

Reimaging Data Governance for a Healthcare Ecosystem

Achieve unified data governance with Microsoft Purview and Neudesic, an IBM Company

Executive Summary

Across all industries, and particularly in healthcare, companies are looking to tap into their data assets and drive intelligent decision making across the organization. Healthcare providers are looking to tap into their data assets to improve patient outcomes, act on innovation opportunities, and drive intelligent decision making across the organization. Simultaneously, these companies are also accelerating their digital transformation agendas and modernizing their IT infrastructure.

This piles on the pressure for information management stakeholders who now must deal with skyrocketing volumes of data, more data-driven priorities, and evolving compliance requirements in an progressively complex and connected healthcare IT environment.

According to IDC, 175 zettabytes of data will be generated annually by 2025 - a 300% increase from 2021. The healthcare industry alone generates over 30% of the world's data and is anticipated to reach 36% within the next three years. If current trends are any indication, this data will span the edge, multiple clouds and the data center while needing to be managed across both operational and analytical healthcare systems.

Most organizations just don't have the right infrastructure to ensure better discovery, trust, and protection of this data. Their existing data governance systems are either system-specific, or use case specific, or limited in lineage capabilities, or are not scalable.

This book shares how building your data governance foundation on Microsoft Purview, a unified data governance service from Microsoft, enables your data consumers to easily find valuable, trustworthy data across your on-premises, multicloud, and software as a service (SaaS) data sources. It also shares how you can take the next step in this journey with IBM Consulting and Neudesic.

Table of Contents

01	Trusted data is key to digital transformation	Page 04
02	Time to act on data governance	Page 05
03	Reimagine your approach with a holistic data governance framework	Page 04
04	Govern data across the entire data estate with Microsoft Purview Automate data discovery Make data easily discoverable with Purview Data Catalog Trace lineage of data assets across the data estate Classify data at scale Get a bird's eye-view of sensitive data	Page 07
05	Accelerate your journey with IBM Consulting's and Neudesic's additional tools and capabilities	Page 13
06	Take the next step to better discover, understand, and trust your data	Page 14
07	Case study: Automating data discovery and enhancing data governance for Grundfos	Page 15

Trusted data is key to digital transformation

Across the entire Health and Life Sciences space, healthcare providers, payers, biotech, and pharmaceutical companies, are investing to accelerate digital initiatives. At the heart of this transformation lies data and the ability to generate analytical and predictive insights from it.

But in order for this data to fuel digital transformation, organizations need to ensure that the right people are able to easily find and use this data, while staying on the right side of regulation. Nowhere is this more evident than in healthcare, where there's an increasing volume of disparate patient data sources across multiple systems and fragmented repositories. This extends to the growing adoption of IoT devices that monitor and deliver real-time insights into health data and patient behavior.

Essentially healthcare organizations must solve two problems simultaneously. First, the data must be timely because digital transformation is all about accelerating time-to-market -- whether that's delivering personalized experiences to customers, providing real-time insights to businesses, or optimizing treatments and outcome-based decision making according to the most current patient data.

Speed and trust often compete with each other and it has always been common for organizations to focus on one or the other.

Today, more than ever, it has become critical for organizations to overcome these traditional data challenges.

Data is expected to grow by 400% from 44 ZB in 2020 to 175 ZB in 2025.¹

45% of enterprise data could be described as 'dark data', i.e. data unused for analytics initiatives.²

Healthcare providers are at the dawn of a new AI-driven era enabled by data from IoT and other digital sources. Product leaders of healthcare provider IoT applications should exploit this opportunity for AI-enabled IoT to create new value propositions leading to new revenue streams.³

Sources:

1. IDC Whitepaper, sponsored by Seagate, The Digitization of the World: From Edge to Core, 2018

2. Voice of the Enterprise: Data & Analytics, Data Platforms, 451 Research, S&P Global Market Intelligence, 2021

3. AI-Enabled IoT — Implications for Healthcare Providers, Gartner, 2022

Reimagine your approach with a holistic data governance framework

For organizations to unlock the value of data, technology is key but not enough. Data governance should become a cross functional discipline, orchestrating people, process, and technology. This allows organizations to be comprehensive, consistent, and coherent in the way they define, discuss, analyze and leverage their data as information. The topics below are the disciplines that need to be addressed in a Data Governance Program that may span one or multiple years. IBM Consulting s and Neudesic's Data Governance Framework provides a holistic approach, helping you to implement the artifacts for each discipline in a structured and complete way.



