AIIM Market Intelligence

Delivering the priorities and opinions of AIIM's 80,000 community

aiim Industry Watch

Content Analytics: automating processes and extracting knowledge

Underwritten in part by:

KOFAX from Lexmark

OPENTEXT™

Rocket.

SPS a Swiss Post company

aiim®
The Global Community of Information Professionals
aiim.org | 301.587.8202

About the Research

As the non-profit association dedicated to nurturing, growing and supporting the information management community, AIIM is proud to provide this research at no charge. In this way, the entire community can leverage the education, thought leadership and direction provided by our work. We would like these research findings to be as widely distributed as possible. Feel free to use individual elements of this research in presentations and publications with the attribution – "© AIIM 2015, www.aiim.org". Permission is not given for other aggregators to host this report on their own website.

Rather than redistribute a copy of this report to your colleagues or clients, we would prefer that you direct them to www.aiim.org/research for a download of their own.

Our ability to deliver such high-quality research is partially made possible by our underwriting companies, without whom we would have to return to a paid subscription model. For that, we hope you will join us in thanking our underwriters, who are:

KOFAX from Lexmark

Kofax, a Lexmark Company
15211 Laguna Canyon Road
Irvine, California 92618
Tel: +1 949-727-1733
Sales Tel: +1 949-783-1333
Email: Contactme@kofax.com
Web: www.kofax.com

OPENTEXT™

OpenText
275 Frank Tompa Drive
Waterloo, Ontario
Canada, N2L 0A1
Tel: +1 519-888-7111
Web: www.opentext.com

Rocket®

Rocket Software
77 Fourth Avenue
Waltham, MA 02451
781-577-4321
Tel sales: Andrew Lee 781-684-2176
Email: alee@rocketsoftware.com
Web: www.rocketsoftware.com/solutions/enterprise-search-and-text-analytics

SPS
a Swiss Post company

Swiss Post Solutions AG
Pfingstweidstrasse 60b
8080 Zürich
Switzerland
Email: global.sps@swisspost.com
Web: www.swisspostsolutions.com

Process Used and Survey Demographics

While we appreciate the support of these sponsors, we also greatly value our objectivity and independence as a non-profit industry association. The results of the survey and the market commentary made in this report are independent of any bias from the vendor community.

The survey was taken using a web-based tool by 238 individual members of the AIIM community between April 17 2015, and May 08, 2015. Invitations to take the survey were sent via e-mail to a selection of the 80,000 AIIM community members.

Survey demographics can be found in Appendix 1. Graphs throughout the report exclude responses from organizations with less than 10 employees taking the number of respondents to 222.

About AIIM

AIIM has been an advocate and supporter of information professionals for 70 years. The association mission is to ensure that information professionals understand the current and future challenges of managing information assets in an era of social, mobile, cloud and big data. AIIM builds on a strong heritage of research and member service. Today, AIIM is a global, non-profit organization that provides independent research, education and certification programs to information professionals. AIIM represents the entire information management community: practitioners, technology suppliers, integrators and consultants.

About the Author

Doug Miles is head of the AIIM Market Intelligence Division. He has over 30 years' experience of working with users and vendors across a broad spectrum of IT applications. He was an early pioneer of document management systems for business and engineering applications, and has produced many AIIM survey reports on issues and drivers for Capture, ECM, Information Governance, SharePoint, Mobile, Cloud, Social Business and Big Data. Doug has also worked closely with other enterprise-level IT systems such as ERP, BI and CRM. Doug has an MSc in Communications Engineering and is a member of the IET in the UK.

AIIM The Global Community of Information Professionals
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
+1.301.587.8202
www.aiim.org

Table of Contents

About the Research

Introduction

Drivers and Adoption

Process Automation and Inbound Routing

Information Governance and Metadata Generation / Correction

Contextual Search, Curation, and E-discovery

Analysis / Business Insight / Customer Input

Big Content Projects

Opinions

Spend

Conclusion and Recommendations

Appendix 1: Survey Demographics

Appendix 2: General Comments

UNDERWRITTEN IN PART BY

Introduction

The capacity of computers to recognize meaning in text, sound or images has progressed slowly and steadily over many years, but with the arrival of multi-processor cores, and the continual refinement of software algorithms, we are in a position where both the speed and the accuracy of recognition can support a wide range of applications. In particular, when we add analysis to recognition, we can match up content with rules and policies, detect unusual behavior, spot patterns and trends, and infer emotions and sentiments. Content analytics is a key part of "big data" business intelligence, but it is also driving auto-classification, content remediation, security correction, adaptive case management, and operations monitoring.

The first step for many analytic processes is capture and recognition – from paper, from emails, and from other inbound channels. This in itself involves validation and some "intelligent guesswork" based on word matching and sentence construct. Similar principles can be applied to search and knowledge extraction, moving beyond simple keywords to contextual analysis, taking into account the significance and use of the search terms.

Humans hate filing. Even more, they hate sifting content for deletion - and they are generally bad at it. Computers are much more consistent in their application of rules, and given suitable criteria for classification, or for deletion, can hugely reduce unwanted content. This improves the searchability and business value of what remains, and also make-safe any sensitive content. Beyond this, we can use meaningful extraction of comments, opinions, diagnoses, reports, claims, social chat, and so on, to gain business insight, improve competitive advantage, or achieve fast response.

In this report we will look at the take-up of analytics applications for inbound routing and text recognition, for content classification and metadata correction, for improved search and knowledge extraction, and to provide business insight. We look at the success factors and outcomes, and the choices being made for deployment.

Key Findings

Drivers and Adoption

- **73% of respondents agree that enhancing the value of legacy content is better than wholesale deletion.** 53% agree that auto-classification using content analytics is the only way to get content chaos under control.
- **54% feel that their organization is exposed to considerable risk due to stored content that is not correctly identified.**
- **73% consider that there is real business insight to be gained if they can get the analytics right.** 63% are being held back by a lack of analytic skills and an absence of allocated responsibilities.
- **34% of responding organizations are using content analytics for process automation, information governance, contextual search or business insight.** A further 44% have plans in place.
- **17% consider content analytics to be "essential" now for their organization, growing to 59% in 5 years' time.** Plus 28% feeling it "is something we definitely need".
- **The biggest issues for adoption are lack of expertise (36%), and a need to set information governance policies first (36%).** 43% admit that their current capability in enterprise search is poor, 33% have problems with BI, and 19% have poor ECM.

Process Automation

- **15% are using OCR data capture of inbound content for process input, 14% are auto-classifying content for archive, and 12% are auto-routing to specific processes or to case-files.** 10% are triggering processes from inbound content, including 5% from mobile device input.
- **5% have fully automated filing or archiving of inbound emails, and 11% user-prompted filing.** 24% have plans in the next 12-18 months.

- **Benefits from inbound analytics include faster flowing processes (50%), happier staff (32%) and improved governance (20%).** 18% are seeing high levels of "hands-off" processing.

Information Governance

- **20% are already using auto-classification to assist staff with filing, metadata tagging, or records declaration, and 17% have immediate plans.** 18% are using automated or batch agents to correct metadata for improved searchability, to better align metadata between repositories, or to detect security and compliance risks.
- **Improved search is the biggest benefit of auto-classification (reported by 52%) along with better staff productivity (40%), and improved compliance and governance (31%).** Defensible deletion and recovered storage space are also reported (19%).

Contextual Search and Curation

- **Only 35% have contextual search, including 11% across multiple internal sources and 7% across external sources.** 8% rely heavily on their contextual e-discovery tools, although a further 10% have them but don't use them.
- **19% have some automated curation tools to create custom libraries and alerts, although 9% are from internal sources only.** 6% have manual curation processes. 59% have neither, but feel it would be useful.

Business Insight

- **24% have at least one "big content" project for business insight, with 10% having several.** Improved product or service quality is the strongest objective, followed by core investigations and research, and then detection of non-compliance.
- **Nearly half have used in-house development, and 17% external custom.** 27% have used cloud or SaaS products and 27% products from their ECM vendor.
- **34% have achieved ROI in 12 months or less, and 68% in 18 months or less.**

Spend

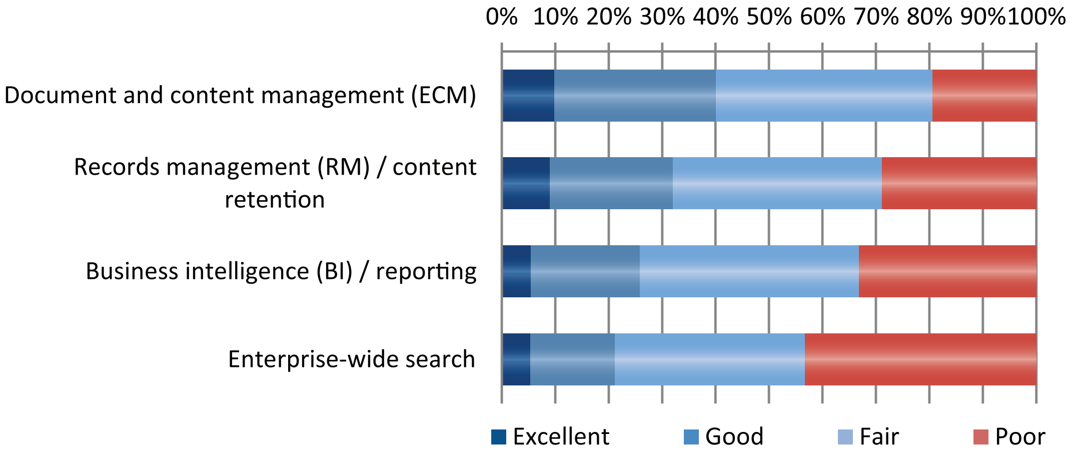
- **Most of our respondents expect to spend more on content analytics in the next 12 months.** Strongest growth is in enhanced or contextual search, analytics for business insight, and automated classification tools or modules.

Drivers and Adoption

Content analytics by its nature places demands on how content is stored and managed within the business. Poorly cataloged content spread out across multiple repositories and file-shares, immature information governance policies, and only basic search and BI tools will make knowledge extraction difficult. This is an area where many of the content correction and re-classification tools that we discuss later can help to improve these situations.

As we can see in Figure 1, 18% of our respondents rate their ECM capability as poor, although only 40% consider it to be good or excellent. When it comes to records management and content retention, 30% admit it is poor and only a third rate it as good or excellent. Business Intelligence (BI) and reporting is a frequent cause for complaint from line-of-business managers in most organizations, and 33% of our respondents would consider it to be poor. But the biggest shortcomings are in enterprise-wide search with 43% having poor capabilities, and only 20% in good shape.

Figure 1: How would you best characterize the following capabilities across your organization? (N=222)



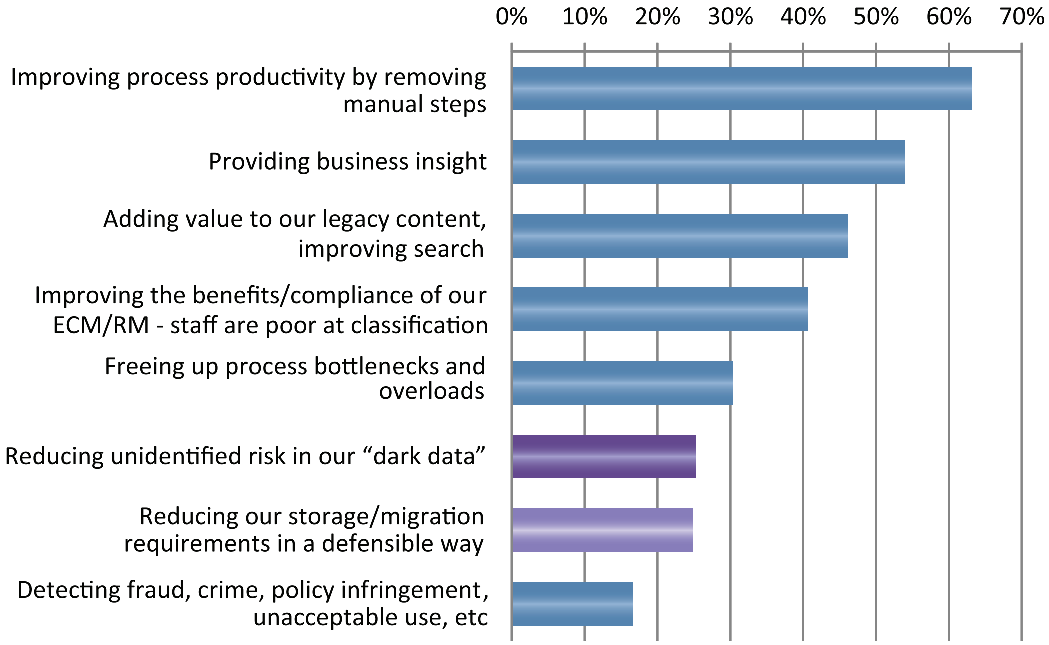
Against this background, it is understandable that many organizations may feel that content management comes first, with content analytics further down the track. However, it may well be that these low ratings come from poorly deployed or poorly used ECM and RM systems. This can be particularly true of many SharePoint implementations[1]. Automated classification, and content correction across existing content, would be a good way to re-vitalize these failed or stalled projects.

