

IBM StoredIQ and GDPR

Accelerate your assessment
of unstructured data sources



What is GDPR?

The General Data Protection Regulation (GDPR) is part of European Union (EU) law that addresses data protection and privacy. Since the regulation went into effect on 25 May 2018, organizations that collect, retain or process the personal data of EU data subjects must follow GDPR requirements. The GDPR is meant to streamline data protection regulations for EU countries and those within the European Economic Area, while addressing the global export of personal data concerning EU data subjects. The regulation may present a major shift in your treatment of personal data.

Under the GDPR, individuals have the **right to choose** what happens with their data, including what information is used, who uses it, and how it's used. Your customers may ask your organization to edit, delete, export or refrain from processing their personal data, and your organization is required to satisfy such requests.

Organizations need content analytics tools to help transform their data, as failure to comply with the GDPR can have steep consequences. Some organizations have struggled to identify and develop readiness solutions, but there are basic steps you can take to begin the process.

GDPR readiness

What can your organization do to prepare?

First, assess the data sources in your organization. Once all personal data sources are identified and inventoried, data managers can evaluate the gap between how personal data is being managed versus how it must be managed according to GDPR standards. To close the gap, your organization may need to change where data is stored, how data moves throughout your system, or how employees respond to customer requests and data breaches. These tasks can seem overwhelming given the vast amounts of data companies retain. Fortunately, **IBM® solutions can help.**

Accelerate your GDPR assessment with IBM StoredIQ

IBM StoredIQ® addresses the issues that can affect your records management by assessing unstructured data sources. This in-depth analysis gives you a clearer sense of where data resides so you can make better informed business decisions. Here's a closer look at how StoredIQ works:

Understand what data resides in your data sources

- Discover and sort data using reports and data maps based on metadata
- Identify hotspots of personal data, potential compliance issues and violations
- **Locate personal and sensitive information** using powerful search capabilities with predefined, customizable patterns

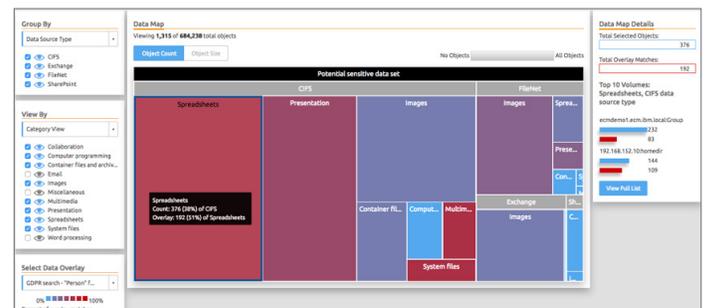


Figure 1. The IBM StoredIQ data map helps you visualize hotspots of personal data.

Simplify data management and analysis

- Locate personal and sensitive information using powerful search capabilities with predefined, customizable patterns
- Preview documents from search results using the highlight feature to validate matches
- Create reports by term for files that match specific patterns or distribution rules
- Learn whether or not documents with sensitive content have been stored in unexpected places

Create a central catalog with governance rules and policies for GDPR

- Integrate StoredIQ with [IBM InfoSphere® Information Governance Catalog](#) (IGC) to sync business terms and rules with your assessment
- Use IGC for unstructured data sources and IBM InfoSphere Information Analyzer for structured data sources
- Populate the results of your assessment into your central catalog

Generate multiple reports to inform decision making

- Create CSV reports showing file details that match searched phrases
- Pull management reports to show high-level distribution and the amount of personal data
- Configure sensitivity levels in each data class and share reports throughout the organization

Support data subject requests

- Use predefined filters in StoredIQ to identify personal data related to a person in connected unstructured data sources
- Use predefined filters to identify and delete selected data when individuals request that their data be removed

StoredIQ supports many different unstructured data sources, including SharePoint, email servers and file shares. For a complete list, [visit the IBM Knowledge Center](#).

How StoredIQ identifies personal data

StoredIQ reads the data from selected unstructured data sources and creates an index identified as metadata and fulltext. The documents remain in place—no data moves to a repository. The index can be enriched with tags to identify personal data.

Analytic plug-ins, also known as [cartridges](#), identify personal data and enrich the index with the findings. These cartridges can contain analysis logic based on different technologies ranging from simple regular expressions to full-blown cognitive approaches like natural language processing (NLP) and machine learning. A cartridge in StoredIQ is designed to determine insights like names, numbers or addresses and makes those insights searchable.

IBM delivers sample cartridges for GDPR

GDPR-focused data discovery cartridges are designed to assist organizations in their efforts to prepare for the GDPR and other data privacy regulations. IBM delivers two StoredIQ GDPR-focused cartridges.

