

EDM Webinar

Best practices for managing the explosion of External Data providers

A conversation with



Scott Albin
GM, International & CSO
Demyst



Jim Halcomb
Head of Product Management
EDM Council



Moderator



Jim Halcomb

Head of Product Management
EDM Council

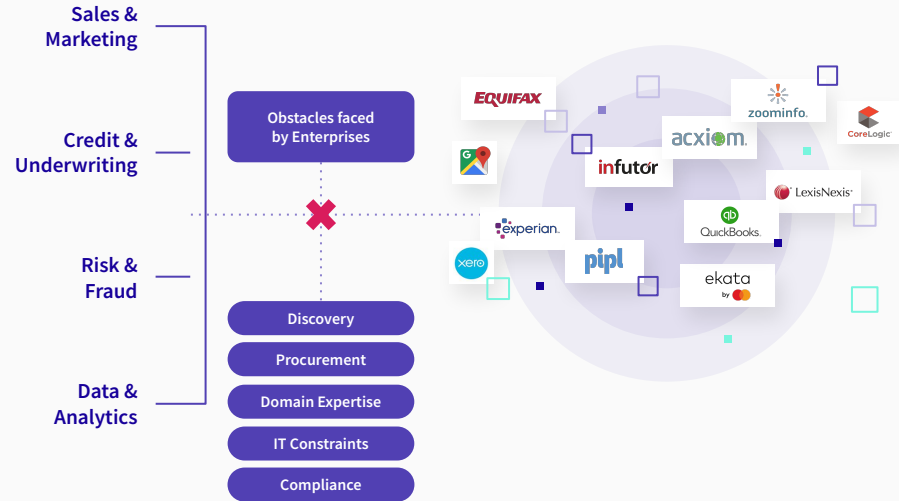


Scott Albin

GM, International & Chief Strategy Officer
Demyst



As data and compliance needs proliferate, large companies are **getting stuck** with the **growing costs** and **complexity** of managing a huge variety of External Data



➤ The fastest growth external data categories are Business (esp. SMEs), Consumer (esp. unbanked), Digital, ESG and Property

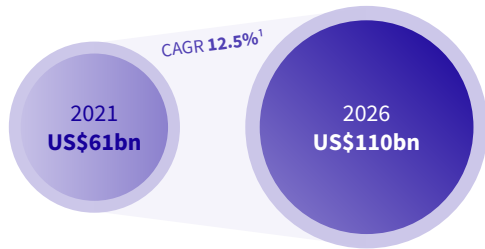
Data Management Challenges	Discovery	Governance	Deployment		Maintenance
Most Common Use Cases	Acquisition / Digital	AML / KYC	Identity / Fraud	Credit / Underwriting	Risk Management
Example Data Providers	<p>Business</p> <ul style="list-style-type: none"> Companies Search Commercial Profile Database Balance Sheets Watch List Search 	<p>Consumer</p> <ul style="list-style-type: none"> Identity Management Identity Verification Social Search Court Search 	<p>Digital</p> <ul style="list-style-type: none"> IP Probe Email Address Investigation Domain Availability Broadcast Transcript Feed 	<p>ESG</p> <ul style="list-style-type: none"> Risk Ratings Carbon Data Controversy Screening ESG Risk Metrics 	<p>Property</p> <ul style="list-style-type: none"> Location Details Sales History Property Search Validate Address

Poll:
Which of the following external data domains are most in demand within your company?

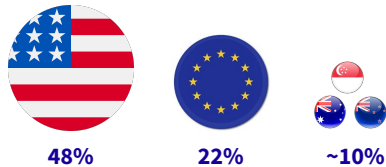
How big is the market for external data?

> The external data market will represent >\$110b in spend by 2026

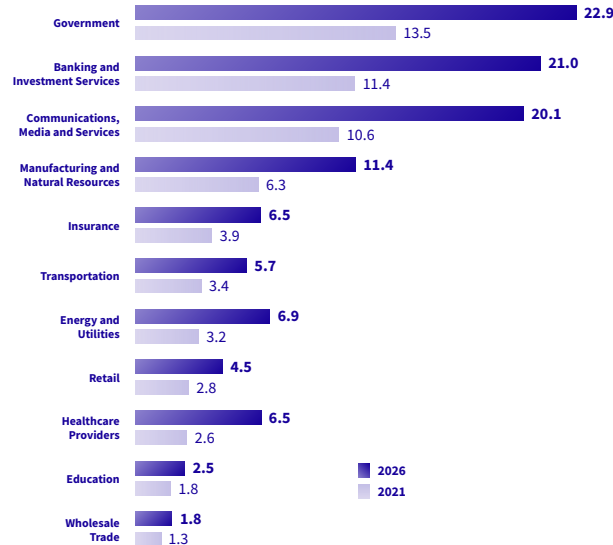
Size and industry distribution of the market for External Data