Drivers

Process productivity, business insight, and adding value to legacy content take the top places when it comes to key drivers. This is followed by improving the benefits and compliance of ECM/RM - by more consistent declaration and classification of records. Reducing unidentified risk in what is termed “dark data” is important for 25%, and this rises to 32% for the largest organizations. This refers to content which may contain sensitive or personally identifiable information about customers or staff, or may have business sensitivity.

In a more general sense, 25% are keen to use content analytics to help them reduce overall storage requirements, or to clean up content before migrating it to newer systems or consolidated repositories.

Figure 2: What would be the THREE biggest drivers for content analytics in your organization? (N=217)



Importance and Leadership

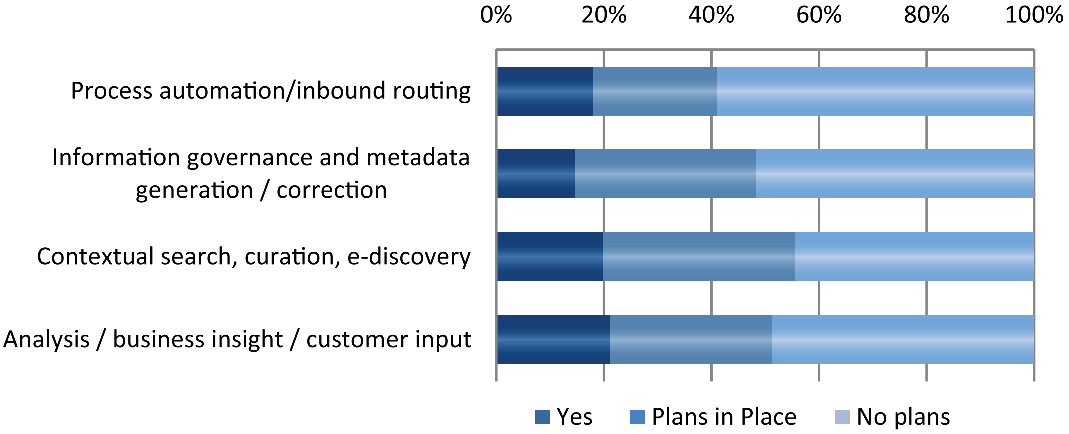
Looked at today, 17% of our respondents consider content analytics to be “essential”, with 48% feeling it is “something we definitely need”, but projecting that to five years’ time, this grows to 59% feeling it will be essential, and 28% a definite need, with only 13% seeing it simply as “useful”.

There has been much talk about the need for a CDO – variously described as a Chief Data Officer or Chief Digital Officer – to raise awareness and realize the potential of analytics or big data projects, but when we asked, only 4% of our sample have such a position, with 1% having a CAO or Chief Analytics officer. 10% said they have plans in place, and 6% felt their organization has such a job role, but not with that job title (CIO is given as the most likely alternative). By implication, therefore, 80% of our responding organizations have yet to allocate a senior role to initiate and coordinate analytics applications.

Adoption and Applications

Taking a broad look at adoption across the four areas that we have identified (and remembering that this is a self-selected survey and will over-read the general population) 38% are using content analytics for one or more types, with around 20% using any one of the types, and 20-30% with plans in place. Contextual search and e-discovery is the most popular overall, but information governance and metadata correction shows the most potential growth. Looking at usage across business sizes, mid-sized organizations (500-5,000 employees) are lagging somewhat, especially in analysis and business insight applications, where 14% have applications in use, compared to 28% of the largest organizations (5,000+ employees). Smaller organizations at 21% are surprisingly active here.

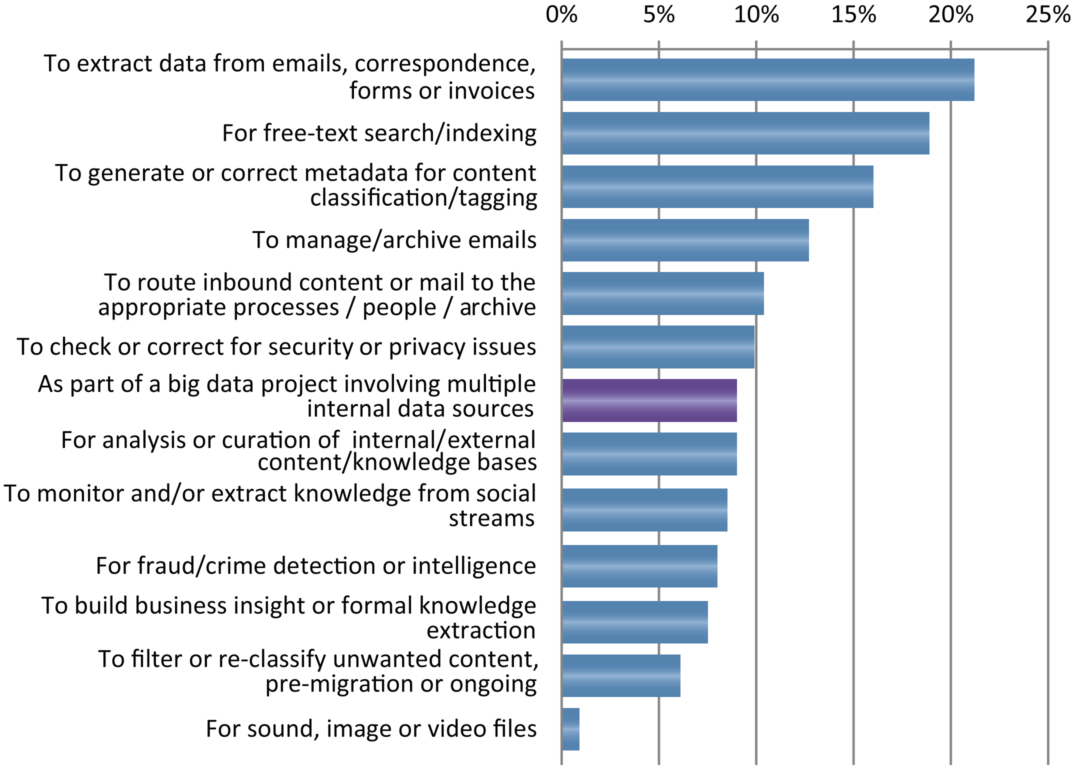
Figure 3: Are you using content analytics for any of the following? *(N=219)*



Looking in a little more detail at specific applications, 21% are extracting data from emails, forms or invoices – most likely invoices - and 19% are using free-text search, although it is likely that many of these applications do not use a high degree of text analysis, relying mostly on keyword extraction.

16% are generating or correcting metadata for content classification or tagging, and 13% are applying this to email management and archiving. 9% are using content analytics as part of a big data project across multiple data sources.

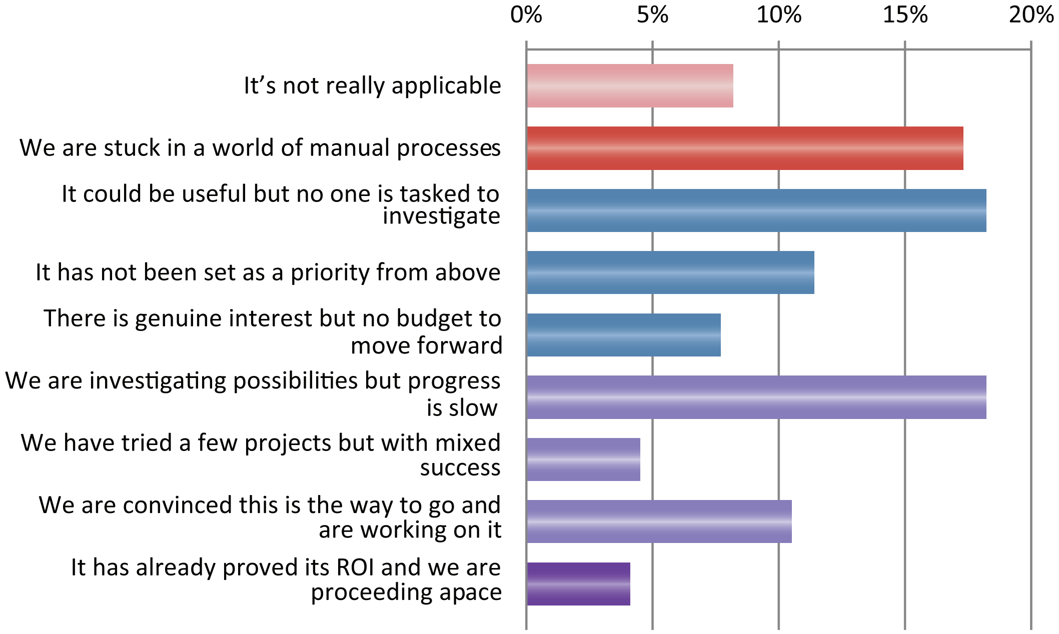
Figure 4: Are you currently using content analytics on unstructured content in any of the following ways? *(N=212)*



Progress and Issues

As with any relatively new software application, interest is high, but progress is mixed. A quarter of our respondents feel it is either not applicable, or that they are stuck in a world of paper processes. 37% either have no one tasked to investigate, no mandate from above, or no budget to proceed (or a combination of these). For 23%, a start has been made, but progress is slow, or of mixed success. 11% are underway and encouraged by the results, and 4% are already showing a return on their investment.

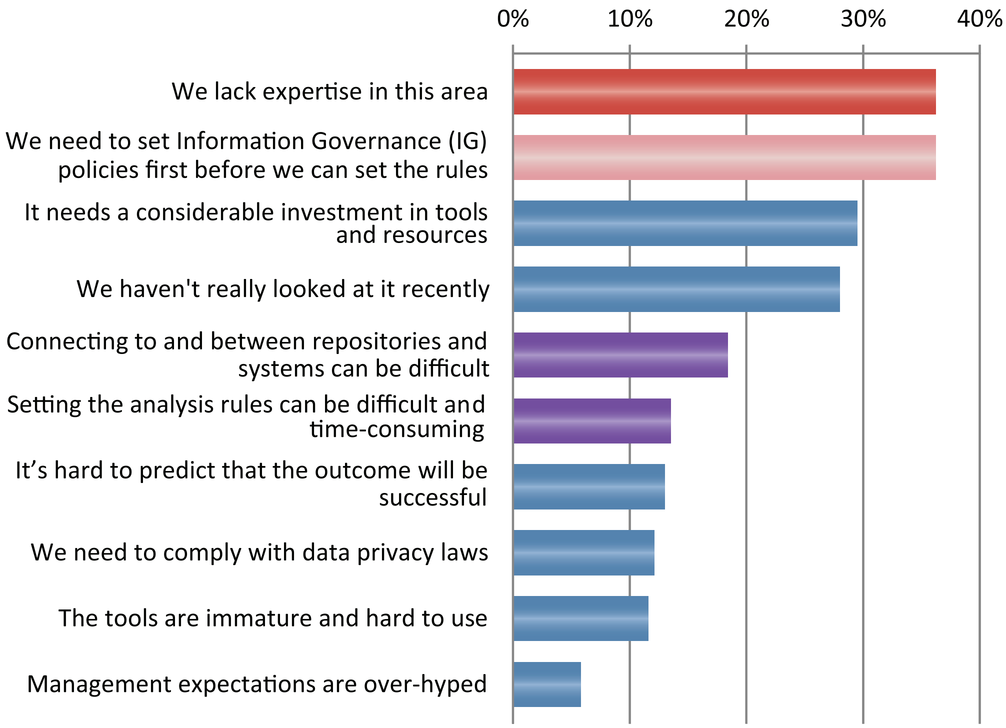
Figure 5: How would you best describe current progress in your organization towards the use of content analytics? *(N=220)*



Issues

Again, as we might expect for a new technology, lack of expertise is a big issue, reported by 36%. As we suggested before, not having firm and agreed information governance and content retention policies is also an issue that needs to be solved before rules-based classification can be implemented. Our respondents are also reporting some technical issues around connecting repositories and setting up the rules. Compared to big data projects in general, "over-hyped management expectations" does not seem to be a significant issue for our early adopters.