Architecture

How well does the organization design, develop, deploy, and manage data architecture?



Business Rules Management

How well does the organization define, develop, and deploy, manage, the application of business rules across the enterprise?



Change Management

Are capabilities in place to adopt required organizational change and drive business value realization?



Data Quality

Is the organization able to consistently define, and measure data quality, and mitigate and data quality issues?



Metadata

How well does the organization capture, manage, and access business, technology, and operational information on key corporate data?



Organization & Stewardship

Are there formal organizations and roles for supporting, managing, and improving governance processes and capabilities?



Policy

How well does the organization define and manage organizational behavior using policy?



Privacy & Security

Are appropriate considerations in place for protections of customer privacy and data security?



Regulation & Compliance

Is the organization correctly prioritizing activities to address people, process, and technology requirements for regulatory and compliance issues?

Govern data across the entire data estate with Microsoft Purview

Microsoft Purview is a unified data governance solution that gives you a holistic, up-to-date map of all your on-premises, multicloud, and software as a service (SaaS) data and helps you establish governance at every step of the data lifecycle. This enables you to spend more time on unlocking the value from your data rather than spending time on integrating and managing infrastructure.

Data Map

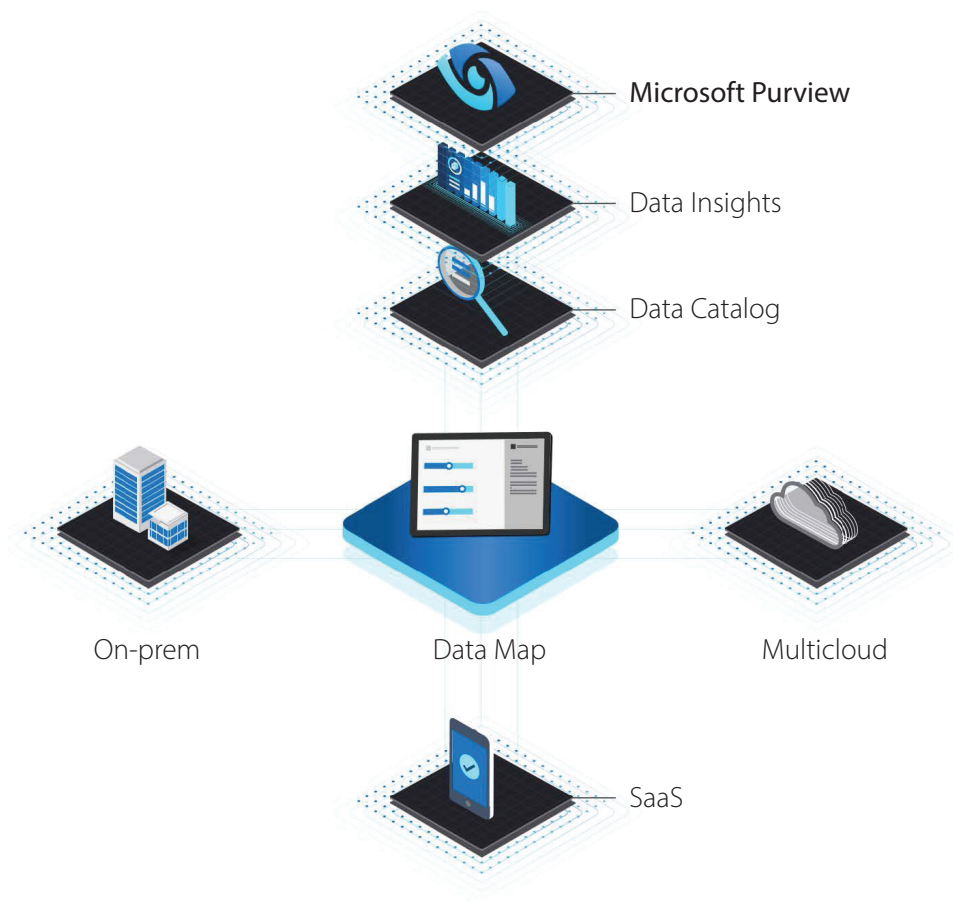
Automate and manage metadata at scales

Data Catalog

Enable effortless discovery for data consumers.