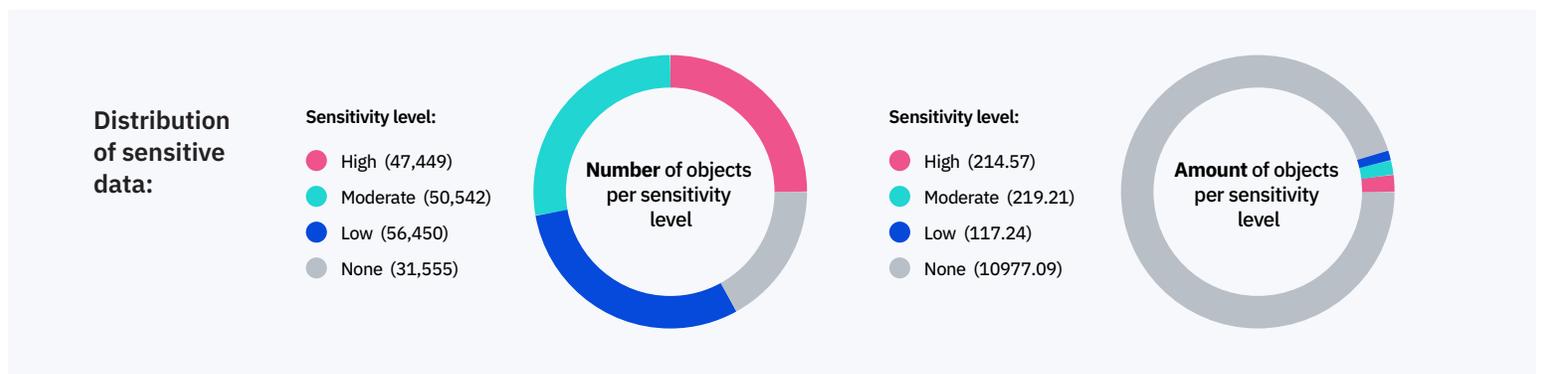


Figure 2. Representation of data privacy report

GDPR-focused basic cartridge

Based on regular expressions for all European countries in which the following personal data types are supported by default:

- Bank account numbers, international IBAN and national versions
- National identification numbers such as tax ID, national ID card or passport number
- Phone numbers, email addresses and IP addresses

GDPR-focused advanced cartridge

Based on NLP or SystemT-based patterns, in which the following personal data types are supported:

- Person names
- Locations
- Organizations
- Addresses
- Dates

Both cartridges can be customized and extended based on your requirements.

For more information on the cartridge concept and how to create cartridges based on your requirements, visit this [IBM developerWorks tutorial](#).

StoredIQ also has predefined macros with additional personal data types to support preparedness for other data privacy regulations:

- Credit card numbers
- US Social Security numbers, which can include dashes and numbers
- US phone numbers, which can include dashes, dots, parentheses, numbers and letters; this macro can filter for mnemonics that use letters to market phone numbers
- IP addresses, which can include dots, numbers and letters
- Email addresses, which can include symbols like @, dots, numbers and letters

Additionally, you can create a filter using word lists. This approach is useful when searching for sensitive personal data.

StoredIQ Insights

StoredIQ Insights offers an interactive search with instantaneously updated counts and percentages in the filter facets. The resulting list is updated dynamically when you type a search string or when you pick a facet from the library. The search criteria is highlighted in the preview of each document. Indexed annotations created with GDPR-focused cartridges are available in the facet pane to filter in or filter out. This feature can help you search for personal data and validate results. For IBM StoredIQ production information, visit the [IBM StoredIQ Marketplace page](#).

The screenshot shows the StoredIQ Insights user interface. On the left, there is a 'Facets' pane with several expandable sections: 'Indexed annotation' (with sub-facets like is.PhoneNumber, is.EmailAddress, is.IPAddress, is.NationalIDNumber, is.BankAccountNumber, is.NationalTaxNumber), 'Category' (with sub-facets like Email, Word processing, Miscellaneous, System files, Presentation), 'File name extension', 'Object size', 'Created date' (with sub-facets like Over 3 years ago, 1 to 3 years ago), 'Modified date', 'Volume', and 'Path'. The main area displays search results for 'personal data'. At the top, there are search filters for indexed annotations: is.PhoneNumber, is.NationalIDNumber, is.NationalTaxNumber, is.NationalSocialSecurityNumber, and is.PassportNumber. Below this, a table shows 83 results. The table has columns for Object name, Size, Volume, Path, Created, and Last modified. The first few rows show documents with titles like 'SUPPLEMENTAL View...', 'Per: Are You Living in...', 'Your mailbox is over L...', and 'Your mailbox is over L...'. The last row shows a PDF document titled 'id-039-0348.pdf' with a size of 21.09 KB, created on 2016-03-10 4:45 PM, and last modified on 2010-02-05 9:38 AM.

Figure 3. StoredIQ Insights user interface

IBM Analytics Unified Governance & Integration

For more information

IBM StoredIQ is part of IBM data governance solutions. These solutions can help you understand your data, whether structured or unstructured, and govern it to help you mitigate your compliance risks and empower your information stakeholders. To learn more about IBM data governance solutions visit ibm.com/analytics/data-governance.

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. For more information, visit ibm.com/financing.

© Copyright IBM Corporation 2019

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
March 2019

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Important information: Clients are responsible for ensuring their own compliance with various laws and regulations, including the European Union General Data Protection Regulation. Clients are solely responsible for obtaining advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulations that may affect the clients’ business and any actions the clients may need to take to comply with such laws and regulations. The products, services, and other capabilities described herein are not suitable for all client situations and may have restricted availability. IBM does not provide legal, accounting or auditing advice or represent or warrant that its services or products will ensure that clients are in compliance with any law or regulation.



Please Recycle