Figure 6: What are the biggest issues for you with content analytics projects? *(N=207)*



60% of our respondents feel that content analytics will become an essential capability for their organization within the next five years, and while initial efforts are a little varied in outcome, users are applying the technology across a range of application areas.

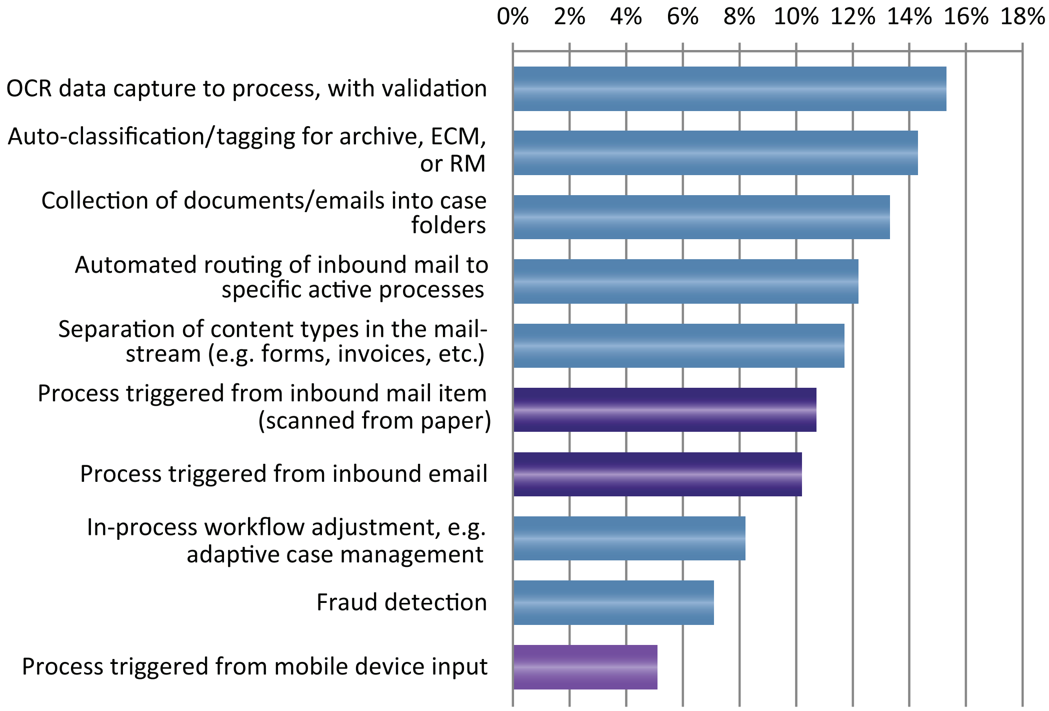
Process Automation and Inbound Routing

More recently tagged as "smart business processes", automated and adaptive processing based on analysis of inbound content has been growing steadily in recent years. As the volume, variety and urgency of multi-channel inbound content has grown, users have been looking at ways to reduce handling loads, speed up response, and embed compliance into their customer or supplier-facing processes. The most popular application has been invoice processing (accounts payable) where invoices are recognized out of the inbound mail, examined for layout of key fields and OCR'd to capture the actual data. This is then validated against the original purchase order data from the finance system.

Varying degrees of analytic capability can be built into this application, and it can, of course, be extended to any number of inbound forms. As the inbound capture extends across more and more types of content, especially where the digital mailroom concept is employed (centrally or distributed), recognition of content type and automated routing to specific processes becomes very useful. In many cases, the arrival of a specific form or piece of customer correspondence (paper or email) can kick off a downstream process such as on-boarding, a support ticket, or a claim.

It then becomes particularly useful if a case-folder is created, and subsequent inbound items, such as proof of identities, assessment reports, income statements, etc., can be automatically routed to the case folder. This is also where intelligent case management can use information derived from the inbound content to adapt the required processes within the case, ensuring that procedures are followed in a compliant way. The most advanced organizations (5%) are even able to trigger processes from mobile device apps.

***Figure 7: Are you using content analytics for any of these inbound content functions?** (N=196)*



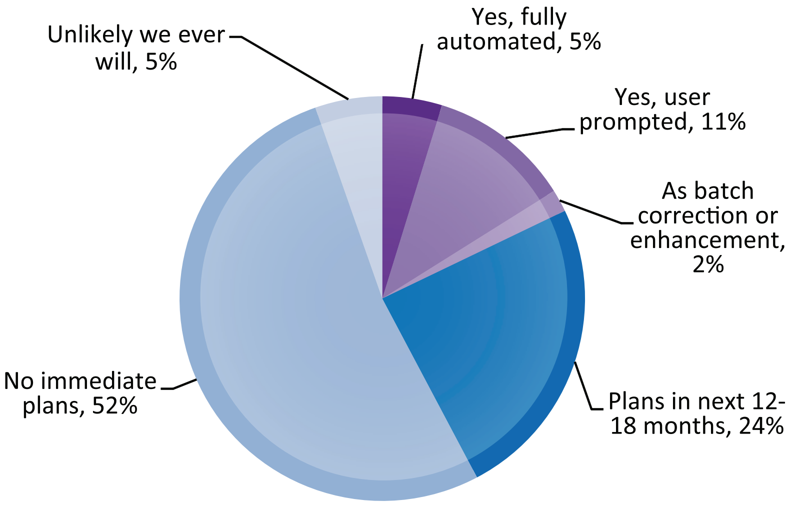
Automating Email Classification

It has been one of the longest running dilemmas of electronic records management systems as to whether to declare important emails as records into the system, and if so, how to rely on staff to do so reliably and responsibly, and how to avoid overloading the system with irrelevant records. As emails now carry full evidential weight in litigation cases, many organizations have implemented bulk email archiving systems, or long-term stored back-ups, in order to cover off potential legal discovery or freedom of information requests. Unfortunately, many of these archives are of the “store and forget” variety with little in the way of applied metadata, and no legal hold and e-discovery tools for contextual searches. They are certainly not optimized for surfacing knowledge or being part of the “corporate memory”.

Given that humans will never become consistent in filing and classification, and that the volume of emails continues to grow rapidly, automation is likely to be the only solution that can provide a usable and defensible way to archive emails. This may be fully automated, or may be a prompting system, asking users to confirm the suggested classification. As we will see later, there will be those who question the accuracy of machine classification, but email is particularly interesting in this context as most of us already rely on (and trust) a degree of spam filtering on our inbound emails, and the latest email clients are making their own judgments as to what emails to prioritize.

Only 5% of responding organizations are currently using fully automated classification of emails, with 11% using user-prompted techniques. However, a further 24% have plans in the next 12-18 months to do so, a sign that this long-running problem may finally be reaching an accepted solution.

Figure 8: Are you using auto-classification for filing or archiving inbound emails?
(N=168, excl. 34 Don't Know)

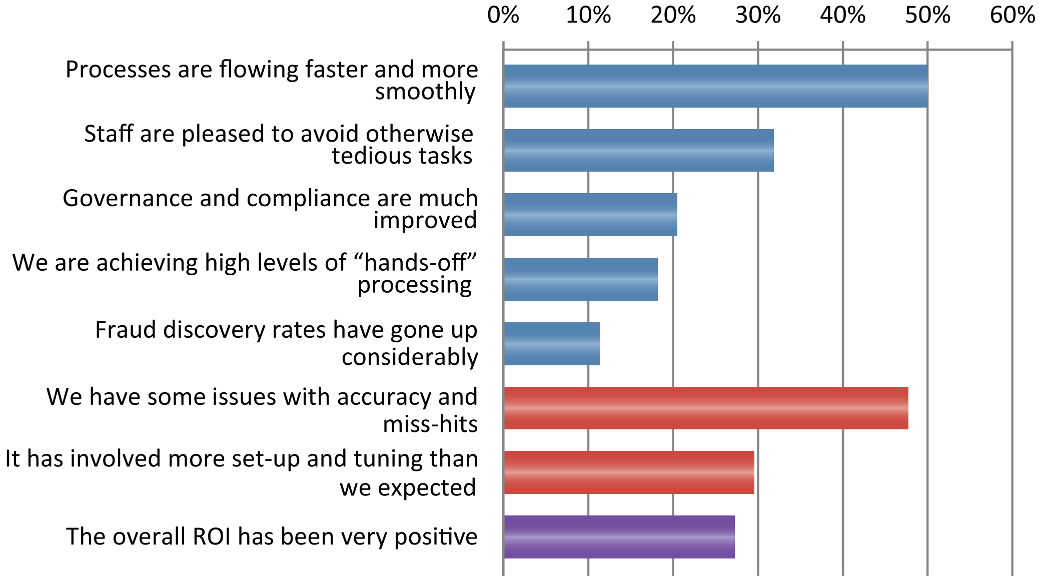


Project Success

The benefits of content analytics for users of inbound processing seem to be well defined. We can see in Figure 8 that processes are flowing more smoothly, staff are happy to avoid the tedious task of filing, and governance and compliance are much improved. As far as productivity improvements, 18% report that they are achieving high levels of "hands-off" processing, where large chunks of the process are handled by the computer.

There have been some issues, particularly accuracy and miss-hits, and to overcome those has involved a higher degree of set-up and tuning than some users were expecting. However, 27% report a positive ROI already.

Figure 9: How would you describe the success of your inbound analytics projects?
***(Check all that apply)** (N=44, excl. 102 "Not applicable", 50 "Too early to say")*

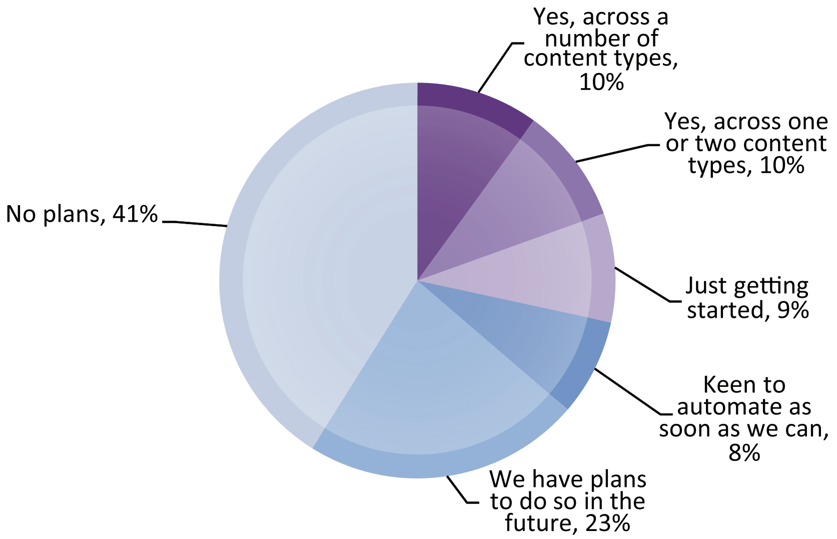


Only 5% of respondents have fully automated classification for filing or archiving emails, with another 11% having user-prompted filing. According to forward plans, this is set to more than double in the next 12 to 18 months.

