



INTELLIGENT BUSINESS STRATEGIES PRESENTS



Mike Ferguson

**AI-Assisted Active Data and AI
Governance**

- The Route to Healthy Data, Analytics and AI

www.intelligentbusiness.biz
training@intelligentbusiness.biz

OVERVIEW

Most businesses today are operating in a distributed computing environment with data spread across multiple types of data stores on-premises, in multiple clouds, in SaaS applications and at the edge. This makes data much harder to find and to govern. Yet, data governance is now very high priority in most organisations not just to remain compliant with legal and regulatory obligations but also to create a high quality, secure data foundation of reusable data products to underpin data and AI initiatives now happening across every department in the enterprise. With data and AI now strategic in the boardroom, data governance has become so important that companies classified as 'leaders' regard it as a strength that gives them commercial advantage and not just an initiative to remain compliant with legislation like GDPR.

To date however, data governance in many organisations has been fractured with different tools being used to support different data governance disciplines. This includes the use of different tools for data quality, data privacy, data access security, data sharing and data retention. Also, data catalogs, that can automatically discover and classify data in many different data stores are often purchased independently of other data governance tools. Therefore, while automated data discovery is possible, to some extent, almost all data governance disciplines are still highly dependent on people manually keying and re-keying policies and other metadata across many different tools to try to ensure data remains governed and that governance policies are consistently enforced across the enterprise.

However, as more data is created and new data sources continue to appear, the challenge of manually understanding all data relationships and manually governing data is becoming almost impossible. In addition, doing this with multiple tools is also challenging as there are no industry standards to exchange metadata across those tools which means data governance tasks often have to be repeated again and again to ensure data is being correctly governed across a distributed data estate.

Given this increasingly difficult challenge, companies are looking for a more automated way to deal with data governance. To do this requires taking data governance to a new level by introducing AI-driven, active data governance. Active data governance is more than just keeping metadata up to date. This is something that uses AI-driven automatic data discovery and data classification, tag-based policy management, and an AI-driven data governance action framework to continuously govern data more efficiently and effectively. An AI-driven data governance action framework needs to include data governance health metrics to monitor progress, data governance events, automated data governance issue detection, automated verification that actions have occurred, different types of governance action services, data governance action processes and automated triggering of data governance action processes to ensure the health of your data continues to improve.

This two-day in-depth class looks at this problem and shows how to successfully implement AI-driven active data governance across a distributed data estate. This includes AI-driven active governance of data access security, data privacy, data sharing, data retention data quality and data usage. The class looks at the business problems caused by poorly governed data and how it can seriously impact business operations, cause unplanned operational costs, and destroy confidence in accuracy of BI, machine learning model predictions and recommendations and Generative AI.

It also looks at requirements for AI-driven active data governance. Having understood the requirements, you will learn what should be in a governance programme. This includes data governance roles and responsibilities, processes, policies, technologies, and data governance capabilities to govern data across a distributed data estate. It looks at how to implement AI-driven active data governance by breaking the data governance problem down into a series of steps that need to be implemented and looks how to take advantage of emerging AI-driven data governance platforms to implement this. The class will cover

- Data governance disciplines including data curation, data quality, data privacy, data access security, data sharing, data retention, and data usage
- Current problems with data governance today

- Requirements to dramatically improve data governance using AI and automation
- The need for an integrated data governance platform, AI augmentation and AI-automation
- Establishing health metrics to measure effectiveness of your data governance program
- Understanding the core AI-assisted data governance services you need to discover, classify and curate data
- Creating a Data Governance Action Framework for your enterprise
- Data governance observability – monitoring the health of your data
- AI-Assisted data governance action automation
- Implementing AI-assisted governance of different data governance disciplines
- Implementing AI governance to manage and avoid risk

AUDIENCE

This seminar is intended for CDOs, CIO's, Heads of Data Governance, CISOs, Business Analysts, data scientists, BI Managers, data warehousing professionals, data architects, solution architects, Data strategists, Database administrators, IT consultants.

LEARNING OBJECTIVES

Attendees will learn how to set up an AI-driven enterprise data governance program to systematically govern data and content across their distributed data estate . Using a data governance framework and key technologies like data catalogs, automated data discovery, automated data classification , machine learning, Generative AI, AI-agents, decision automation, data marketplaces and workflow. attendees will learn what is needed to discover, classify and govern data and content. This includes creation of health controls and how to implement AI-assisted data and AI governance. This includes automated data discovery, data curation data access security, data privacy, data loss prevention, data sharing, data retention, and data quality.

