

TECHNOLOGY TRANSFER PRESENTS

# MIKE FERGUSON

---

## DATA CATALOGS

---

**governing and Provisioning Data in a Data Driven Enterprise**

---

**ONLINE LIVE STREAMING**

**DECEMBER 12, 2024**



info@technologytransfer.it  
www.technologytransfer.it

## **ABOUT THIS SEMINAR**

This one-day class looks in detail at what a Data Catalog is and what the main reasons are for needing one. In particular, we will look at the challenge companies are dealing with in respect to data complexity and the need to implement both an enterprise Data Governance program and provide technologies to help people find, engineer, provision and consume 'business ready' data for both analytical and operational use cases.

### **WHO SHOULD ATTEND**

This seminar is intended for business and IT professionals responsible for data engineering, data provisioning and enterprise data governance including data access security, data privacy, data sharing, data usage, data retention, data quality of both structured data and content.

This includes Chief Data Officers, citizen and professional IT Data Engineers, Data Architects, Data Scientists, Heads of Data Governance, Data Stewards, Solution Architects and Enterprise Architects. It assumes a basic understanding of data governance, data management, metadata, data warehousing, data cleansing, data integration etc.

### **WHAT YOU WILL LEARN**

Attendees will learn how Data Catalogs work and what their capabilities are. They will also learn how to use Data Catalogs to discover, classify and catalog data in multiple data stores both on premises, across multiple Clouds, and at the edge.

For people involved in data Governance, they will learn how to use Data Catalogs in an enterprise Data Governance program to systematically classify data and set policies to govern classified data and content across their distributed data estate from a single place. This includes automatic discovery and classification of sensitive data, Governance of data access security, data privacy, data loss prevention, data sharing, data usage, data retention, and data quality.

Also, data engineers and data scientists will learn how to use Data Catalogs to discover data they can engineer in data integration pipelines to produce data products.

# OUTLINE

## 1. Introduction

The session looks at typical existing setups and challenges that companies are facing with respect to data to explain why Data Catalogs are needed.

- The ever-increasing complex distributed data landscape – on-premises, multiple Clouds, SaaS applications and the edge
- The growth in new data sources
- Disparate operational transaction systems – SaaS applications, on-premises and Cloud
- Existing siloed analytical systems – Data Warehouse, Data Lakes, Data Lakehouses
  - Data Engineering problems in a siloed environment
- The emergence of Data Mesh and its impact with respect to Data Engineering and Data Architecture
- The impact of ungoverned data on
  - Business operations, Decision Making and risk
  - Business profitability and ability to respond to competitive pressure
  - Ability to comply with Data Privacy legislation in one or more jurisdictions
  - Data security, the possibility of Data Breaches and their Business impact
- Major requirements facing companies with respect to data
  - The need for end-to-end data governance to:
    - Make data easy to find and understand no matter where it is
    - Know where sensitive data is so it can be governed
    - Comply with multiple data privacy regulations and legislation
    - Avoid Data Breaches in a complex distributed data estate
    - Understand where data quality is poor
    - Govern data access, privacy, sharing, usage, retention and quality
  - The need to accelerate Data Engineering and provision business ready data to share and reuse across the enterprise
  - Key technologies needed – the Data Catalog and Data Fabric

## 2. What is a Data Catalog and why have one?

This session looks at what a Data Catalog is, why you need it, software vendors in the market and outlines Data Catalog capabilities.

- What is a Data Catalog
- Why have them?
  - Data Catalog use cases
- The Data Catalog Marketplace
  - Alation, Ataccama, AWS Glue Data Catalog, BigID, Cambridge Semantics Anzo Data Catalog, Collibra Data Catalog, data.world, Google Data Catalog, Hitachi Vantara Lumada, IBM Watson Knowledge Catalog, Informatica IDMC Data Governance and Catalog, Microsoft Purview, Oracle, SAP, Talend Data Catalog, TopQuadrant TopBraid, Truist Zaloni Data Catalog
- Core Data Catalog capabilities
  - Business Glossary
  - Automated data discovery
  - Manual and automated mapping to Business Glossary
  - Manual and automated Data Governance classification
  - Automated data quality profiling
  - Data Governance policy management
- What types of metadata and asset types are stored in a Data Catalog?
- Supporting different types of personas
- Skills needed to use one
- Data Catalog maintenance – manual, automated

## 3. Data Catalog Capabilities: the importance of a Business Glossary

This session looks at the need to understand your data landscape from a business perspective. The key to making this happen is to establish a common business vocabulary in the Business Glossary of a Data Catalog to create common data names and definitions for your data. This enables you to search for and govern data across your data estate from a business perspective.

- Data standardisation using a shared business vocabulary

- The purpose of a common vocabulary in Data Governance
- Business Glossary software – now a capability of a Data Catalog
  - Alation, Amazon Glue, Collibra, Informatica Axon Business Glossary, IBM Watson Knowledge Catalog, Microsoft Azure Purview, Talend Business Glossary and Data Catalog, SAS Business Data Network, TopQuadrant Top Braid 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 Data Modelling, ETL, BI, ESB, APIs, & MDM

#### **4. Data Catalog Capabilities: understanding your Data Landscape using auto data discovery, cataloguing and mapping to a Business Glossary**

Having defined your data, this session looks at discovering what data you have, where it is and how it maps to your Business Glossary to provide a business understanding of your data landscape.

