# IBM Data Governance solutions

Andrzej Wróbel
IBM Analytcis
Kraków, Poland
andrzej.wrobel@pl.ibm.com

Konrad Komnata
IBM Analytics
Kraków, Poland
kkomnata@pl.ibm.com

Krzysztof Rudek
IBM Analytics
Kraków, Poland
krudek@pl.ibm.com

*This paper covers aspects of governing information data on enterprise level using IBM solutions. In particular it focus on one of the key elements of governance – data lineage for EU GDPR regulations.*

## I. Introduction

IBM Information Governance Catalog enhances an organizations ability to build and maintain a strong data governance and stewardship program that can turn data into trusted information. Information Governance is becoming more important and is receiving more attendance mostly from financial institutions (Fig. 1). This trusted information can be leveraged in various information integration and governance projects, including big data integration, lifecycle management, and security and privacy initiatives. By leveraging the comprehensive capabilities in Information Governance Catalog, you are better able to align IT with your business goals [1].


Figure 1 – Governance metadata infographic [7]

## II. What is Information Governance?

Information Governance is a phrase that describes way, how organization is managing all the data and information that is going through all its processes [2]. Governance is not one-time project, but constant program. "Information governance is the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information. It includes the processes, roles, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals [3]." According to Generally Accepted Recordkeeping Principles [4], linking duty with value to information assets should lead to more efficient and effective management. Diagram (Fig. 2) picturing Information Governance Reference Model shows how information should flow during Governance process.

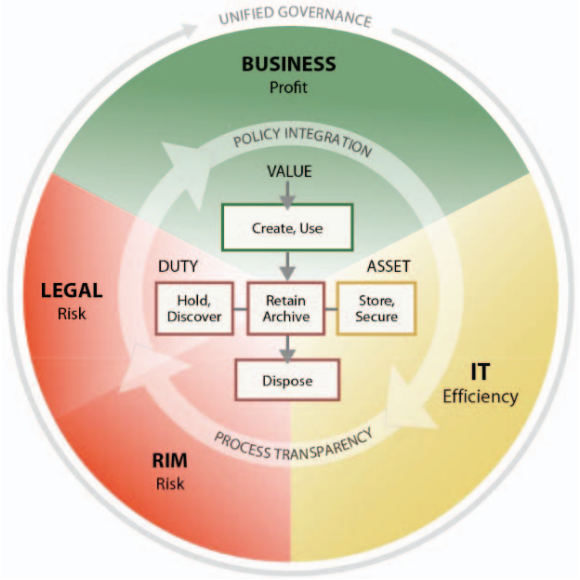


Figure 2 – Information Governance Reference Model [4]

Organizations are subject to regulatory requirements. In particular European Union General Data Protection Regulation (GDPR) seeks for data protection law framework across the EU and primarily focus on giving citizens back the control of their personal data anywhere in the world. IBM Data lineage is one of the key elements that help to achieve full compliance with EU regulations. Having full knowledge about data assets and information on data flows is critical to meet GDPR requirements.

978-1-5386-2366-4/17/$31.00 @2017 IEEE

Authorized licensed use limited to: UNIVERSIDAD DE GRANADA. Downloaded on January 14,2025 at 16:49:47 UTC from IEEE Xplore. Restrictions apply.

EU General Data Protection Regulation (GDPR) is “most important change in data privacy in 20 years” [5]. One of key additions to data protection are introductions of Breach Notification, Right to be Forgotten, Data Portability, Privacy by Design and Data Protection Officers [5].
To be compliant with all those expectations, organizations need to introduce Information Governance solutions, that will have necessary functionalities available for users.

There are many known solutions for Information Governance on the market. Among them, only few can be considered as good choice for organizations that need to be compliant with GDPR, for example SAP Data Governance, Collibra Data Governance Center and IBM InfoSphere Information Governance Catalog [6]. All that software differ from each other, but couple of functionalities are common: glossary and possibility of generating data lineage. What differs them are additional possibilities provided by vendors. In this field IBM wins with its InfoSphere Information Governance Catalog (IGC). First of all it is fully compatible with all IBM offerings, which makes it natural choice for companies that use at least one IBM tool for their data. Despite of IBM compatibility, IGC offers “various levels of sophistication and interconnectivity options [6].” Some companies are highly regulated – like financial institutions. For them IBM has prepared “workflow and day-to-day administration of high-visibility data governance processes [6].”

