Conquer GDPR by Leveraging Policy and Technology:

Intelligent Content Analytics to the Rescue

Authors: Jim Sinur and Jim Lundy

Summary

Enterprises face a looming deadline to become GDPR compliant. Many of the compliance issues are tied to information that is locked inside of documents. Content Analytics can assist with speeding up the process of ensuring GDPR compliance. This research note focuses on content even though many of the methods and approaches apply to all types of data.

Key Findings

Enterprises that conduct business in the European Union must be GDPR Compliant in 2018.

Prediction: By May 2018, Aragon Research estimates that only 55% of Enterprises will be GDPR Compliant.

Intelligent Content Analytics offers a path toward ensuring that GDPR Compliance can be achieved.
TABLE OF CONTENTS

Introduction…………………………………...3
Costs of Ignoring GDPR……………………3
Lawful Uses of Data…………………………4
Intelligent Content Analytics………………4
Key Selection Criteria to Evaluate………5
  Concept Identification………………….5
  Concept-Driven Aggregation………….6
  Event-Driven Aggregation…………….7
  Track and Compare…………………..7
GDPR in the Enterprise………………….9
  Contracts…………………………….9
  Finance, Legal and Procurement…..10
  HR and Employee Data………………10
  Sales and Marketing………………11
ICAP Vendor Selection………………….12
GDPR Capture……………………………12
Update Privacy……………………………14
Aragon Advisory……………………………16
Bottom Line………………………………16
Introduction

Enterprises face risks with the new General Data Protection Regulation (GDPR) effective May 2018. Preparation is needed to ensure that existing enterprises are GDPR compliant. This issue involves data, content management, and content analytics. This research note overviews the process of becoming GDPR compliant for incoming data and content, as well as for preexisting content.

GDPR will become effective in May 2018. This regulation will impose significant obligations on businesses that handle the personal data of EU residents. If organizations want to conduct business in the EU but are not GDPR-compliant, then they risk significant fines. GDPR applies to all organizations of any size inside or outside the EU. Even though countries are still working out the balance between collection and protection, GDPR will continue to march forward.

GDPR is mandatory in that organizations cannot keep data and content forever. GDPR also encourages better records management and strong information governance. The total cost or risk of data and information retention needs to be measured and managed. All the back up and historical data sources could be in the scope of any request for correct data usage. Organizations will be required to change or remove current or past data. The implications of this policy are huge.

Costs of Ignoring GDPR

Potential risks of non-compliance are costly, both tangibly and intangibly. The tangible costs may be upwards of 20 million Euros, or 4% of global gross revenue. These numbers are huge, so the cost of each event is high. The intangible cost is the distrust created by the misuse of personal data.

As mentioned above, with the deadline to be GDPR Compliant coming in May 2018, Aragon estimates that many enterprises will not be GDPR compliant by May 2018.

Prediction: By May 2018 only 55% of Enterprises will be GDPR Compliant.
Lawful Uses of Data

GDPR requires the protection of personal data, which is defined as any data or information that can directly or indirectly identify a person. This means that companies must know and protect complete data supply chains, including transfers. Therefore, organizations must have prior knowledge of all subject data location and its lineage. If consent is given, then the processing of the data is lawful, and one of six legally approved reasons must be present (see Note 1).

It is important to note that a number of reasons will be subject to interpretation. Item one, relative to consent, is crucial for marketers. Perusing a website (where permission is asked for and given) may yield a different result than signing up for ongoing information where formal data are provided and confirmed via unambiguous consent.

Content Focused Technologies that Deliver on GDPR

*Intelligent Content Analytics*

Content Analytics refers to the use of analytics to derive insights from content where the text or a higher-level abstraction of meaning – called a concept – has been organized in a model that can be mechanically processed. Content Analytics generally includes, but provides more value than, simple text analytics because Content Analytics can provide predictions and analysis based on these higher-level concepts. Intelligent Content Analytics Platforms (ICAPs) will allow enterprises to search and manage this content to drive better and more insightful decision making through predictive analytics. ICAPs will also allow business professionals to understand the content that is working in a business process, such as clauses that allow deals to close sooner versus ones that get rejected (see Figure 1).