Information Governance and Metadata Generation / Correction

We have seen a very rapid acceptance of the idea of auto-classification[2] for the purposes of improving compliance over the last three years, although as we will see, improving searchability is also a prime driver. In this survey 20% are already actively using it, with a further 9% just getting started. An additional 31% have plans to do so, including 8% in the short term. Overall, this represents nearly two-thirds of our respondents.

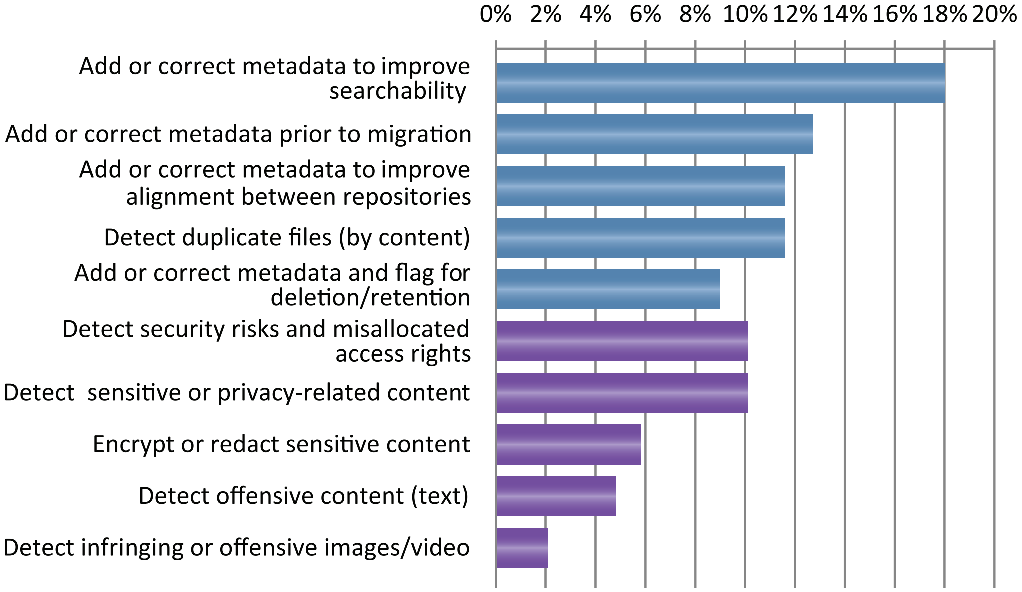
***Figure 10: Are you using auto-classification to assist staff with content filing / metadata allocation / records declaration?** (N=190)*



Although what we might call the classic view of auto-classification is that content is classified based on analysis of its text (or sound, or imagery) at the point of creation or ingestion, there is a strong application area that uses batch agents to crawl over existing content in whatever repository it exists, and to apply or correct its metadata based on a set of rules aligned to the information governance policy, and/or to the current taxonomy.

Once the metadata has been sorted out, many useful management controls can be applied. Searchability is improved, particularly in terms of accuracy and completeness. This can hugely benefit knowledge sharing and maximizes the value of stored information for research, reuse and audit, as well as speeding up the legal discovery process. Aligning metadata and taxonomies between repositories will also facilitate enterprise-search or content federation. If content is to be migrated between systems, aligned metadata is essential, and, of course redundant, obsolete and trivial content (ROT) can be left behind and deleted.

Figure 11: Do you use automated or batch agents to perform any of the following functions? *(N=189, 59% "None of these")*



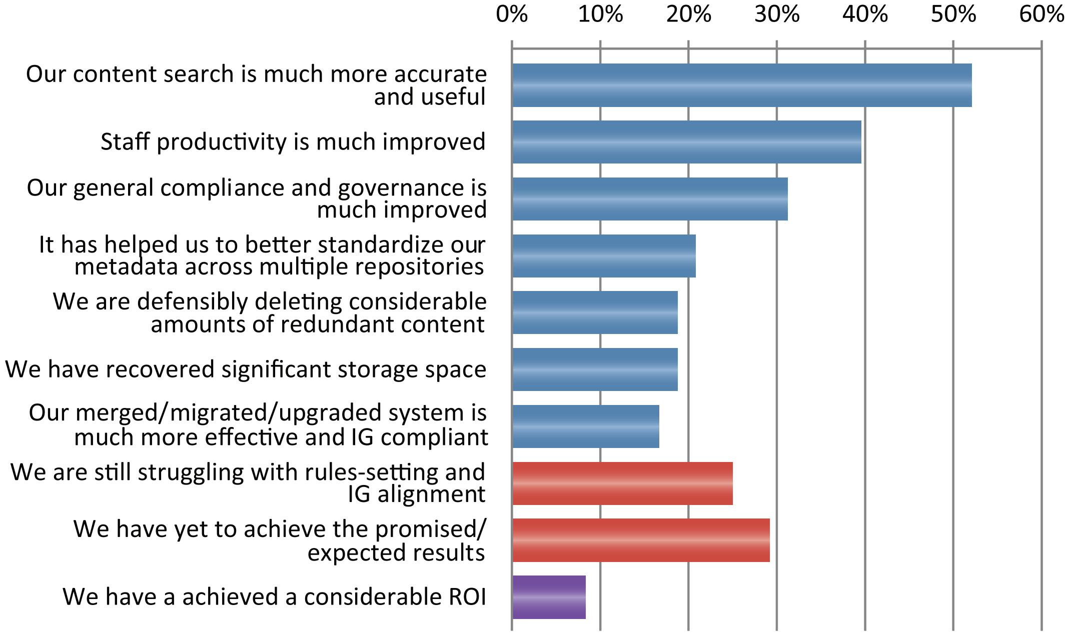
This removal of ROT, and also detection of duplicate content (even if filenames are different), can recover considerable amounts of storage space, which in itself speeds up and improves search. Content type-classification and correctly set metadata will be an essential step in determining retention periods, with the knock-on effect that potentially risky or non-compliant content can be defensibly deleted. If sensitive content is detected, it can be tagged for a higher access level and even encrypted or redacted for enhanced security.

Finally, offensive or unacceptable content can be detected, and dealt with immediately. For some organizations, this capability alone is sufficient to justify the purchase of a content remediation tool.

Project Success

52% of those using auto-classification report much improved content search, 40% have seen an improvement in staff productivity, and 31% feel that their general compliance and governance is much improved - a strong endorsement across a number of important goals within the business. The benefits continue: defensible deletion, recovered storage space and better optimized systems are all cited. On the issues side, some experienced difficulties with rules-setting to align with IG policies, and it is taking time for some to see the expected results.

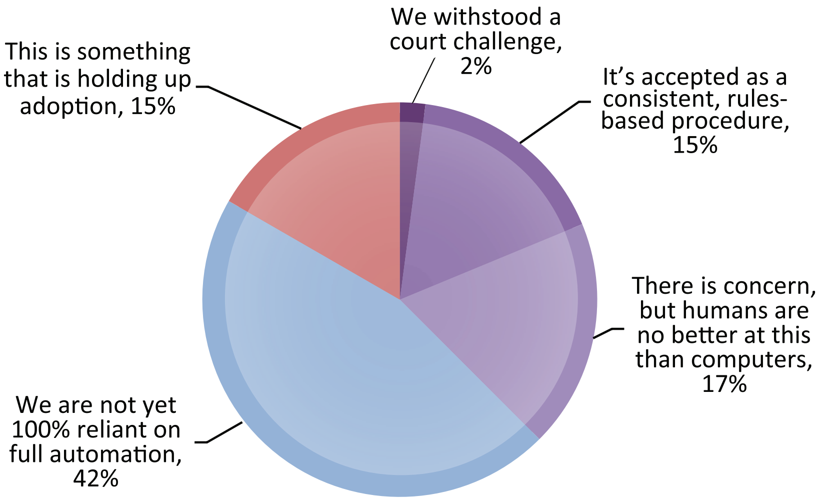
Figure 12: How would you describe the success of your auto-classification / metadata correction projects? (Select all that apply) *(N=48, excl. 99 "Not applicable", 43 "Too early to say")*



Legal Judgment

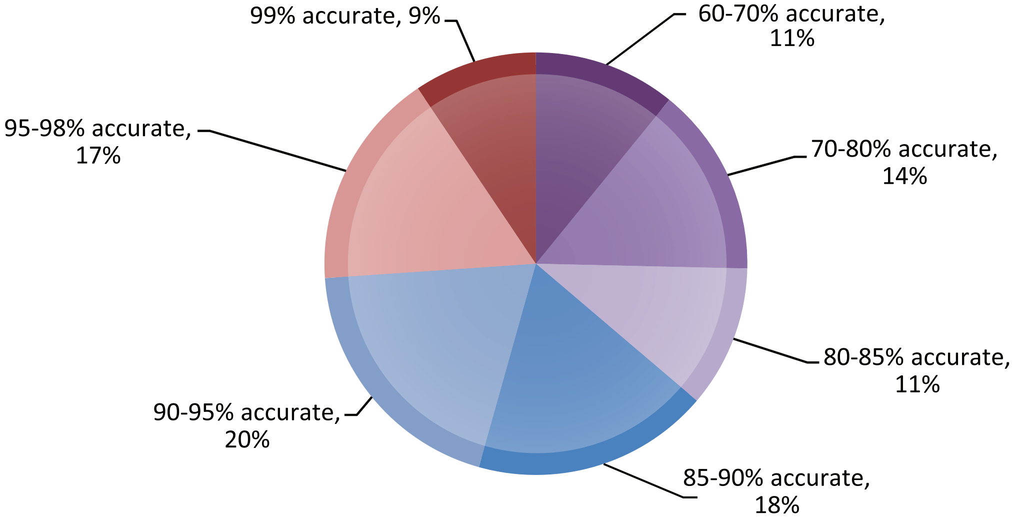
Knowing that some legal advisors might take a view that automated classification is not sufficiently accurate to rely on, particularly as regards deletion of emails, we asked if our respondents had encountered any legal resistance. 34% indicated wide acceptance within their organization, including 2% who withstood a challenge in court. Of the remainder, 42% are not in full operation, and only 15% report that this issue is holding up adoption.

Figure 13: Have you encountered any legal resistance or compliance questions regarding auto-classifying emails or other records pre-deletion? *(N=52, excl. 136 Don't Know, N/A)*



As a follow up question, we asked what degree of accuracy of classification, both for emails, and for general content, might be deemed acceptable in their organization. We also suggested that this should apply to human classification as well as automated. More than a third (36%) are OK with an 85% accuracy or less, another third (38%) with 95% or less. Only 26% feel that greater than 95% accuracy is needed, including 9% who are seeking 99% accuracy. It would be interesting to audit the content systems in these companies to see if human accuracy can actually achieve these levels!.

Figure 14: For emails and general content, what would you consider to be an acceptable accuracy of classification within your organization (human or automated)? *(N=138, excl. 47 Don't know)*



37% are using or just getting started with auto-classification, and are seeing the benefits of corrected metadata in searchability, productivity and compliance. 74% are looking for an accuracy of 95% to avoid any legal resistance.

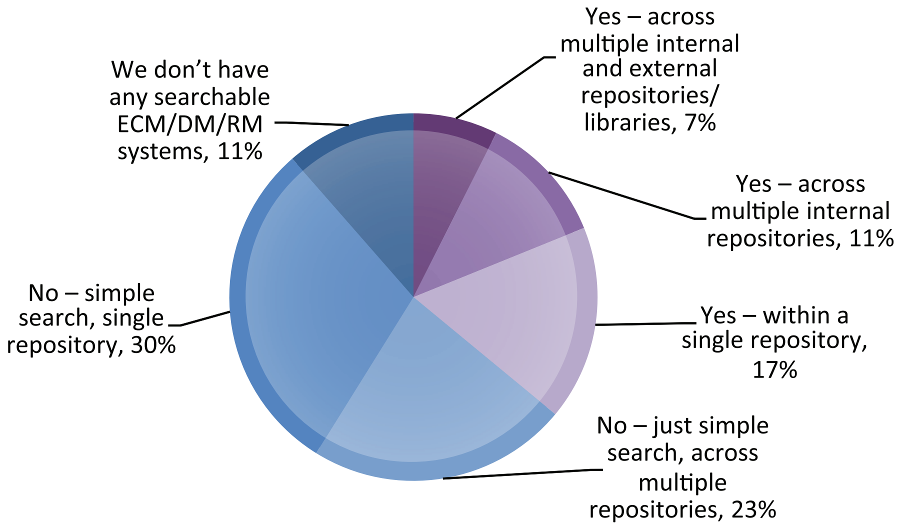
Contextual Search, Curation, and E-discovery

As we mentioned earlier, many content search engines rely on simple keyword searches, perhaps extended with some Boolean capabilities. Users are increasingly frustrated that these search methods fall so short of what is available with Google search on the web. Of course, indexing web pages, with their links and popularity, is somewhat less demanding than searching across multiple corporate repositories for important but little-referenced documents.