- Understanding your data landscape - the critical role of Data Catalog software
- The data discovery process
- Registering data sources for discovery
- Automated data discovery, profiling, using a Data Catalog
- Mapping data assets to a Business Glossary

#### **5. Data Catalog Capabilities: classifying data and content to know how to govern it**

This session looks at manually and automatically labelling data using a Data Catalog to know how to govern it using predefined classifiers, user-defined classification schemes and trainable classifiers. It then looks at how 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 data classification?
- Automated sensitive data type detection and classification using pre-defined trained classifiers
- Creating your own data classification schemes for data confidentiality and retention
- Manually classifying content using your own classification scheme, e.g. Office Documents, SharePoint, Email, Chat, Microsoft Teams or Zoom Meetings
- Training classifiers to automatically label content
- Using trained classifiers to auto label content in the Cloud and on-premises
- Using your own classification schemes with a Data Catalog
- Automatically classifying sensitive structured data and objects using a Data Catalog
- Using classification insights to understand sensitive data proliferation and data redundancy across your estate
- Setting policies in a Data Catalog to govern data across your data estate

#### **6. Data Catalog Use Case: what's needed to govern data across distributed Data Landscape?**

This session looks at what the requirements are to govern data in a modern enterprise and how the requirement can be met using a Data Catalog.

- Key requirements for governing data and content across a distributed data landscape
- What do you need to know to govern data?
- Introducing a Data Governance framework to help meet the challenge
- People
  - Key roles and responsibilities
  - Getting the organisation and operating model right
  - Data owners, data stewards, data governance control board and working groups•
- Core processes needed to establish and govern commonly understood data
- Types of policies and rules needed to govern:
  - Data quality
  - Data access security
  - Data privacy
  - Data retention
  - Data loss prevention

- o Data sharing
- o Data use and maintenance
- The role of the Data Catalog
  - o Automated sensitive data discovery
  - o Trainable classifiers
  - o Data Catalog policy creation
  - o Integration with policy enforcement technologies
    - Dynamic data masking
    - Data loss prevention
- Core Data Governance capabilities needed
- Tasks involved in governing a distributed data landscape

### 7. Data Catalog Use Case: finding, engineering and provisioning data

This session looks at what the requirements are to accelerate Data Engineering in a modern enterprise and how the requirement can be met using a Data Catalog and Data Fabric software.

- Defining data products in a Business Glossary
- Automatically discovering, mapping and classifying data in a Data Catalog
- Integration with Data Fabric
- Building pipelines to produce data products
- Creating a data marketplace as a Data Catalog application to share business ready data
- Publishing and consuming data products using a data marketplace and Data Catalog metadata

## SPEAKER

**Mike Ferguson** is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in Business Intelligence and Enterprise Business Integration. With over 39 years of IT experience, he has consulted for dozens of companies on Business Intelligence Strategy, technology selection, enterprise architecture, and data management.

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 and European Managing Director of Database Associates.

He teaches popular master classes in Operational Business Intelligence, New Technologies in DW and BI for the Agile Enterprise, Big Data Multi-Platform Analytics, Master Data Management and Enterprise Data Governance

# INFORMATION

## PARTICIPATION FEE

€ 800

The fee includes all seminar documentation.

## SEMINAR TIMETABLE

9.30 am - 1.00 pm  
2.00 pm - 5.00 pm

## HOW TO REGISTER

You must send the registration form with the receipt of the payment to:  
info@technologytransfer.it

TECHNOLOGY TRANSFER S.r.l.  
Piazza Cavour, 3 - 00193 Rome (Italy)

## PAYMENT

Wire transfer to:  
Technology Transfer S.r.l.  
Banca: Cariparma  
Agenzia 1 di Roma  
IBAN Code:  
IT 03 W 06230 03202 000057031348  
BIC/SWIFT: CRPPIT2P546

## GENERAL CONDITIONS

### DISCOUNT

The participants who will register 30 days before the seminar are entitled to a 5% discount.

If a company registers 5 participants to the same seminar, it will pay only for 4.

Those who benefit of this discount are not entitled to other discounts for the same seminar.

### CANCELLATION POLICY

A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.

### CANCELLATION LIABILITY

In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.

## MIKE FERGUSON

Data Catalogs

December 12, 2024

Registration fee:  
€ 800

first name .....

surname .....

job title .....

organisation .....

address .....

postcode .....

city .....

country .....

telephone .....

fax .....

e-mail .....



Stamp and signature

Send your registration form with the receipt of the payment to:  
**Technology Transfer S.r.l.**  
Piazza Cavour, 3 - 00193 Rome (Italy)  
Tel. +39-06-6832227 - Fax +39-06-6871102  
info@technologytransfer.it  
www.technologytransfer.it

If anyone registered is unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.