Data Insights

Assess data usage across your organization



01 Automate data discovery

Out of the box support for 35 data sources and growing...¹



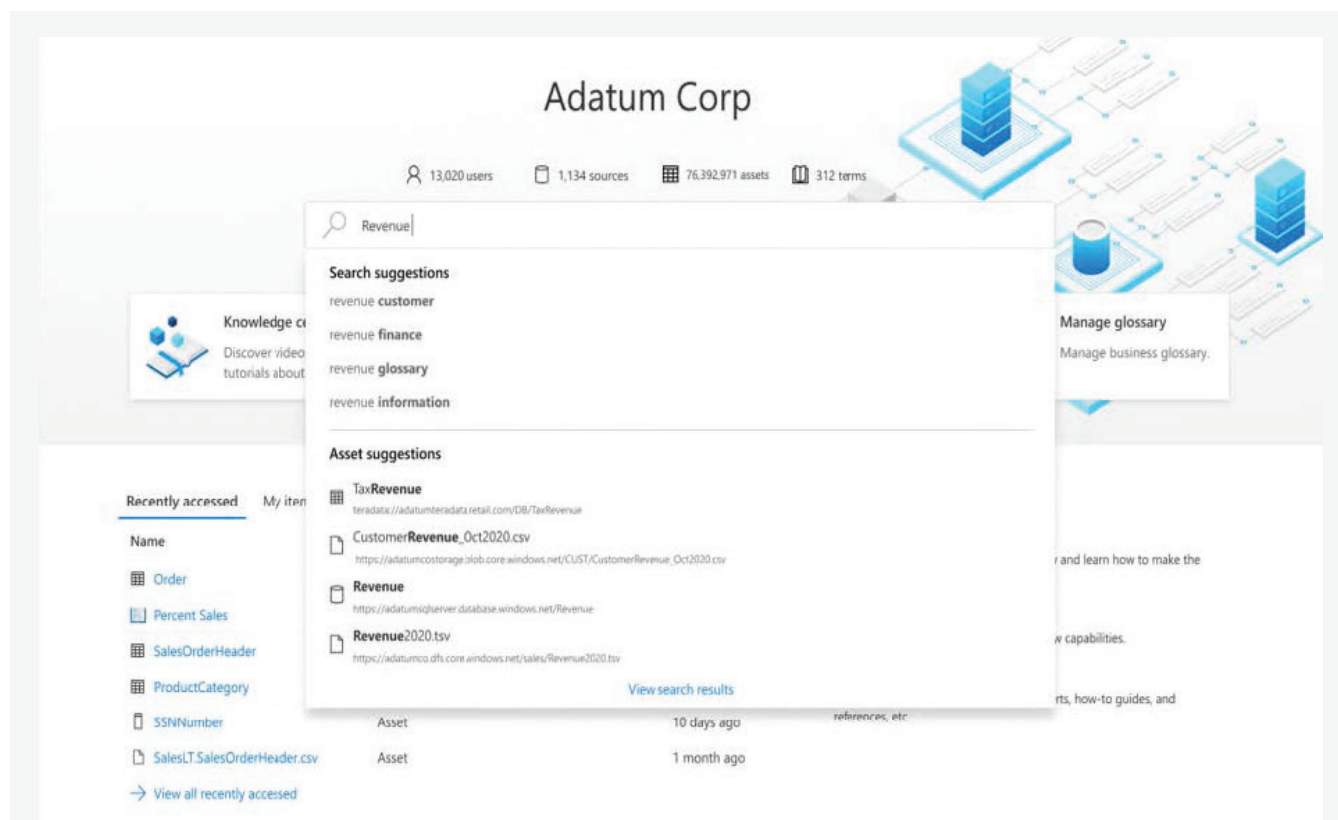
The platform automates today's highly manual data discovery processes by providing data scanning as a fully managed service. This enables data engineers to scan petabyte-sized data estates across on-premises and multi-cloud data sources, to build a data map, that captures key pieces of information such as technical metadata, semantic metadata, and business metadata.

As of today, Purview supports native support for 35 data sources covering on-premises, cloud and SaaS sources, with others in Preview. In addition, Purview provides you the capability to integrate the non-native data systems using the Apache Atlas APIs. At IBM Consulting and Neudesic, we are also developing "Custom Scanners" to collect characteristics of these non native sources to make it easier for you to integrate your data estate in Microsoft Purview.