Note 1: Six Legal Reasons for Data Collection that Conform to GDPR.

1. The data subject has given consent to the processing of his or her personal data.
2. Processing is necessary for performance of a contract.
3. Processing is necessary under legal obligation to which the controller is subject.
4. Processing is necessary to protect the vital interest of the data subject or another person.
5. Processing is necessary for public interest or vested authority of the controlling organization.
6. Processing is necessary for the legitimate interests of the controller or a third party.
Content Analytics could identify potential GDPR violations automatically and route situations that would require human judgment to resolve non-compliance situations to GDPR savvy folks. The amount of time and money saved with this automated approach would more than pay for the investment in Content Analytics.

**ICA and GDPR Providers - Key Selection Criteria to Evaluate**

Aragon has identified four key functional areas that comprise an Intelligent Content Analytics Platform that could help in GDPR efforts. Note that there will be a variety of Content Analytics Solutions that will come to market to help solve the GDPR problem.

1. **Concept Identification**

   An Intelligent Content Analytics Platform must support the identification of key concepts contained within every document, file, and object to be analyzed. Concepts must
include any terms that have specialized meanings or significance within the domains (for example, concepts related to standard clauses in real estate contracts, or terms related to protected classes in HR documents). Ideally, these concepts will be available to the user in an easily traversable, hierarchical concept map. The key shift in ICAP is that concept identification can happen in real-time. The element of real-time means that actions can instantly be suggested which gives the enterprise an advantage.

Intelligent Content Analytics (ICA) will change the game due to the fact that the system is proactively extracting, classifying, and presenting critical insights to the user. So, in reality, the need for search is significantly reduced because business users will automatically deliver the analytics and suggested actions to the user.

Determining Personally Identifiable Information

In the course of doing business, business units capture data that contains PII and they need to recognize those procedures and ensure they are compliant.

Identifying those business units and their processes that capture PII data is the first step. ICA, in conjunction with capture tools, will be able to auto-classify the data and then apply a record policy to that data to help meet regulations.

If there is PII data that does not conform to the established policy, an automatic alert could be generated relative to the PII being identified. One of the providers that can help in this step of GDPR compliance is ABBYY (see Note 2).

2. Concept-Driven Content Aggregation

All content within the enterprise should be accessible in aggregate form based on concepts. ICAPs that normalize these concepts across input source types will be able to connect to existing content repositories to identify the patterns in existing content.

For disparate source types, this method of analysis would be impossible without a common representation dependent upon concept normalization. This suggests that the development of a basic taxonomy will help to speed up the classification of GDPR related content. We’d note that more advanced

Note 2: ABBYY at a Glance

ABBYY focuses on Intelligent Capture and Content Analytics. ABBYY Solutions help to reduce risk, accelerate decision-making, and drive faster time to revenue.

Locations: Milpitas, CA; Kiev; Munich; Moscow; and multiple other locations.

CEO: Ulf Persson

Key Offerings: Comparator, FineReader, FlexiCapture, FlexiCapture for Invoices, FlexiCapture for Mailrooms, Recognition Server.

Availability: Available now

Website: www.abbyy.com
Content Analytics offerings will be able to generate a taxonomy automatically in the future.

3. Event-Driven Aggregation

An ICAP can identify all content relevant to a specific internal or external event, such as a merger or a sale to a specific customer. Event-driven aggregation can happen in real-time so the enterprise has a view of things as they unfold (e.g. fraud) and can react immediately.

4. Track & Compare

An ICAP should enable a user to track changes to a unit of content and assess versions or different content files to compare them with respect to specific concepts or clauses.

Records Management

Records Management gains a new use case with the advent of GDPR. Records Management has two primary components: to uphold developing retention policies that conform to business laws, and to not violate GDPR throughout the course of business.