MODULE 1: WHAT IS ACTIVE DATA GOVERNANCE AND WHY DO WE NEED IT?

This module looks at what active data governance is, how ungoverned data can impact business operations, decision making and increase risk, and where most companies are in terms of implementing data governance. It also looks at how this challenge has grown to encompass multiple different governance disciplines including data quality, data privacy, data access security, data sharing, data retention, and data usage across a hybrid, multi-cloud distributed data estate. Finally, it looks at the fractured approach many companies have been taking to tackle this challenge, the problems caused by using best of breed tools and why it is no longer practical to expect people to do everything manually when AI can help automate tasks. .

- The ever-increasing distributed data landscape – on-premises, multiple clouds, SaaS applications and the edge
- The impact of ungoverned data on compliance with legislation and regulations, business profitability and ability to respond to competitive pressure
 - Data Governance – Where are we?
 - Beyond data quality to multi-disciplinary data governance
 - Problems caused by a fractured approach to data governance using multiple best of breed tools
 - Why AI is needed, and a data catalog is not enough .

MODULE 2: WHAT ARE THE REQUIREMENTS FOR AI-DRIVEN ACTIVE DATA GOVERNANCE TO GOVERN DATA ACROSS DISTRIBUTED DATA ESTATE ?

This module looks at what the requirements are to govern data in terms of people, process, technologies, policies and capabilities and what we need to change to move from manual data governance to AI-driven active data governance automation

- Key requirements for governing data & content across a distributed data estate
- People
 - Data governance roles, responsibilities and groups
- Core processes & tasks to standardise data governance activities, approvals and actions

- Technology requirements
 - Universal data governance platform
 - Data catalog
 - Trainable ML classifiers
 - Data Governance co-pilots & AI agents
 - Dynamic data masking
 - Universal data access control
 - Data marketplace
 - Data loss prevention software
 - Data governance observability
 - Data governance enforcement agents
 - The Data Catalog Marketplace
 - Alation, Ataccama, Atlan, AWS Glue Data Catalog, BigID, Cambridge Semantics Anzo Data Catalog, Collibra Data Catalog, data.world, Databricks Unity Catalog, Google Data Catalog, Hitachi Vantara Pentaho Data Catalog, IBM Knowledge Catalog, Informatica IDMC Data Governance and Catalog, Microsoft Purview, Oracle, SAP DataSphere, Qlik (Talend) Data Catalog, TopQuadrant TopBraid
 - Data governance platforms
 - Data access security tools
- Foundational AI-assisted active data governance services
 - Data catalog business glossary
 - Data source discovery scanners
 - Data catalog data map and data relationship knowledge graph
 - Data governance classifiers
 - Generative AI for automated metadata enrichment and conversational data search
- Active data governance applications, policies & policy types needed to govern:
 - Data quality
 - Data access security
 - Data privacy
 - Data retention
 - Data loss prevention
 - Data sharing
 - Data use and maintenance
- Data governance action framework
 - Data governance health metrics to measure data health, how it's processed, protected and used
 - Critical data elements

- Example data governance events & incidents to be monitored
- Likely triggers of governance actions
- AI-models for data governance alerting, recommendation, action and task automation
- Data governance observability Agents to monitor & detect incidents
- Using metadata lineage to understand the impact of events and incidents
- Data Governance action services
 - Data governance verification service to check that governance & data curation activities have occurred
 - Data governance rectification services to rectify metadata
 - Data governance action invocation service
- Example actions to be performed on items e.g. on a data source, a business term, a data asset, a policy, a data product
- Data governance action processes to ensure actions are standardized and consistently executed
- Organising human actions using an in-box and human task monitoring

MODULE 3: FOUNDATIONAL ACTIVE DATA GOVERNANCE SERVICES - THE IMPORTANCE OF A BUSINESS GLOSSARY

This module looks at establishing a common business vocabulary in a business glossary of your data catalog to create common business data names and definitions for your data so you can understand your data. This enables you to search for and govern data across your data estate from a business perspective