Users expect the indexing to include the significance of the keywords, as set by their position in headlines, body text, and so on. They are looking for differentiation between authoritative documents (and authors) and others. They only want the final version of a contract, or the customer letters that threaten legal action. They may like captions and annotations on drawings, or even photos, to show up in the keyword index.

Only 35% of our respondents have any form of contextual search, and this includes 17% who are restricted to a single repository. 7% have sophisticated search across multiple internal and external repositories or libraries. A third are restricted to simple search across a single repository, or do not even have a searchable ECM/DM/RM system.

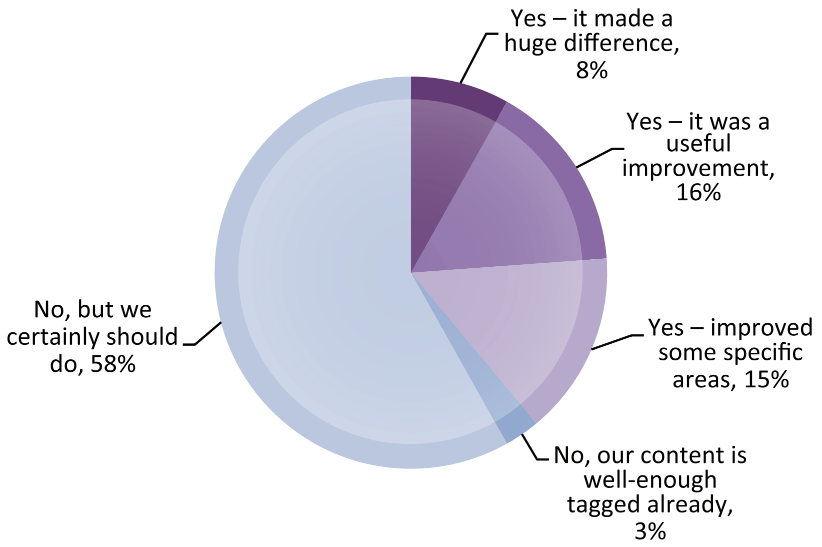
***Figure 15: Do you have a search capability that includes contextual analysis (as opposed to simple free text or keywords)?** (N=175, excl. 16 Don't Know)*



Metadata Creation/Correction

We talked earlier of adding value to the dark data that exists in most organizations, and the way to do this is to use content remediation or correction tools to trawl through the content, and intelligently add metadata, or fix metadata that is wrong or doesn't match the current classification scheme. In this way, even less sophisticated search tools can be made much more effective. 39% have improved their search capability this way, with 8% feeling that it made a "huge difference".

Figure 16: Have you used metadata creation/correction on existing content to improve searchability? *(N=191)*

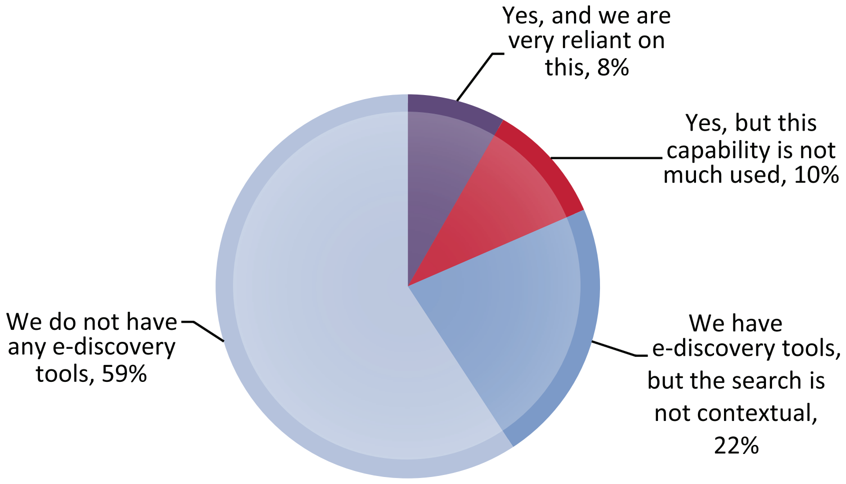


E-discovery

Contextual analysis can be particularly useful for pre-trial e-discovery work, picking up on contract terms, intellectual property, survey reports, complaints, etc. Internally, it can also be used for compliance audits. For example, price-fixing, tax avoidance, money laundering, fraud, etc., will all have a likely vocabulary and context that can be detected using much the same techniques as external fraud detection.

Having said that, it would seem from our results that half of those who have such a tool (10%) do not use it very much. 22% have e-discovery tools that are not contextual, 59% have no tools, including 29% of the largest organizations.

Figure 17: Do you have e-discovery tool(s) with contextual analysis capability? *(N=157, excl. 35 Don't Know)*

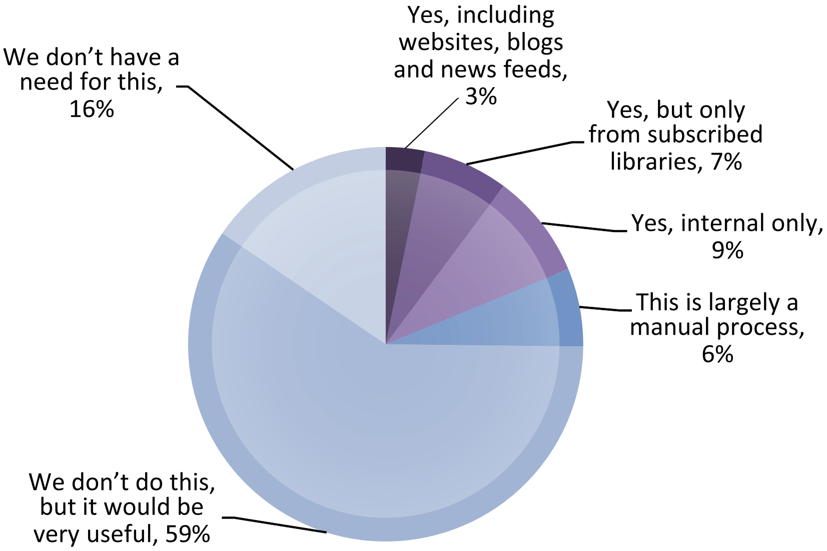


Curation

In many industry sectors such as medical, pharmaceutical, legal, aeronautical, it is important to stay abreast of published content from elsewhere, and in the past the curation of this content would be the role of the company librarian, often with a physical library of books, research reports and periodicals. Today, that sifting or curation role can be assigned to computers, collecting electronic content, and feeding specific references on defined topics to those that need them. However, to truly replace the previous role, the content needs to be collected from outside the business, and include websites, blogs and news feeds.

19% of our respondents have some automated curation, although half of those are internal only. 6% have the traditional manual process. Of the rest, 59% feel it would be very useful to have such a service for their key knowledge workers.

Figure 18: Do you use content curation to automatically create custom libraries and alerts from multiple external and internal sources? *(N=187)*



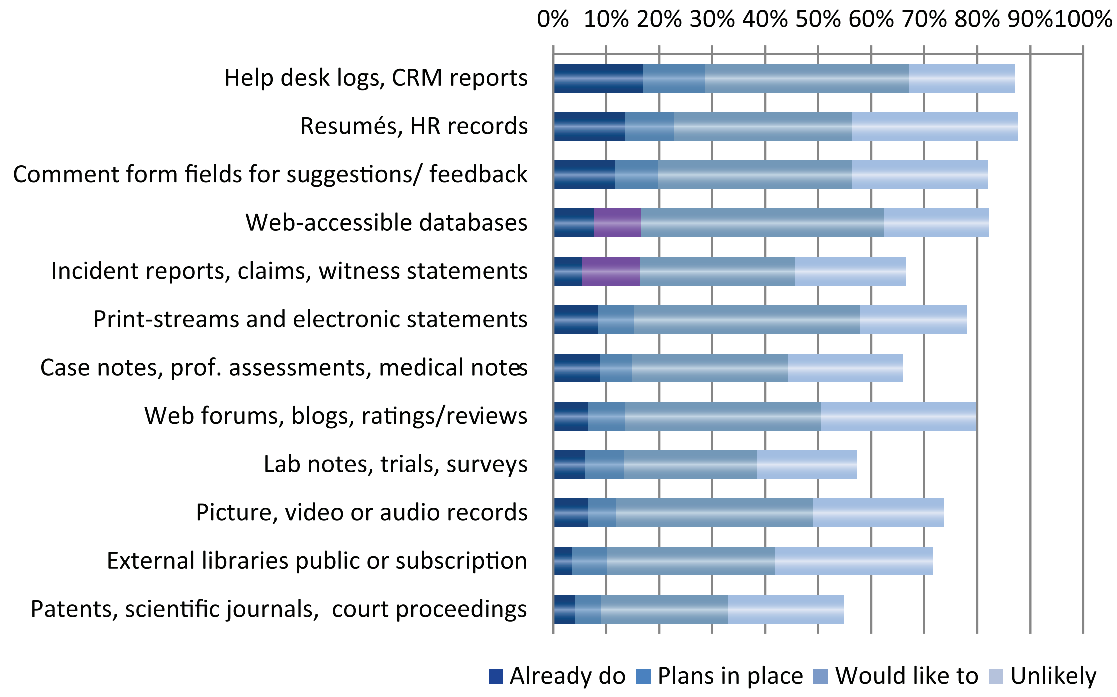
Only a third of organizations have contextual search, but half of those are restricted to one repository. 39% have improved their search with some form of automated metadata creation or correction.

Analysis / Business Insight / Customer Input

AIIM first reported on content analytics 5 years ago. Our subsequent reports picked up on the big data theme, or "big content" as we prefer to call it. The problem then, as it is now, is to come up with a pick-list of the most common applications. Then it was mostly based on blue-sky thinking: what would be the most useful thing for your business to know? Now we have a much more established set of applications, although that is not to say that there aren't plenty of innovative uses yet to come.

Now, as then, help-desk logs and CRM reports are the most popular source for analysis, picking up on customer experience and marketing insights, and a little further down, the free-form comment fields from feedback forms. Next come HR applications, particularly screening résumés for match with job specifications. Web accessible databases figure highly for plans-in-place, and this is often a curated feed, or might be a check of publicly available data, e.g., FBI records for previous convictions as part of a loan application. Similarly, incident reports, claims and witness statements are all part of fraud detection or due diligence.