Top 3 regions²



External Data Market by Industry (\$bn)



Familiar industry themes

- Regulatory pressure** and growing scrutiny on consumer privacy and data compliance is forcing adoption of new ways to manage data
- Data analytics & mobilization** are driving and accelerating business transformation and increasingly requiring new data sources
- Cloud adoption is accelerating** due to scalability, cost efficiency and flexibility, increasing data consumption and connectivity options to new sources
- Dramatic growth in data volumes** leading to rapid growth in data providers, which in turn drives up buyer search costs for high value data
- Improved accessibility of data** across enterprises fuels appetite for External Data – e.g. CRMs, decision engines, BI platforms, data platforms and more
- Increased complexity** due to tightening regulations makes it uneconomical to build capabilities in-house / no competitive advantage to build in-house

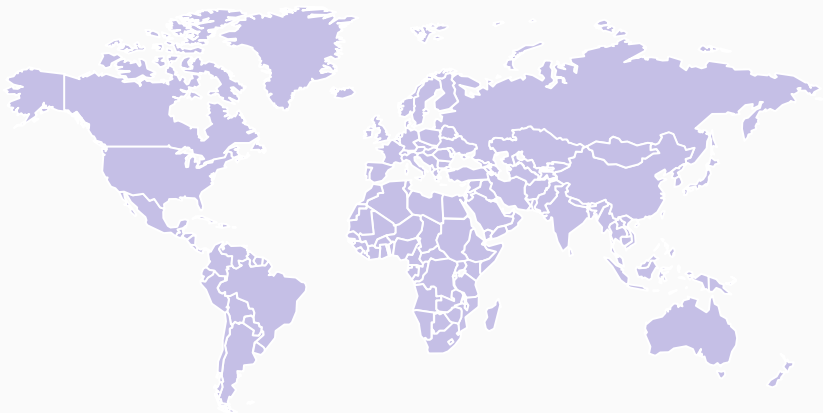
What are the challenges business and tech leaders face with external data?

Demyst's 2022 External Data Research

Demyst's conducted in-depth interviews on:

Challenges in getting value from external data

Best practices in maximizing external data ROI



3
Continents

5
Industries

27
Companies

Over 100
Data, Technology and Business Leaders

Pervasive in all customer interviews were three primary pain points:

Low trust in external data

Observability

Uncertain Performance

Controls

Discovery is cumbersome

Simple Recommendations

Effective Testing

Cataloging

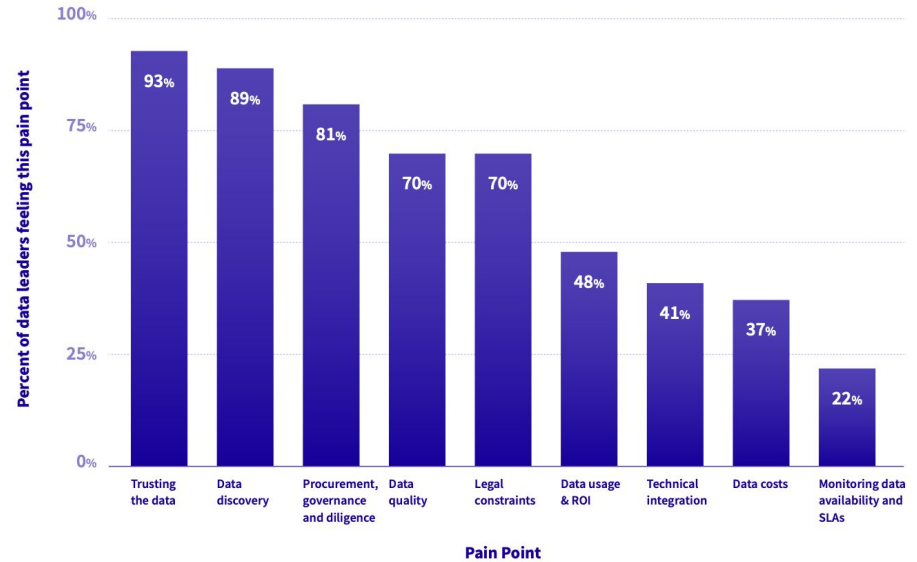
Compliant onboarding is complex

Compliant Testing

Best Practice

Ethics

External Data Pain Points Distribution for Data Leaders



Poll:
Which of the following is the top challenge for you when you think about external data?