Sources

1. <https://docs.microsoft.com/en-us/azure/purview/azure-purview-connector-overview>

02 Make data easily discoverable with Purview Data Catalog

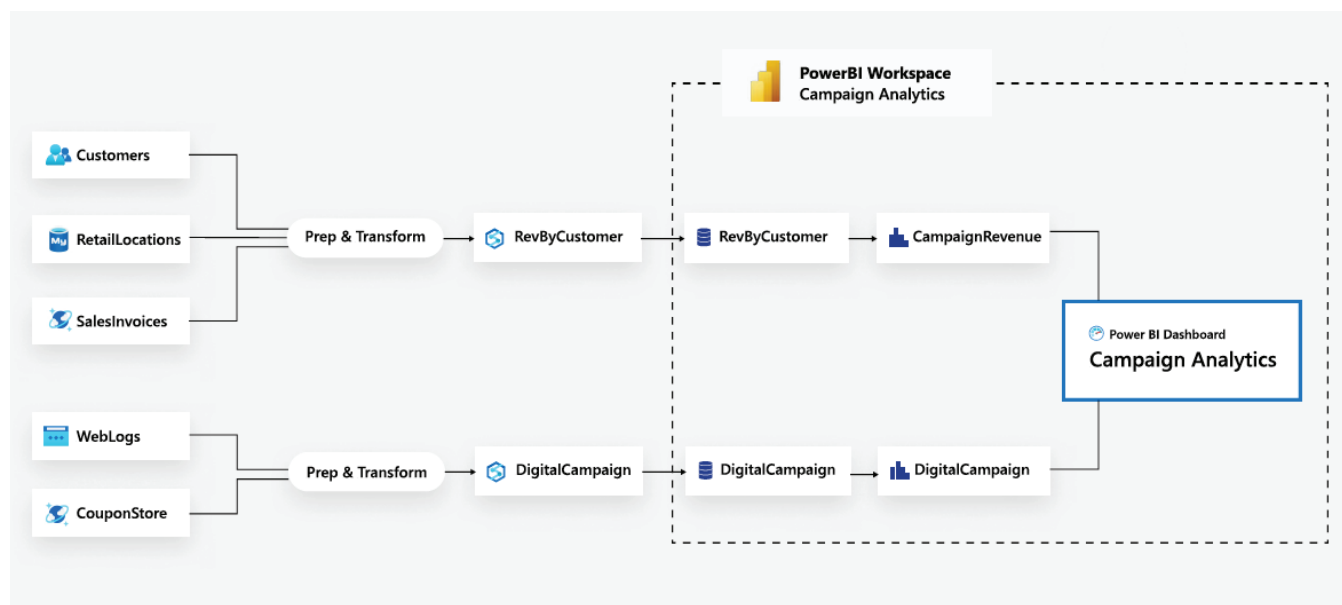


Microsoft Purview Data Catalog with Semantic Search

Most data stewards or data scientists don't know how their data estate is structured. Purview provides a browser experience with powerful data search capabilities, empowering all data consumers to easily find trusted data.

Microsoft Purview also provides an enterprise-grade integrated business glossary. This allows your subject matter experts to provide the business context associated with their data assets and empowers business users to easily discover data assets using the language and terms that are most familiar to them.