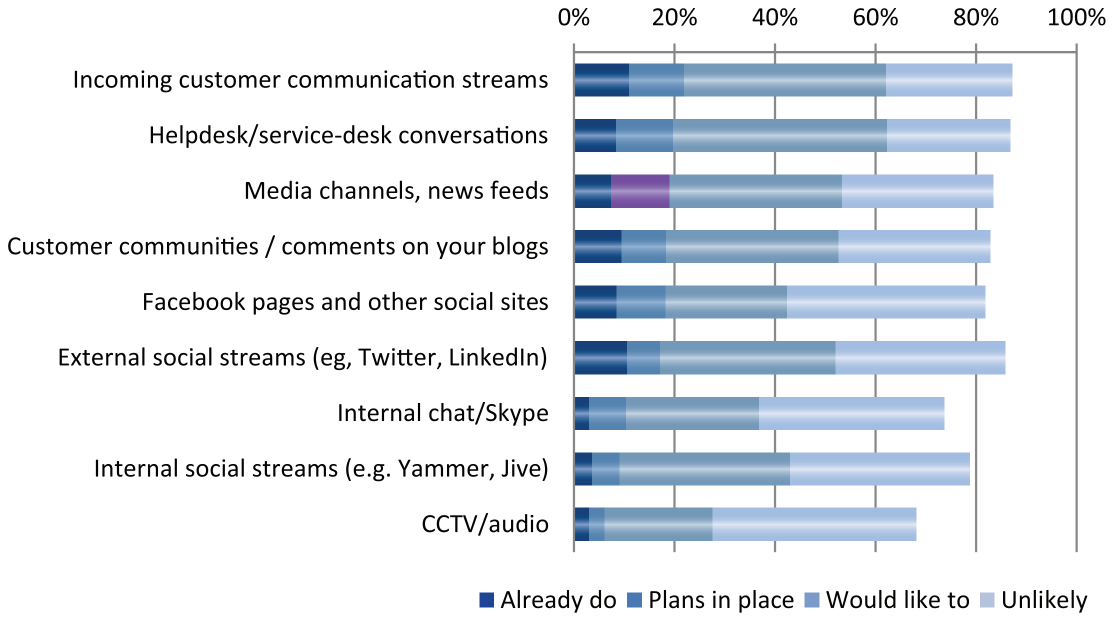
***Figure 19: Have you considered analyzing any of the following document or content types to extract business intelligence or solve problems?** (N=178. Line-length indicates "N/A")*



Real-Time or Near-Time

Incoming customer communications and help-desk streams also top the list for live or near-time alerting, along with an increasing interest in media channels and news feeds. There is, quite rightly, as much interest in what customers are saying on the organization's own community pages as on external social streams, and the former is set to grow more. CCTV and audio monitoring obviously have their place, but this is a more difficult technology.

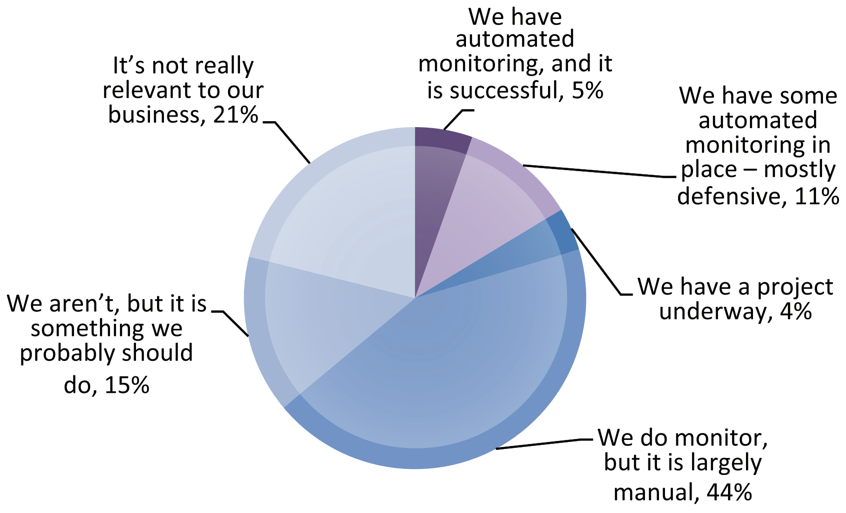
***Figure 20: Have you considered automated analysis of any of the following to extract live or near-time business intelligence?** (N=178. Line-length indicates "N/A")*



Social Media Monitoring

Looking in more detail at social media, the importance of monitoring these fast-moving streams has soared in the past few years, and as a result most organizations have implemented a monitoring mechanism (64%) but only 14% have an automated system . Relying on (designated) staff to alert the marketing or customer service department when complaints (or praise!) show up can be somewhat hit-and-miss, and the speed of response can be crucial in these situations. Automated monitoring using sentiment analysis is a much more reliable way to alert the appropriate people to make a response.

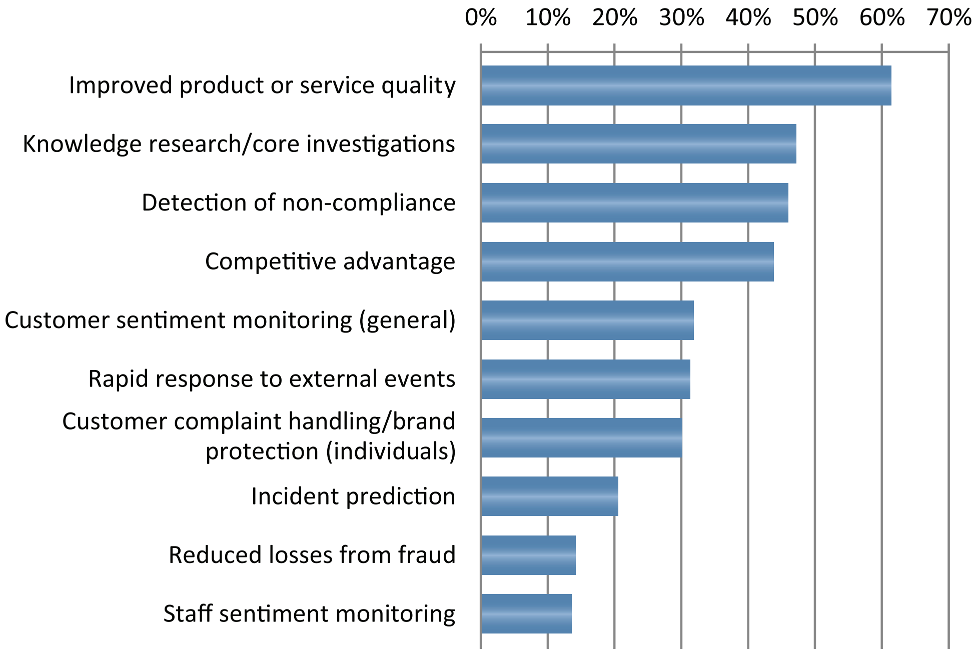
Figure 21: How are you monitoring external social streams (e.g. Twitter, LinkedIn, Facebook)? *(N=147, excl. 35 Don't Know)*



Business Advantage

Improved products or services comes out as the top benefit from business intelligence derived from content analytics, followed by core investigations and knowledge research. Detection of non-compliance rates highly, as do general customer sentiment monitoring, and individual customer complaint handling.

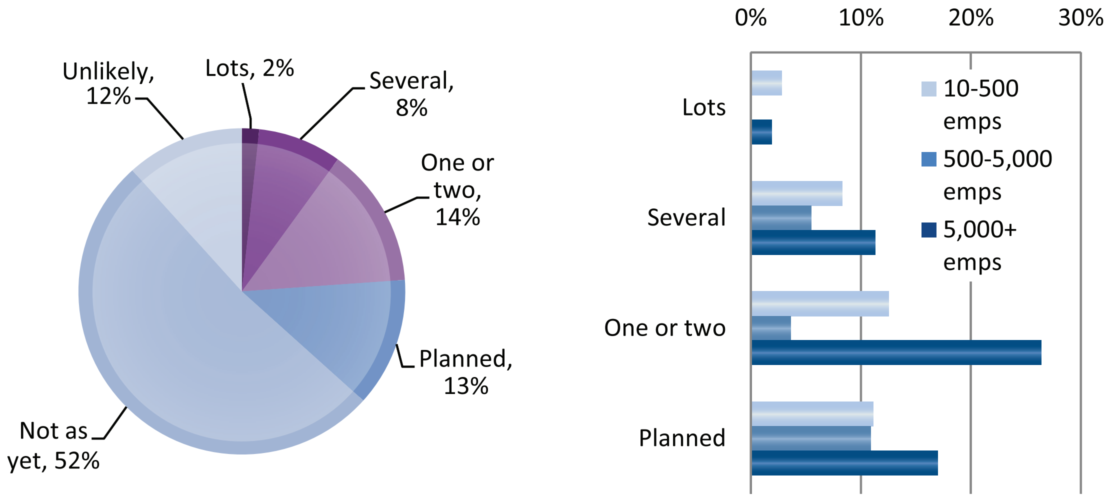
Figure 22: Which of the following business advantages would be the most useful to you based on intelligence derived from content analytics? (Max 4) *(N=176)*



Progress

As we indicated early on, around 25% of our respondents have active projects in the “business insight” category, with 10% having several. Across company sizes, the mid-sized businesses are lagging with only 9% active as yet, compared with 40% of the largest, and an encouraging 24% of the smallest, indicating a readiness to jump in with competitive advantage where possible, or in some cases, build a business on this.

Figure 23: Do you currently have one or more active “big content” or “content analytics” applications making use of unstructured or textual data for business insight? *(N=180)*



Mid-sized companies are falling behind in the take up of business insight projects involving content analytics, with only 1 in 10 having any active projects, compared with 1 in 4 of smaller organizations, and nearly half of larger ones.

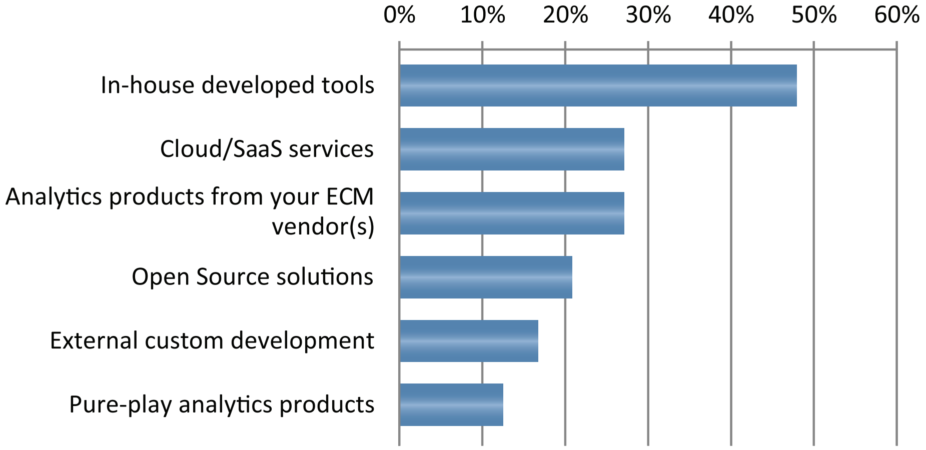
Big Content Projects

In seeking to characterize the projects being worked on, we asked which of the “three Vs” they involved – volume, velocity, variety. There is a fairly even split, with 11% involving volume and velocity, 36% high volume, 15% high velocity, 23% high variety, and 17% neither, but using complex techniques.

We also asked if the big content project involves a link to transactional or structured data, such as CRM systems, financial systems, data logs, etc. 53% are linked to one or more internal systems, and 5% are linked to external data sets.

When it comes to how the projects have been deployed, or what tools are being used, nearly half have used in-house development and 17% external custom (rising to 27% for the largest organizations). 27% are using cloud products, and 17% products from their ECM vendor, with 13% using analytics products from a pure-play vendor. 21% are using open source in some form, which is quite prevalent in this area.

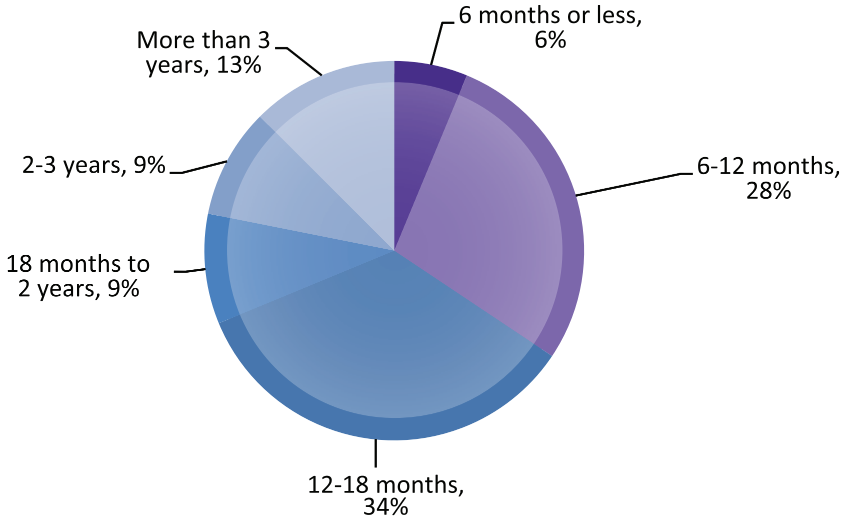
Figure 24: Are you using any of the following for your big content project(s)?
(N=48 with projects)



ROI

With any new technology, there are likely to be those who have latched on to it to solve a very specific problem, or to gain a big business advantage, and there will be others with over-ambitious plans, or who are hampered by lack of analytical skills. 34% of our respondents achieved a return on their investment in 12 months or less, and 68% in 18 months or less. This is a solid expectation of success, although from the 22% taking 2 years or more to show a return, we can infer that some projects will need a little longer to bed down and show a return.