Most enterprises already have Records Management as part of an existing ECM Platform. However, given the criticality of GDPR, an assessment should be made if the right licensing is in place to support GDPR, or if a new offering is needed to meet the needs of multiple business units. In most cases, existing RM solutions will meet the needs of the business, as long as RM policies can be updated or developed to be GDPR compliant.

RM Retention Policies

Retention policies enable documents that have PII data to be deleted after the mandatory retention timeframe has expired. Enterprises need to realize that they may need to provide their clients with the right to ensure that their PII data will be deleted after the government-mandated retention period.
The Content Lifecycle and GDPR – New and Existing Content Accessible By Business Units

Enterprises must have an understanding of data that are collected by different business units in the enterprise.

Content Audit

When first embarking on a course for becoming GDPR compliant, existing contracts will need to be audited. Ideally, a solution for ongoing Content Analytics will be used. However, in some cases, there may need to be a separate initiative to ensure that an audit is done in a timely fashion.

Content Capture and PII

In the course of doing business, business units capture data that contain PII and they need to recognize those procedures and ensure they are compliant.

Identifying those business units and the processes that capture PII data is the first step. Using Capture tools to classify and then apply a record policy to the data will help meet the policies. One of the providers that can help in this step of GDPR compliance is ABBYY (see Note 2).

Making Existing Content GDPR Compliant

For current processes that have been utilized for years, enterprises will need to identify the content repositories, i.e. the applications that contain data and could be at risk of being non-GDPR compliant.

Outlined below are the processes for existing data compliance and applying policies so that new or updated documents will be GDPR Complaint.
GDPR in the Enterprise

GDPR will touch nearly every business unit in the enterprise and nearly every contract, including financial, legal and procurement contracts. For GDPR, ICAP offerings are ideal for ensuring GDPR Compliance. Some of the critical documents and business units to evaluate for GDPR compliance include:

1. Contracts

Today, there is tremendous value inside of contracts, which are usually maintained in an ECM or Contract Lifecycle Management system. These legacy documents contain GDPR data that need to be identified for legacy documents and for all new contracts. Because Content Analytics has built-in intelligence, the enterprise can gain insight into revenue growth opportunities, and compliance and fraud risks. The net is that insights contained in these contracts can be unleashed by an ICAP.

As mentioned above, an audit of existing contracts needs to be completed so that the enterprise can certify that it is compliant. This will need to be done on both electronic and hardcopy documents and records.

The key things that ICA can enable for contracts include:

- Creating a central contract repository to facilitate search and analysis while reducing risk
- Discovering domain-specific terms and provisions to enable accurate, consistent, and fast views into contract data, including a GDPR view
- Normalizing concepts across input types, aiding discovery of associations that would otherwise be impossible
- Managing extraction policies to discover contact information specific to an organization, event, or need
- Reducing the time and cost associated with finding, managing, and reviewing contracts, provisions, and terms
- Identifying PII Data that might be non compliant
2. Finance, Legal and Procurement

Finance teams will be most responsible for contract compliance, but legal teams also need to ensure that any client data they retain is GDPR compliant.

Most importantly, procurement teams, which often report to either supply chain or finance, deal with a significant amount of data pertaining to suppliers and contractors. This often involves data related to the people at those suppliers. Today, the need to be compliant and avoid risk is critical. Intelligent Content Analytics can help ensure compliance and also identify finance, legal, and procurement documents that might make the enterprise noncompliant. Key benefits for finance, legal and procurement include:

- Ensuring compliance with reporting language requirements from various reporting agencies
- Tracking and comparing filing documents from internal and public sources
- Gaining insights by tracking filing clauses with auditing results
- Understanding documents that might put the enterprise at risk and suggest preventative actions

3. HR and Employee Data

Human Resources departments will be significantly impacted by GDPR in part because they have a significant amount of PII data for their employees and contractors.