03 Trace lineage of data assets across the data estate

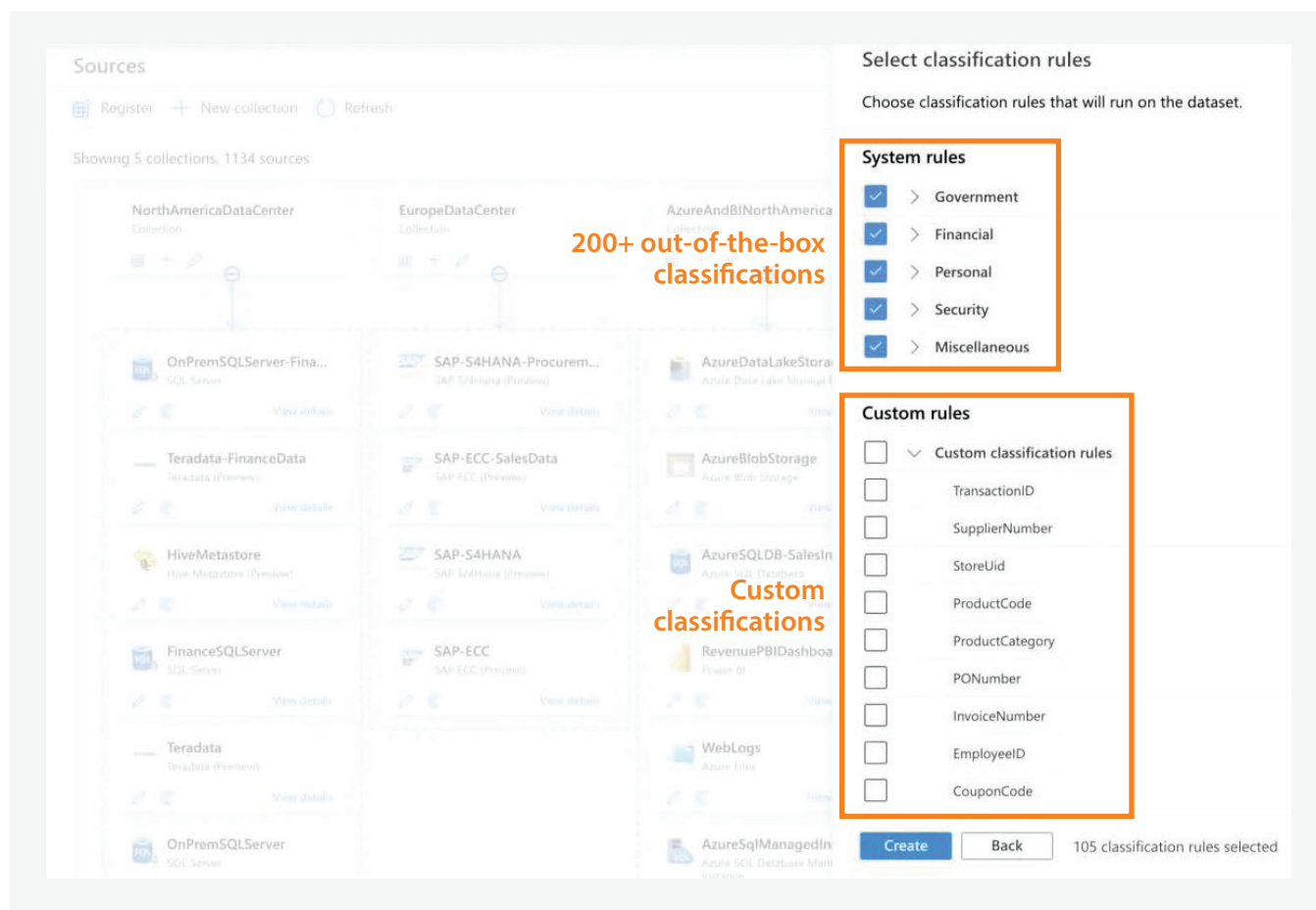


Your data estate may include systems doing data extraction, transformation (ETL/ELT systems), analytics, and visualization systems. Each of the systems captures rich static and operational metadata that describes the state and quality of the data within the systems boundary. For your entire data teams, be it data analyst, data engineer, data scientist or data officer, clearly understanding data lineages can provide critical insights into your data and simplify regulatory compliance.

Purview automates extraction of the movement, transformation, and operational metadata from each data system at the lowest grain possible. Parallely, many of Microsoft data products*, such as Azure Data Factory, Power BI and Synapse pipelines also automatically report data lineage to the data map. The data engineers can easily run the data pipelines and verify if lineage is complete and not broken due to absence of metadata of components or links between components.