Figure 25: How would you rate the ROI from your big content project(s)?
(N=32, excl. 13 "Not Measured" and 12 "Too Early to Say")

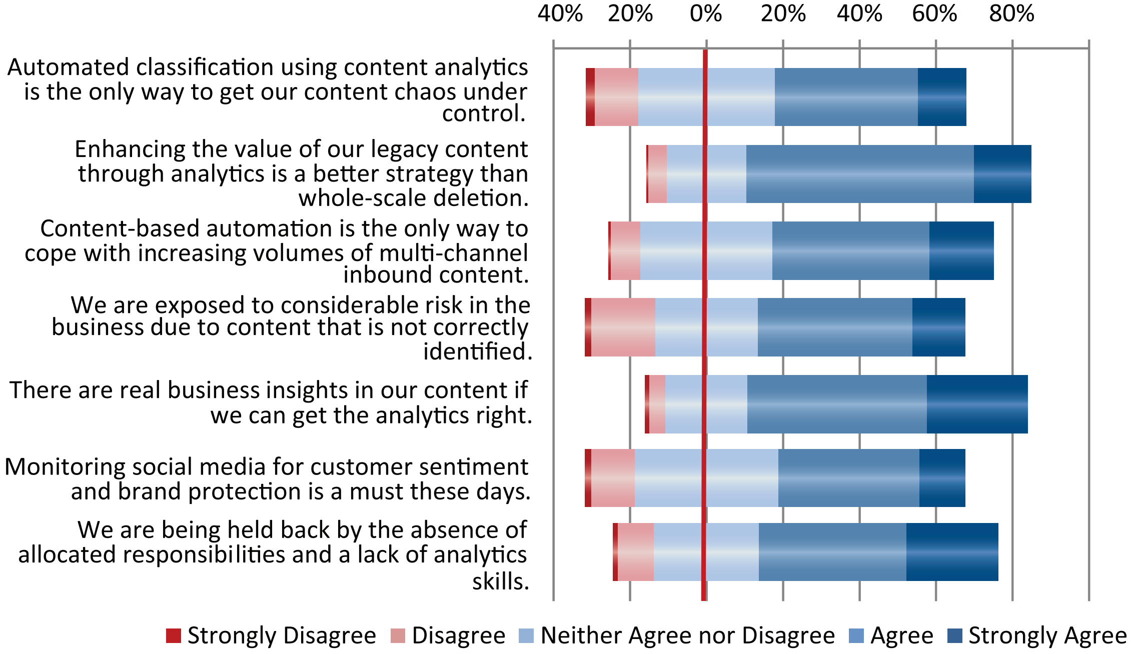


Opinions

Our "opinions" question is intended as a way to take the pulse of active practitioners, and those who are aware of the possibilities but may have more pragmatic issues to solve.

- 53% agree that auto-classification is the only way to get chaos under control.
- 75% agree that enhancing the value of legacy content is better than wholesale deletion.
- 73% know there are real business insights to be gained.
- 54% feel they are exposed to risk from non-identified content.
- 63% being held back by lack of skills and allocated authority.

Figure 28: How do you feel about the following statements? *(N=171)*

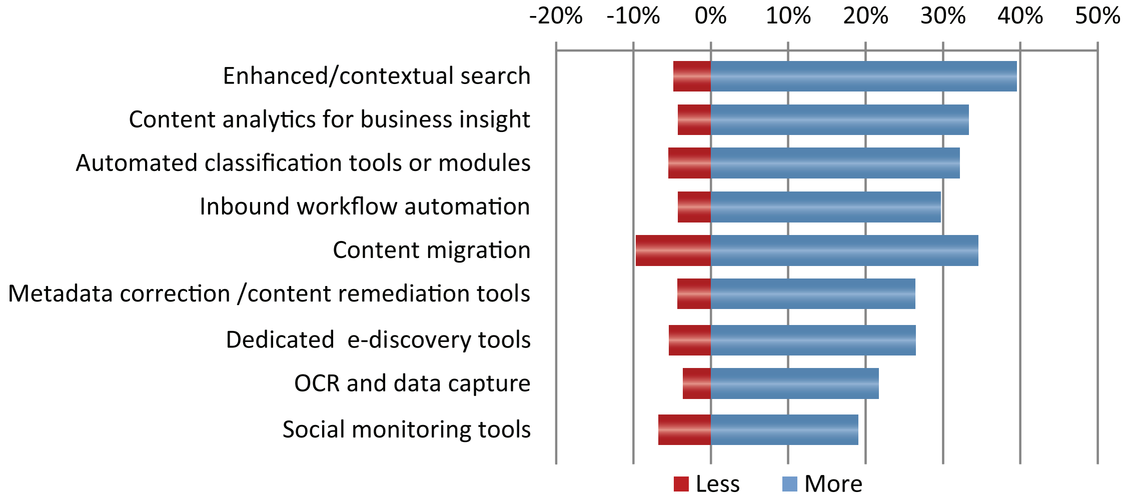


In summary, content analytics is generally considered to be a promising and useful technology, particularly as a way to increase content value and deal with increasing volumes of inbound content. For most, a lack of designated leadership and a shortfall of analytics skills is holding back exploitation of these new tools.

Spend

The indications are for growth in all areas, particularly enhanced/contextual search, analytics for business insight, and automated classification tools or modules. Inbound workflow automation shows demand as organizations build up their multi-channel inbound capabilities. Content migration tools have been buoyed by SharePoint 2007 to 2010 migrations, but are still showing strong growth for the 2010 to 2013 upgrade.

Figure 29: How do you think your organization's spending on the following areas and applications in the next 12 months will compare with what was actually spent in the last 12 months?
(N=168, excl. Same ~ 40%)



As we might expect with a new technology, growth forecasts are strong, as early adopters make way for more mainstream users, driven partly by the need to control content chaos, but also by the refinement of analytics tools and their ability to provide actionable business insight.

Conclusion and Recommendations

Content analytics is rightly taking its place amongst the corporate toolset, but while the business insight (or big data / big content) projects are still in something of an early-adopter phase, there are a number of other applications based on content analysis techniques that are already showing strong benefits in smoother workflows, improved search, and better compliance. We have seen increasing interest and adoption in recognition and routing of inbound content, automated classification of records and email, metadata addition and correction, and all of the improvements in access, security, de-duplication and retention that flow from this.

Staying on top of high volume, multi-channel, inbound content is increasingly difficult if relying on manual processes, and users are coming to accept that automated handling is as accurate but more consistent than humans. Email archiving in particular presents a dilemma, and content analytics offers a way to carry out defensible deletion in line with information governance polices. Dealing with dark data elsewhere in the business, and adding value to content rather than deleting it is a common objective.

Projects to derive business insight from content analytics are proceeding ahead, with 20% of our survey respondents already active, and a further 30% with plans. With some of these early projects coming on stream, 68% are reporting ROI within 18 months or less. Improving products or services is the top-rated benefit, followed by knowledge research or core investigations, and then improved compliance.

Recommendations

- If your content or records management deployment is stalled due to poor decisions early on regarding classification, metadata and taxonomies, or if you are migrating content from multiple repositories to a single system, take a look at metadata correction agents that can sort ROT from valuable content, and align content types and metadata.
- If you have access to contextual search, ensure that it is properly tuned, and that staff know how to use it. If you are reliant on more basic search, consider improving the searchability, and therefore the value of your content, by correcting and enhancing the metadata using analytic agents.
- Unless your staff are diligent and consistent at declaring, classifying and tagging records, consider providing auto-classification assistance or full auto-classification. Be aware that your information governance policies need to be updated and consistent as they will provide the rules for automated agents.
- Take control of your emails. If you have no archive, or the archive is “file and forget” you are losing potential corporate knowledge, but are also exposing the business to risk, and creating a potential e-discovery nightmare.
- Look at your retention policies as a way to control increasing storage requirements. Accurate metadata and enforced retention policies are the only way to limit storage, but will also improve your compliance and risk exposure.
- Inbound content handling can rapidly overload process staff, and reduce speed of response to customers. Implement a digital mailroom philosophy, and use automated recognition, routing and data extraction.
- Look across the range of your business activities to see where content analytics could provide business insight to understand customer needs, improve competitive advantage, help to solve cases and investigations, or prevent non-compliance and fraud.

References

1 “Connecting and Optimizing SharePoint” AIIM Industry Watch, January 2015. www.aiim.org/research

2 “Automating Information Governance – assuring compliance” AIIM Industry Watch, May 2014. www.aiim.org/research

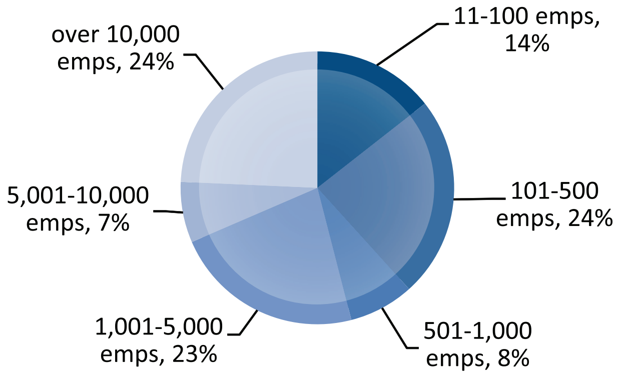
Appendix 1: Survey Demographics

Survey Background

The survey was taken by 238 individual members of the AIIM community between April 17 2015, and May 08, 2015 using a Web-based tool. Invitations to take the survey were sent via email to a selection of the 80,000 AIIM community members.

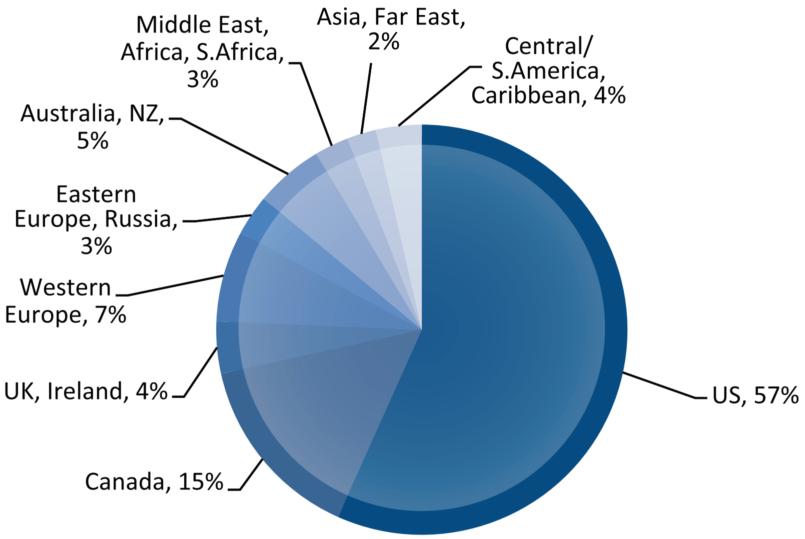
Organizational Size

Survey respondents represent organizations of all sizes. Larger organizations over 5,000 employees represent 31%, with mid-sized organizations of 500 to 5,000 employees at 31%. Small-to-mid sized organizations with 10 to 500 employees constitute 38%. Respondents from organizations with less than 10 employees have been eliminated from the results, taking the total to 222 respondents.