Current employee records need to be compliant, and HR and talent management providers need to provide a Data Processing Addendum (see Note 3).

Enterprises also need to validate that their providers are masking PII data fields and encrypting the data (see GDPR Security and Notifications Section).

For HR data outside current systems—such as information relative to former employees—retention rules apply. Records Management Policies should be updated to delete that data.

Note 3: A Data Processing Addendum

A data processing addendum needs to be developed and included for any firm that offers a Cloud Service that processes data or content for a client. This is to notify what the provider is doing to show proper GDPR Compliance.

The elements of this agreement should include:

1. Describe eligible customers impacted by this DPA.
2. Describe how to execute the DPA.
3. Clearly provide the definitions of different terms (e.g. Adequacy, Controller, Customer Personal Data, Data Controller Affiliates).
4. Roles and responsibilities of processing customer data.
5. Data security policy. Describe how the provider will perform data security.
6. Incident notifications. Describe how customers will be notified if there is an incident.
7. Deletion and return of customer data. Describe procedures available to delete customer data and/or services available to do this.
8. Describe the process for audits and reports.
4. Sales and Marketing

Marketing teams strive to support their sales counterparts by filling the pipeline with qualified leads.

As Content Analytics comes online for marketers, we expect that ICAPs will be able to help marketers understand where there is client data that needs to be managed and protected.

Removing unverified and/or old contacts from both Marketing Automation and CRM systems is important. Additionally, verifying that the capture of personal and enterprise PII data is volunteered by users and also confirmed (via single or double verification) is important to ensure compliance.

The systems that need to be GDPR compliant and also have Data Processing Addendums signed off include:

- Customer Relationship Management (CRM)
- Marketing Automation Platforms
- Advertising Platforms
- Websites

Privacy Policies are covered below, but websites and customer portals need to be updated to include GDPR language.
Preparing for GDPR – Selection, Deployment, Policy, Training and Notifications

GDPR compliance means that processes and policies need to be updated, and business users and managers need to be trained. The competitive nature of the race to the digital enterprise and the need to be GDPR compliant underscore the urgency to deploy an ICAP Solution so that the enterprise is not at risk.

The key steps that need to be taken include:

- ICAP Selection
- ICAP Deployment
- Document Inventory
- Capture, Analytics, and Records Management
- Update Policies
- GDPR Security and Notifications

ICAP Vendor Selection

In addition to the general advice regarding vendor history, financials, and customer satisfaction, enterprises should keep the following in mind when evaluating ICAP vendors:

- Look for platforms that demonstrate experience and excellence within one or more of the key content domains that are critical to your enterprise, and that offer a consistent user experience within the enterprise and across the ecosystem.

- Evaluate vendors using the actual requirements for your organization rather than generic checklists.

- While some requirements will be present for all buyers – like the need to analyze contracts and regulatory documents – it is important to prioritize your needs for special requirements, like sales content analytics, to ensure that your priorities align with those of the vendor.

- Vision/Product Plan: No vendor is likely to offer intelligent analytics with pre-constructed concept maps for every type of content and every domain of interest to your enterprise. Make sure planned releases
align with your anticipated needs in the next 36-60 months.

Implement GDPR Capture, Analytics, and Records Management

There are multiple ways to implement GDPR. Some approaches by best of breed providers are really only targeted at assessment, not ongoing day-to-day management.

Aragon feels that ICAP offerings will help with solving issues in cross content management repositories. This includes having real-time reports that show any discrepancies in compliance.

• Initial capture needs to conform to one of the six legal capture criteria. Developing methods of flagging capture procedures that don’t conform will be key.
• Records Management will be key to update so that retention schedules are updated and managed by current GDPR.
• Real-time alerts need to be configured so that IT staff and business units are notified of non-compliance. Mitigation needs to take place and then reporting to government agencies needs to occur in a timely manner.