**Note: Purview does not yet support lineage for Azure Databricks. However, IBM Consulting and Neudesic can create a custom lineage for when Databricks is used for transformation & propagation of data (ETL).*

04 Classify data at scale



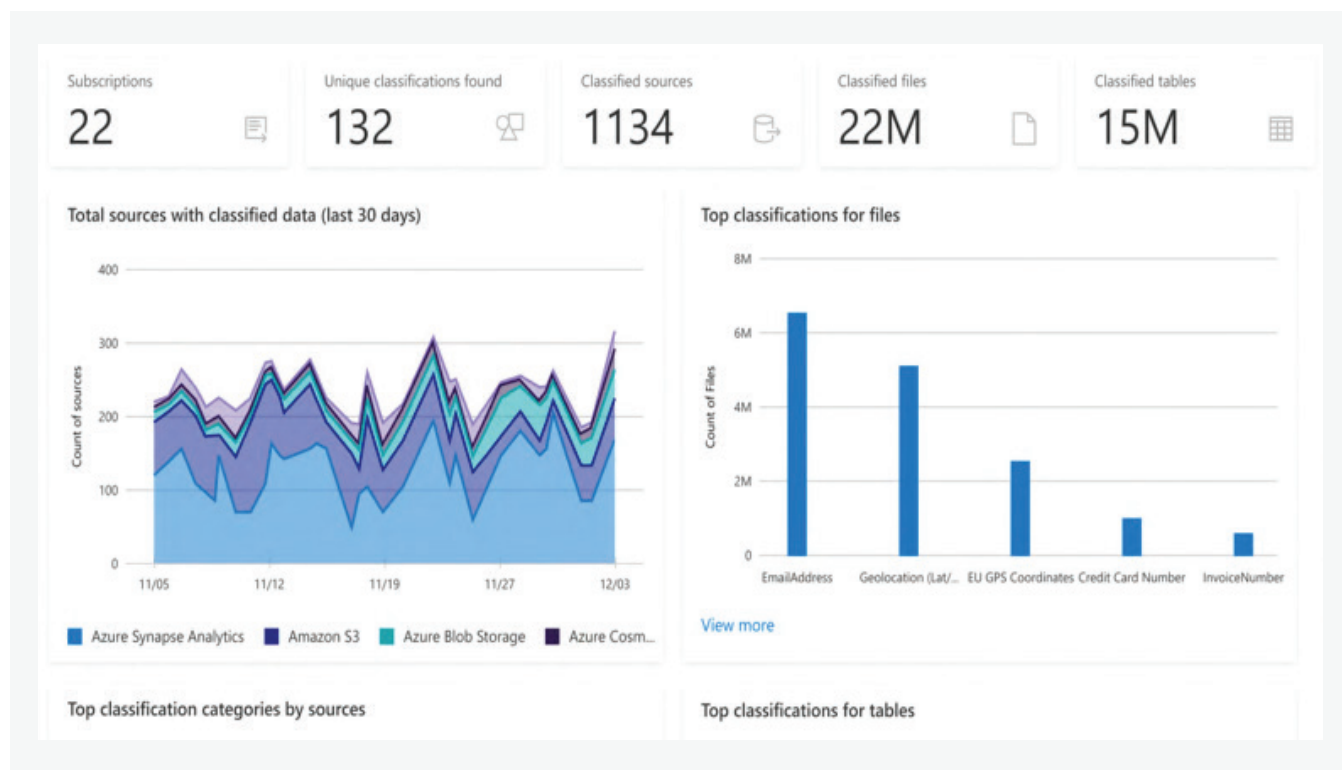
Classifications in Azure Purview

Microsoft Purview doesn't just automate data discovery. It also provides data classification as a fully managed service. There are 200+ built-in classifiers to help data stewards and data engineers to automate the classification of data to specify sensitivity, compliance, industry, business and company-specific value.

For example, if there is a file named multiple.docx and it has a National ID number in its content, Microsoft Purview automatically adds the classification 'EU National Identification Number' to the file asset's detail page.

In some scenarios, the data or compliance teams might want to manually add data classifications. Purview also enables them to create these custom classifiers on a file, table, or column asset level.

05 Get a bird's eye-view of sensitive data



Azure Purview Data Insights

Finally, Purview Data Insights provides a holistic understanding of your data estate, along with the ability to drill down so you can pinpoint issues and maximize the compliant use of data.

This application aims to provide specific insights to the data source administrators, business users, data stewards, data officer and, security administrators.

Asset insights: Distribution of your data estate by source type, by classification and by file size, as some of the dimensions.

Scan insights: Overall health of the scans - how many succeeded, how many failed, how many canceled.

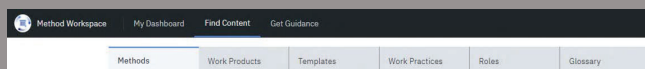
Glossary insights: Distribution of glossary terms by status: how many glossary terms are attached to assets and those not yet attached.