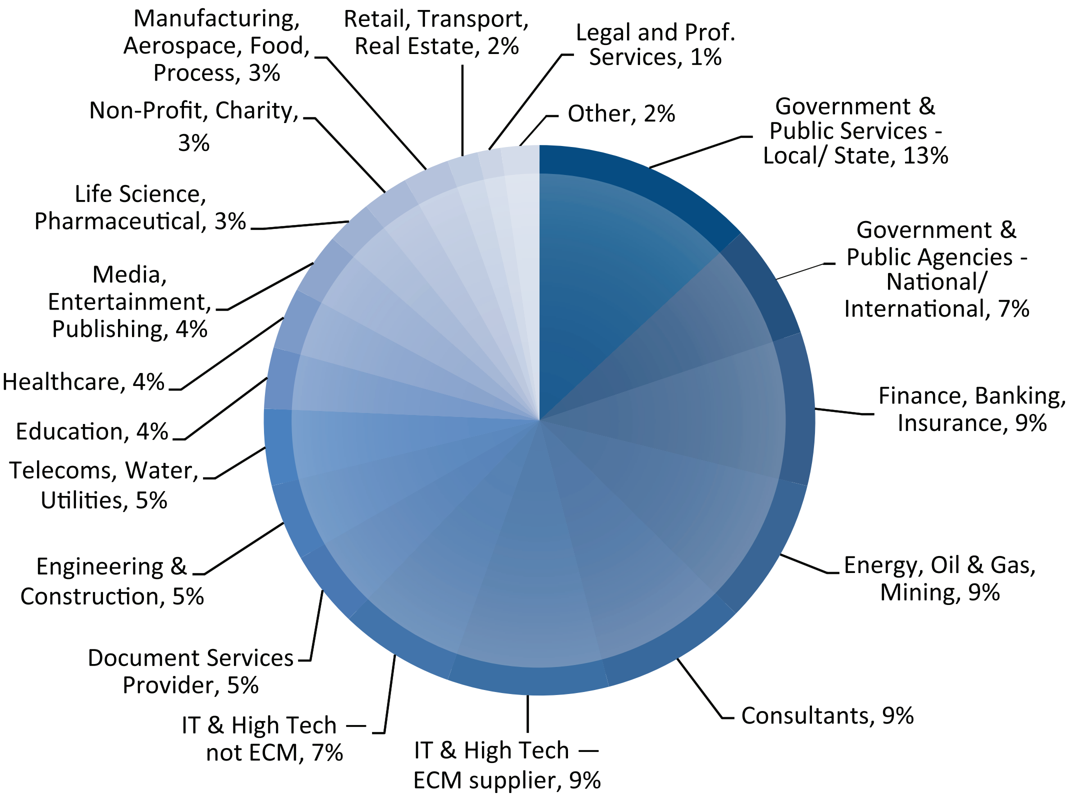
Geography

72% of the participants are based in North America, with 14% from Europe and 14% rest-of-world.



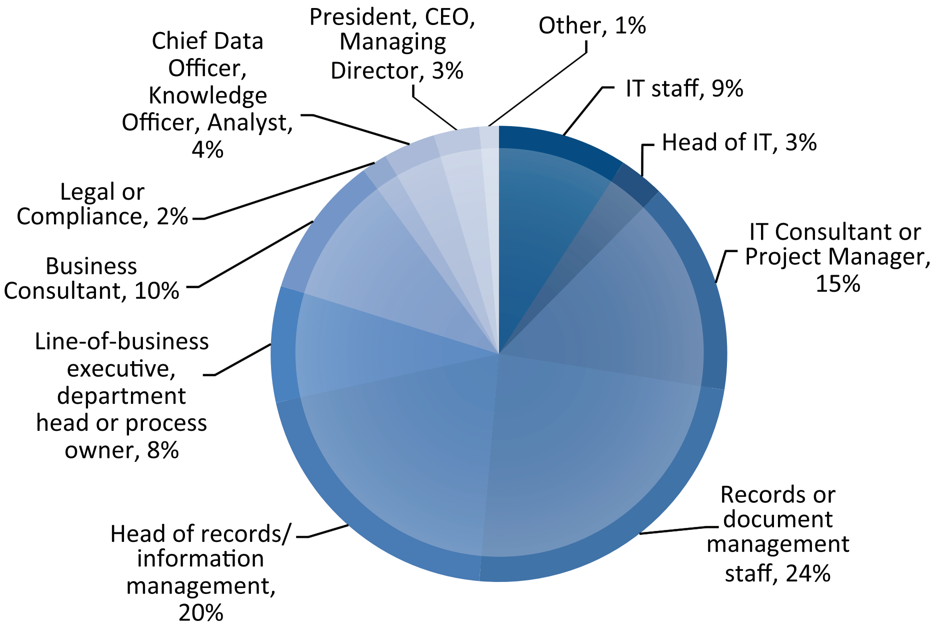
Industry Sector

Local and National Government together make up 20%, Finance and Insurance 9%, and Energy 9%. Suppliers of ECM services have been included as their responses are in alignment with other IT and High Tech. Other sectors are evenly split.



Job Roles

27% of respondents are from IT, 44% have a records management or information management role, and 21% are line-of-business managers or consultants.



Appendix 2: General Comments

Do you have any general comments to make about your content analytics projects? (Selective)

- This survey has shown how much I do not know about content analytics.
- Our organization will definitely benefit from content analytics, but we need to show some success to get management support.
- Our organization does not understand what that is. So when I bring it up they do not know how to respond other than “that would be nice”.
- We have only just started the two projects related to this, so although we may find that we have enhanced capability (e.g. content analytics for business insight) this is not one of the drivers, and don’t yet really know how much more we can achieve once the tools are in place.
- Remove the “human element” to establish consistency.
- Unfortunately it’s not applicable for small companies. But some thoughts brought by this survey are quite useful.
- Not enough attention is paid to unlocking unstructured content. Even a “simple” word doc can be very hard to understand/contextualize.... analysis is almost the ‘easy’ part... its the preparation / organization that is tricky.
- Hadoop is great and worth the cost.

UNDERWRITTEN IN PART BY

KOFAX from Lexmark

About Kofax

Kofax, a Lexmark company, is a leading provider of software to simplify and transform the First Mile™ of customer engagement. Success in the First Mile can dramatically improve the customer experience, greatly reduce operating costs and increase competitiveness, growth and profitability. Kofax software and solutions provide a rapid return on investment to more than 20,000 customers in financial services, insurance, government, healthcare, supply chain, business process outsourcing and other markets. Kofax delivers these through its direct sales and service organization, and a global network of more than 800 authorized partners in more than 75 countries throughout the Americas, EMEA and Asia Pacific.

For more information, visit www.kofax.com

www.kofax.com

OPENTEXT™

About OpenText

OpenText is the leader in Enterprise Information Management (EIM), providing comprehensive software solutions that enable organizations to achieve optimal information governance and transform their operations to succeed as a Digital Enterprise.

OpenText Content Analytics solutions leverage search, analytics, and discovery capabilities to:

- *Reduce the volume of information*
- *Break down siloes and integrate information across the enterprise*
- *Amplify the value of information through improved understanding, access, and collaboration*

The result is enriched business insight drawn from structured and unstructured information: new opportunities for growth, beneficial relationships, improved efficiencies and process speeds.

In addition, OpenText auto-classification combines Records Management with semantic capabilities for classification of content, eliminating the need to manually identify records and apply obligatory classifications. Organizations can demonstrate a transparent and defensible approach to classification based on statistically relevant sampling and quality control, minimizing the risk of regulatory fines and eDiscovery sanctions

Over 100,000 customers use OpenText solutions either on premises or in our cloud. We're helping enable a Digital World by simplifying, transforming, and accelerating the path to success as a Digital Enterprise. To learn more, please visit www.opentext.com

www.opentext.com

UNDERWRITTEN IN PART BY

Rocket®

Rocket Software

Rocket Software provides Enterprise Search and Text Analytics solutions that help users find the most accurate, relevant content they need to make smart decisions. Our integrated search platform gathers content from structured and unstructured sources, incorporates sophisticated indexing and analytic engines and combines powerful search capabilities to deliver an exceptional user experience.

Our Enterprise Search solutions are built with the IT Director in mind. No software or upfront development expertise is required for implementation. The solution comes with pre-built html user templates and pre-configured software enabling IT to implement the solution in 2-3 days, or less. Security and rights access can be customized, by user or at the document-level, to control access to sensitive and confidential information.

Our intelligent search engine uses semantic evaluation and advanced content analytics to understand the user's intent and the contextual meaning of the terms to display the most accurate and relevant results immediately. Built-in HTML 5 user interface templates feature a responsive web design providing a consistent user experience across all modern browsers and devices.

Rocket's Enterprise Search and Text Analytics team includes seasoned search and support professionals that work with you to understand your business and technology objectives from the onset and provide best practices guidance and white glove service throughout implementation and beyond.

www.rocketsoftware.com

SPS

a Swiss Post company

Swiss Post Solutions AG

Swiss Post Solutions, a division of Swiss Post, offers a comprehensive range of document and business process outsourcing services. With 7400 people working across Europe, North America and Asia and with access to an extensive partner network, we are able to support our clients across the globe.

Private and Public sector organizations have chosen to outsource their physical and digital document processing needs to us, utilizing our extensive knowledge of people-based outsourcing and our capability to deliver document processing services on, near or offshore. Our corporate information management system is a unified delivery platform that provides organizations with the ability to cost-effectively on-board and distributes documents throughout the organization. It provides our clients with the capability to:

- *Simultaneously improve productivity and reduce operational costs*
- *Take an enterprise-wide approach to automating business processes*
- *Enable improved decision making and customer satisfaction by accelerating business transactions*
- *Reduce the risk of non-compliance and achieving legislative and regulatory requirements*

Regardless of document type, physical or electronic medium, format, language or geographic location, Swiss Post Solutions offers an end-to-end solution from document creation to content management, production, distribution and business intelligence.

www.swisspostsolutions.com

AIIM Content Analytics Resource Centre

Learn how to combine content analytics, collaboration, governance and processes with anywhere, anytime access to deliver value to your customers, partners, and employees. That's what ECM -- and these best practices resources -- are all about.

PUBLICATIONS

Industry research reports, whitepapers, and toolkits

- **Valuable Content or ROT: Who Decides?**
- **IG Policy versus IG Reality - bringing your wild content under control**
- **Big Data and Content Analytics: measuring the ROI**
- **Toolkit: The A to B of Big Data**

PERSPECTIVES

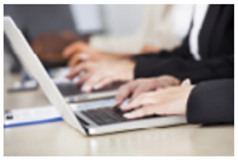
Community insights, opinions, and discussions

- **GTA**
- **Digital Conversion Strategies Applications & Benefits (Jun 22 2015, 7:00 PM - 8:00 PM (IST))**
- **Take the Terminology Challenge and Win! (Jun 11 2015, 11:30 AM - 1:00 PM (ET))**
- **The Fight Between the Cloud and On Premise Software**

WEBINARS

On-demand webcasts led by industry experts

- **Transform your Company through Modern Process Applications**
- **Transform your Company through Modern Process Applications**
- **The Future of ECM: new models for success**
- **Get "On the Case" for Better Business Outcomes**
- **My 2015 Predictions for Information Management: Who Will the Leaders Be?**
- **Webinar: Unleash the Power of Big Data**
- **Webinar: Big Data, Big Hype: Why the Business Should Care**



TUTORIALS

How-to videos developed by industry experts

- **How to Automate Metadata Collection and Classification**
- **How to Start Planning a Taxonomy**
- **How to Start Planning a Semantic Network, Ontology, or Topic Map**
- **How to Start Planning a Metadata Model**
- **How to Automate Records Identification**
- **How to Automate Records Disposition**
- **How to Automate Records Capture**
- **How to Start Planning a Content Model**

EVENTS

Upcoming conference, seminars, and webinars

TRAINING

Courses based on industry standards and best practices

www.aiim.org/Resource-Centers/Content-Analytics

aiim®
The Global Community of Information Professionals

AIIM (www.aiim.org) AIIM is the global community of information professionals. We provide the education, research and certification that information professionals need to manage and share information assets in an era of mobile, social, cloud and big data.

AIIM
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
+1 301.587.8202
www.aiim.org

AIIM Europe
The IT Centre, Lowesmoor Wharf
Worcester, WR1 2RR, UK
+44 (0)1905 727600
www.aiim.eu