable users to use and share the content in a widespread way and paying extra for departmental or global licensing rather than city-based licensing.

Ultimately it is more time- and cost-efficient to be confident about widespread use than it is to track down licensing agreements for many disparate pieces of content or for how that content can be used in different locations.

6.2 CASE 2 - International Manufacturing Company: Consistency and quality across the organization

An international technology manufacturer with approximately 15,000 employees has multiple information centers serving parts of the company related to the different sectors in which they have product portfolios.

Each information center had a different solution for tracking market intelligence, some of which were in-house solutions. Since there was no cooperation or communication between these systems, a steering committee was formed to select one product that could be used across the corporation. They have been using Intelligence2day corporation-wide for nine months and employees are excited about the platform.

Which sources?

Each of the four information centers using Intelligence2day® has its own sources for market intelligence; in total 1300 different feeds are established throughout the system.

Big competitors and industry topics are covered through intelligence provided by the company's Factiva license. However, it's important to note that not everything can be found in licensed sources.

For example, one center's competitive landscape is heavily weighted toward small private companies that cannot be effectively tracked through licensed sources. To capture information about these competitors, the information center has set up RSS feeds of new information on the websites of those private companies, Google alerts on competitors, and Google alerts on specific topics.

How does it run?

While they were running a pilot program, the information centers were on the phone weekly with Comintelli to get feeds set up, to tweak settings and to troubleshoot after software upgrades were made. However, now that the system is up and running, the only thing that consistently takes time to maintain is finding and fixing broken links.

One challenge they have is figuring out how to get better information about markets in China and Japan. News monitoring in these languages is difficult, and company websites seem to change constantly, requiring the information centers to update their feed settings.

Content tip!

The information centers at this company recommend looking at three factors when examining potential content sources:

- **Right content:** Consider from which sources you can get the best content. If you consistently need intelligence from licensed sources, you will need to pay for that content. But there are several situations where licensed sources will not offer what you need. For example, if you need videos and images in your content, you can only get those from public sources. If your competitors are privately-held companies, the licensed sources will probably not deliver the desired content.
- **Cost:** Consider the financial implications of how many free sources versus licensed sources you will need to employ.
- **Efficiency:** Using free sources is not as efficient as using licensed sources. With a licensed source, you will get the full text of an article delivered to the platform, whereas free sources will pull in only a few sentences of an article, and you will have to copy-paste the full text manually into the platform.

7 Next steps

Having the right information is a game-changer, in terms of the power, relevance and overall ROI of your intelligence system.

In addition to the insights offered through this White Paper, we invite you to request a copy of “**The Comintelli eWorkbook on External Content**”, which provides worksheets and guidance to help you:

- Analyze your need for content
- Find appropriate content
- Balance TCO and impact
- Develop a calendar to manage the content that feeds your system.

We live in a world with a vast wealth of information at our fingertips. The more discerning you are about what constitutes “value” for your intelligence requirements, the better you’ll be able to extract the good stuff and make the most of it.

8 About Comintelli

Comintelli is a Swedish software company which sells Intelligence Software that converts unstructured Big Data content into organized, digestible information for decision-making.

The award-winning solution Intelligence2day® acts as an insight engine to help customers make faster and more confident decisions.

Founded in 1999 and with extensive intelligence experience, Comintelli continues to develop user-friendly solutions that shortens Time-To-Insights.

For more information, please contact:

Phone: +46-8-663 76 00, US/Canada: 1-800-485-6402

E-mail: contact@comintelli.com

Web: www.comintelli.com | www.intelligence2day.com

(Visit our Resource Center for more White papers and Cases)

Comintelli makes no warranties, either expressed or implied, in this document. Information in this document is subject to change without notice. The entire risk of the use or the results of the use of this document remains with the user. No part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means or for any purpose without the expressed written permission by Comintelli. The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

© Comintelli AB. All rights reserved.