- Data standardisation using a common business vocabulary
- The purpose of a business glossary in data governance
- Business glossary capability in data catalogs
 - Alation, Amazon Glue, Collibra, Informatica IDMC Business Glossary, IBM Knowledge Catalog, Microsoft Purview, Qlik (Talend) Business Glossary and Data Catalog, TopQuadrant TopBraid EDG Business Glossary

- Planning for a business glossary
- Glossary roles and responsibilities
- Glossary term submission, voting approval and dispute resolution processes
- Approaches to creating a common vocabulary
- Organising data definitions in a business glossary
- The role of a data concept model
- Utilising a common vocabulary in BI tools semantic layers, data modelling, data fabric, MDM and APIs

MODULE 4: FOUNDATIONAL ACTIVE DATA GOVERNANCE SERVICES - AUTO DATA DISCOVERY, CATALOGUING AND MAPPING TO A BUSINESS GLOSSARY

Having defined your data, this module looks at automatically discovering what data you have, where it is and how it maps to your business glossary to provide a business understanding of your data estate

- Understanding your data estate - the critical role of AI-driven data catalog software
- Registering data sources for discovery
- Automated data discovery, and data quality profiling, using a Data Catalog
- Automating metadata enrichment using Generative AI
- AI Automated suggestion of business glossary terms during data discovery
- AI-Assisted mapping of physical data assets to a business glossary terms
- Setting policies and health controls to monitor business glossary creation and data curation
- Using AI agents to monitor data curation activity and automate curation actions

MODULE 5: FOUNDATIONAL ACTIVE DATA GOVERNANCE SERVICES - AI-DRIVEN DATA & CONTENT CLASSIFICATION

This module looks at manually and AI-driven automatic labelling of data and content to know how to govern it using predefined AI classifiers, user-defined classification schemes and trainable AI classifiers. It then looks at how AI-driven automatically classified data shows up in a data catalog and how policies can be assigned to labelled data to govern it across your data estate.

- What is AI-driven data classification?

- Automatically detecting and classifying sensitive structured data using pre-defined AI classifiers in a data catalog
- Creating your own data confidentiality and retention classification schemes
- Manually classifying content using your own classification scheme, e.g. Office Documents, SharePoint, Email, Chat, Microsoft Teams or Zoom Meetings
- Training and using AI classifiers to auto label content across your data estate
- Using classification insights to understand sensitive data proliferation and data redundancy across your estate
- Understanding policies, policy groups and tag-based policy management

MODULE 6: IMPLEMENTING AI-ASSISTED GOVERNANCE OF DATA SECURITY ACROSS YOUR DISTRIBUTED DATA ESTATE

Having classified the data and content in your data estate, this module looks at protecting data and content in your data estate with focus on that which is classified as sensitive or confidential. It looks at setting and enforcing policies to govern data access and usage security as well as governing data loss prevention.

- Data security objectives
- Key technologies in governing data security
 - Policy establishment
 - Policy enforcement
- Steps to implement data security
- Setting health controls and enterprise wide and domain specific policies in your data catalog to govern data access across your data estate
 - Attribute based access control
- Unifying data access control across multiple data stores
- Universal authorisation fabric software (e.g. IBM, Immuta, Databricks (Okera)) and how they integrate with data catalogs
- Using cloud application security brokers
 - Auto discovery of cloud app data usage
 - Setting policies to govern access to and use of sensitive data and content from applications
 - Monitoring cloud application activity
- Dealing with insider risk management and internal information barriers

MODULE 7: IMPLEMENTING AI-ASSISTED GOVERNANCE OF DATA PRIVACY ACROSS YOUR DISTRIBUTED DATA ESTATE

This module looks at AI-assisted governance of personal and financially sensitive data across your data estate to remain compliant with legislation in multiple jurisdictions that your company operates

- Data privacy objectives
- Data privacy legislation – GDPR, CCPA, HIPAA and more
- Steps involved in enterprise-wide data privacy risk management
- Using AI to automatically identify where unprotected personal data is located
- Setting policies and health controls in your data catalog to govern data privacy across your data estate
- Data privacy insights on sensitive data location, how it moves and where you are at risk
- Data privacy policy enforcement across a distributed data landscape
 - Linking your data catalog to other technologies
 - Encrypting and de-identifying personal data
 - Using data loss prevention (DLP) to avoid loss of personal data
 - Protecting personal data in email, chat, documents, file shares, cloud storage, and endpoints
- Using AI agents to monitor data privacy violations and automate actions
- Monitoring AI-Agent effectiveness
- Managing subject action requests

MODULE 8: IMPLEMENTING AI-ASSISTED GOVERNANCE OF DATA RETENTION ACROSS YOUR DISTRIBUTED DATA ESTATE

This module looks at governing the lifecycle of data across your data estate and how you can use AI to automatically classify data to label it for retention, set policies to control how long data is kept for and what happens to it on expiry. It also looks a special purpose retention condition known as “legal holds” placed on data by legal departments.