Content and Document Inventory

Understanding the types of documents that exist in each business unit is key to understanding where the ICAP rollout should begin. The ideal business units to use for the first phase of an ICAP implementation project are ones that have critical documents. A Content Analytics Platform can help to auto create the inventory, assuming that indexing can be done across multiple content repositories.

Visualization

Creating a visual catalog will be an essential deliverable to consider in creating a consent management for both content and data. The catalog and search functionality could help organizations find and summarize all relevant information of a
citizen on digitized paper. This could be disposed of after a 30-day window, once citizens are informed.

**Update Privacy and Data Processing Policies**

All company policies need to be updated to be GDPR compliant. This is one of the most critical parts to ensure that the enterprise is compliant with GDPR. Legal departments will play a major role in ensuring that policies conform to the GDPR Guidelines. Most importantly, this includes publicly available privacy policies.

The checklist for policies includes:

1. Draft updates of all documents that pertain to use of Data. This includes policies, contracts and specific privacy policies
2. Legal review of updated policies
3. Changes and final review of all policy documents
4. Publish final updated policy and privacy documents

For Technology Providers, a GDPR data processing addendum needs to be included for existing customers (see Note 3).

**Train Business Units Users**

All associates that are involved with customers need to understand GDPR and how their enterprise will address the regulations. This means basic training for managers so that they understand the risks of being non-compliant.

Many regulated industries will require the completion of this training. All enterprises that do business in Europe should invest in training.

**GDPR Security and Notifications**

Document and data security is key to ensuring GDPR compliance. Security breaches are one of the top ways that PII information is stolen and then misused. Enterprises need to ensure that they are GDPR compliant and that they *implement appropriate technical and organizational measures.*
Enterprises need to ensure that their overall IT security is current and that content and data are part of a larger Disaster Backup and Recovery initiative (DBAR).

Key security factors tied to GDPR include:

- **PII Masking and Encryption:** The pseudonymization and encryption of personal data. This should be expected, but today many firms do not mask and encrypt data.

- **Ongoing Processes:** Ensuring that the ongoing processes maintain confidentiality, integrity, resilience and availability of processing systems and services.

- **Backup and Restoration:** The ability to restore personal data and the access to it in a timely manner in the event of a physical or technical incident.

Enterprises need to ensure that their overall IT security is current and that content and data are part of a larger Disaster Backup and Recovery initiative (DBAR).

- **Ongoing Security Testing and Assurance:** Ensuring that there is a consistent process for testing, assessing, and evaluating the effectiveness of technical and enterprise measures for ensuring the security of processing.

When breaches do occur, processes need to be in place internally and inform government and customers that a breach has occurred. Notifying the parties that are affected must occur within 72 hours. Again, for many firms involved with processing credit cards, this is a similar notification process. For firms that are new to GDPR, this notification process will need to be established.
Aragon Advisory

- Enterprises should not wait to develop a GDPR compliance plan, even though there is little on the ground experience with GDPR.
- Given timeframes, enterprises may want to have a GDPR audit done to find critical exposed business units.
- Enterprises have to initiate training for their staff, as personal identity data and content require new safeguards.
- Enterprises should prioritize sources that need the most help. Likely, marketing content and data will rise in priority for most enterprises. Sales data will likely be another critical focus.
- While GDPR is a new approach to guarding individual data and content, it is a move that promotes personal control. The obvious long-term solution is shared digital identity with personal permissions for each use.
- Enterprises need to look at their content analytics as a new way to understand and evaluate their business.
- Enterprises need to deploy contract analytics as part of an overall digital business strategy.

Bottom Line

Keep in mind that GDPR is new, and it is just one step in the journey towards identity management. GDPR gives protection to individuals, and Content Analytics is one of the best ways to ensure GDPR compliance. Given the looming 2018 deadlines, enterprises should deploy a Content Analytics offering to speed up the process of ensuring compliance and avoid what could be hefty fines. Also, because so much information is locked inside documents and contracts, enterprises need to begin to leverage content analytics as part of a digital business strategy.