Classification insights: Classifications found during a scan, location of classified data, and a drill-down to the classified files themselves.

Sensitivity labeling insights: Details about the sensitivity labels found during a scan, as well as a drill-down to the labeled files themselves.

Accelerate your journey with IBM Consulting's and Neudesic's additional tools and capabilities

Methodology



A method workspace to use a structured approach for all deliverables in a project. Enables content customization based on applicable Growth Platform & Service Line.

Ideation and Agile

An end-to-end model to chart your transformation journey with an iterative framework and equip you with the experts, practices, and technologies to bring that journey to life.

Frameworks

Organizational and technical approaches to drive consistency and implement best practices across the end-to-end data governance lifecycle:

- Data Governance Framework
- Data Quality Framework
- Data Catalog Use Cases



Architectures

Technical approaches to meet both your current as well as future data governance needs:

- Reference Platform Architecture
- Metadata Architecture; patterns and integration and exchange of MTD between various Data Management components

Add-Ons:

1.Data Quality

Capability to check data integrity based on your project specific data quality rules as input

2.Master Data Management (via Profisee)

Out-of-the-box tooling to ensure alignment to standards for master data

3.Custom lineage for Azure Databricks

4.Automation via Atlas

Automations to upload glossary terms, build custom lineages, and import sensitivity labels

5.Cognitive Classifier

Algorithms to map source data attributes and assess classifications

6.Custom scanners for data sources not available out of box from Purview

7.Customized scanners for selective data extraction and cost savings

8.Patient and Provider indexes using Microsoft Dataverse and Microsoft Cloud for Healthcare

Take the next step to better discover, understand, and trust your data

Gain an understanding of IBM Consulting's and Neudesic's framework and technology solution for all three aspects (People, Process and Technology) of your Data Governance journey, get a thorough demonstration in your environment by starting with 3-5 data sources, and implement the framework at scale, with Microsoft Purview as the key technology component.



Case study: Automating data discovery and enhancing data governance for Grundfos



Situation

Founded in 1945, Grundfos manufactures pumps that provide critical services such as heating, cooling, and clean water to people around the world. Over time, the company started using Internet of Things (IoT) solutions to add value.

As data volumes grew, the company's data became hard to track and manage across the hybrid environment. To find data, Grundfos employees had to know that it existed in the first place, then identify the right colleague and send an email or knock on their office door to ask for the information.

Solution

To drive better data discoverability and governance, Grundfos implemented Microsoft Purview and integrated it with SAP HANA environment and with other Microsoft Data sources, including Azure Synapse Analytics, Power BI, Azure Data Lake Storage, and SQL Server.

Impact

By implementing Microsoft Purview, Grundfos is reducing cumbersome manual tasks, boosting productivity, simplifying compliance, and making it fast and easy for employees to gain full value from the company's data.

"Adopting Microsoft Purview helps us free up time for our resources to be used for more purposeful, proactive work. We'll accelerate our ability to track and interpret a wide range of data, which we'll use to create solutions at scale that truly help improve the environment and people's lives."

Thomas Asger Hansen,
Senior Manager, Enterprise Data and AI,
Grundfos

Contact Us

Lisa Schnitzius

Offering Leader, IBM Consulting
lisa.schnitzius@ibm.com

Tyler Suss

Director, Data & AI Solutions Marketing & Alliances,
Neudesic, an IBM Company
tyler.suss@neudesic.com

Steve Kasten

Director, Solutions Marketing & Alliances Healthcare,
Neudesic, an IBM Company
steve.kasten@neudesic.com

© 2022 IBM All rights reserved. IBM, the IBM logo and ibm.com are trademarks of IBM Corporation in the United States, other countries or both. Merge and its respective logo are trademarks of Merge in the United States, other countries or both. All other company or product names are registered trademarks or trademarks of their respective companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Why IBM Consulting and Neudesic



Global Practice focused on
Azure Data, AI, and IoT



Microsoft's fastest growing GSI



2,000+ Certified Azure Data and
AI Specialists



1,000+ Cloud Migrations in the
past year



30 global centers, 55 languages,
24x7 operations



17 Microsoft Competencies



Analytics on Azure Advanced
Specialization



AI and Machine Learning on
Azure Advanced Specialization