- Creating a data retention classification scheme
- Complying with country and region-specific legislation
- Training AI classifiers to label you data

- Automatically classifying data & content using AI to create retention labels
- Setting policies and health controls in your data catalog to govern data retention across your data estate
- Using AI agents to monitor data retention expiration and automate actions to destroy, archive and hold data

MODULE 9: IMPLEMENTING AI-ASSISTED GOVERNANCE OF DATA SHARING ACROSS YOUR DISTRIBUTED DATA ESTATE

This module looks at producing trusted, compliant data products to be shared across the enterprise and beyond and how data sharing can be governed

- Data sharing objectives
- Key technologies to help produce high quality, compliant data products for sharing
- Steps to creating data products using a data catalog business glossary, automated data discovery and mapping and AI-assisted data engineering
- A unified approach to producing high quality data products using Data Fabric and DataOps pipelines
- AI-assisted publishing of certified, high quality, compliant data products in a data marketplace
- Potential metadata standards for data products e.g., DPROD
- AI assisted governance of data sharing and consumption using data contracts in a data marketplace
- Consumer use of AI-assisted conversational data search in a data marketplace
- Creating a standard data sharing approval process for consumers
- Using AI agents to monitor and track shared data consumption and usage

MODULE 10: IMPLEMENTING AI-ASSISTED GOVERNANCE OF DATA QUALITY ACROSS YOUR DISTRIBUTED DATA ESTATE

This module looks at AI-Assisted governance of data quality across your data estate.

- The business impact of bad quality data
- Common data quality health metrics
- AI-assisted creation of data quality validation, matching and survivorship rules in your data catalog using Generative AI
- Using your data catalog to automatically profile and validate your data quality
- Setting data quality health controls and thresholds in your data catalog to govern data quality across your data estate
- Leveraging Data Catalog AI-suggested data quality rules
- Integrating data observability with data catalogs to monitor and report data quality issues
- Using AI to monitor data quality validation rules against health control thresholds across your distributed data estate
- Auto generating validation rules when data quality thresholds are breached
- Using the data catalog for AI-assisted data cleansing & generation of data integration pipelines
- AI-assisted MDM & the data catalog

MODULE 11: IMPLEMENTING AI-GOVERNANCE

This module looks at AI governance to manage AI models and AI risk in your organisation. It looks at:

- AI Governance best practices including:
 - Creating an AI inventory & risk registry
 - Setting up accountability for AI
 - Evaluating & mitigating AI risk
 - Governing AI Development
 - AI Observability
- Formalising processes and enabling auditing
- Explainable AI
- AI Observability
- Avoiding PII leakages when building vectors for generative AI LLMs
- Establishing AI Guard Rails for GenAI
- AI Governance tools

PRESENTER



Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant, he specialises in BI / analytics and data management. With over 40 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, data strategy, technology selection, data architecture, and data management. Mike is also conference chairman of Big Data LDN, the largest data, analytics and AI conference in Europe and a member of the EDM Council CDMC Executive Advisory Board. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the

Relational Model, a Chief Architect at Teradata on the Teradata DBMS. He teaches popular master classes in:

- Building a Data and AI Strategy for A Data-Driven Enterprise
- Data Catalogs – Governing and Provisioning Data in a Data Driven Enterprise
- Data Virtualisation & the Logical Data Warehouse
- Data Warehouse Modernisation
- Embedded Analytics, Intelligent Apps, AI Agents & AI Automation
- AI-Driven Active Distributed Data and AI Governance
- Migrating your Data Warehouse to the Cloud
- Modern Data Architecture
- Practical Guidelines for Implementing a Data Products (Data Catalog, Data Fabric, Data Product Development, Data Marketplace)
- Real-Time Analytics

ONSITE TRAINING

All training can be given as onsite education, tailored to your company's requirements. For further details please contact us at training@intelligentbusiness.biz