## III. Power of Integrated Metadata

IBM Information Governance Catalog outlines the complete lineage of fields from applications, reports or data warehouses back to source systems all in a single user interface. Knowing the origin of data and what happens to it as it moves through numerous systems is fundamental to establish confidence in the completeness and correctness of that data. Data validation and traceability is a critical factor in complying with regulations that require companies to confirm the veracity of their data.

Information governance initiatives address enterprise concerns about the quality and reliability of information in particular it answers following questions:
- Where does the data come from ?
- Where does the data flow to ?
- Is the data up to date ?
- Who owns this data ?
- Can I trust this data ?

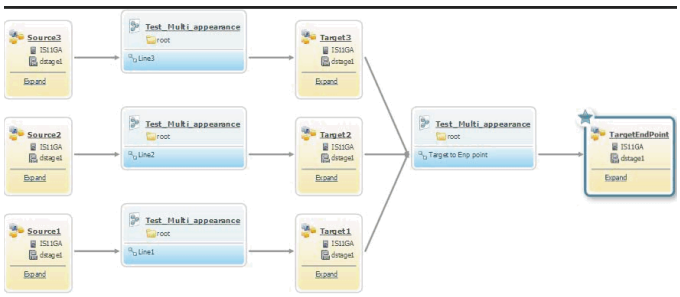

Figure 3 – Data Lineage in IGC

In order to build the flow of the meta-data IBM implemented following approach:

- Collecting meta-data from source objects like databases or any object that is providing meta-data and participates in the data flow.
- Normalization of incoming meta-data
- Reconciliation of incoming meta-data with existing objects or providing placeholder information for unknown objects
- Data traversal algorithm to connect meta-data based on multiple criteria.
- Enrichment of the graph through additional actions based on the user knowledge
- Validation of the graph
- Constant update of the graph through monitoring of the meta-data changes

Example of such flow is pictured on Figure 3 – it is direct screen shot from newest version of IBM InfoSphere Information Governance Catalog.
Above steps are well proven in the field delivering robust solution for enterprise based customers. One of the common use cases where lineage supports business decision is in logistics domain. In the past when information was controlled manually and was on the much smaller scale the confidence in data coming from reports was high. However with huge number of data being created every day, much more complex environments that are difficult to be completely controlled by human it's also difficult to say that data we have in our reports is up to date. Therefore having tools to control enterprise systems and governing them increase confidence in the data we obtain from the reports. Governance in this case supports the process but also gives huge money savings when we can trust the information being provided. In addition we can verify that data coming from multiple sources is coming from reliable sources and not being injected by unknown parties. All those processes are supporting organization to address standards and requirements defined by European Union data protection legislation. European Union new requirements are challenging enterprises to bring all the processes into line for the usage of data in a new way that was previously unforeseen.

Authorized licensed use limited to: UNIVERSIDAD DE GRANADA. Downloaded on January 14,2025 at 16:49:47 UTC from IEEE Xplore. Restrictions apply.

[1] https://www.ibm.com/us-en/marketplace/information-governance-catalog/details#product-header-top
[2] Smallwood R. F., Information Governance: Concepts, Strategies, and Best Practices, Wiley 2014
[3] http://blogs.gartner.com/debra_logan/2010/01/11/what-is-information-governance-and-why-is-it-so-hard/
[4] EDRM, How the Information Governance Reference Model (IGRM) Complements ARMA International's Generally Accepted Recordkeeping Principles (GARP®), EDRM 2011
[5] http://www.eugdpr.org
[6] http://searchdatamanagement.techtarget.com/feature/Ex
[7] http://www.ibm.com

Authorized licensed use limited to: UNIVERSIDAD DE GRANADA. Downloaded on January 14,2025 at 16:49:47 UTC from IEEE Xplore. Restrictions apply.