93%
of Data Leaders

Low trust in external data

Data leaders must be able to answer the following questions to build trust

Observability

- ✓ What data are we buying?
- ✓ What are our costs?
- ✓ Is it being used effectively?
- ✓ What is business critical?

Performance

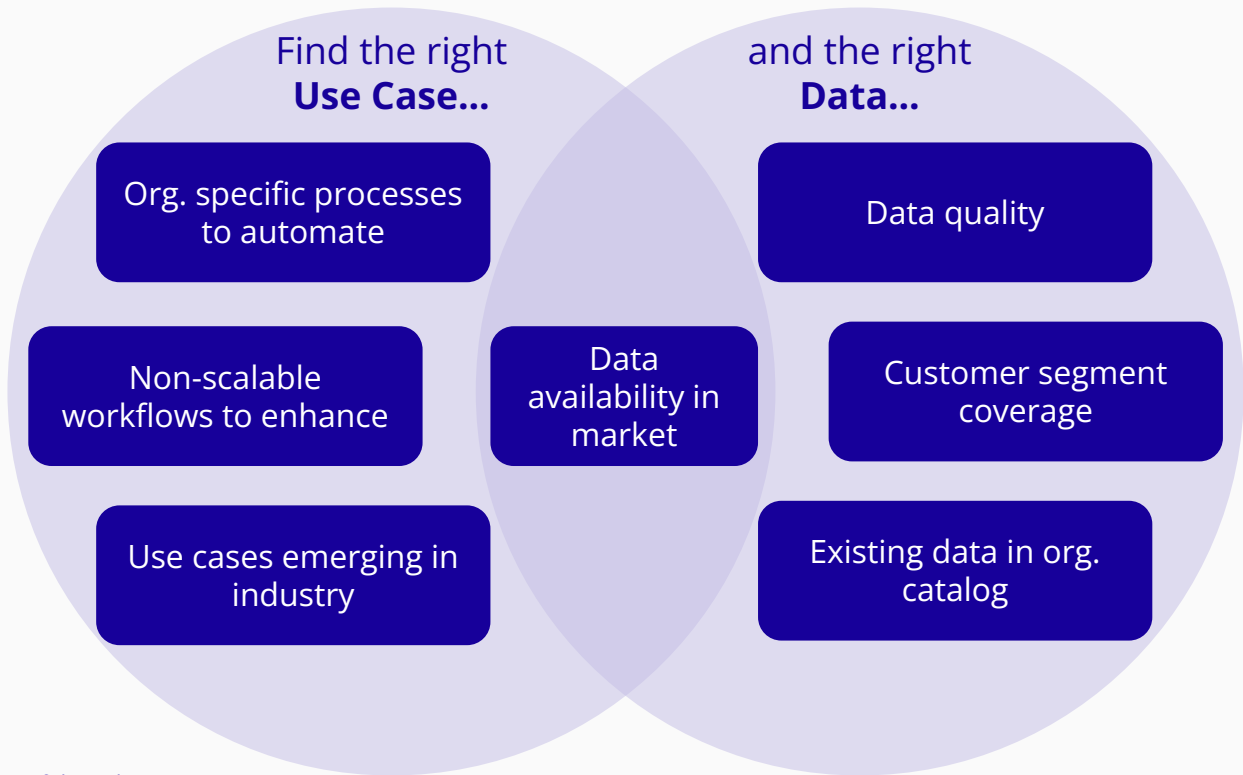
- ✓ What is this data's est. ROI?
- ✓ Is this source accurate?
- ✓ What happens if it goes down?
- ✓ How can we optimize it?

Auditability and controls

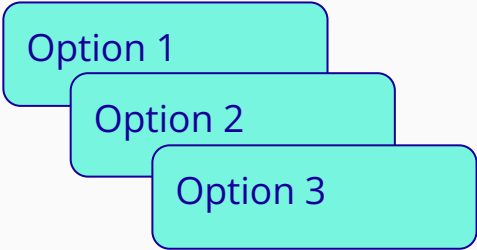
- ✓ Who is accessing each source?
- ✓ Is access in line with policy & budget?
- ✓ Have you implemented granular user access?
- ✓ Was the data actually deleted?

Discovery is cumbersome

89%
of Data Leaders



to create simple
recommendations...



with clear
trade-offs...

- ✓ Cost
- ✓ Performance
- ✓ Availability

81%
of Data Leaders

Compliant onboarding is complex

A fail early & learn fast mentality...



Can feel at odds with best practice compliance

Compliance, Procurement and Legal

- ✓ SME knowledge of industry best practice frameworks
- ✓ React to evolving regulations & review existing data

Ethics

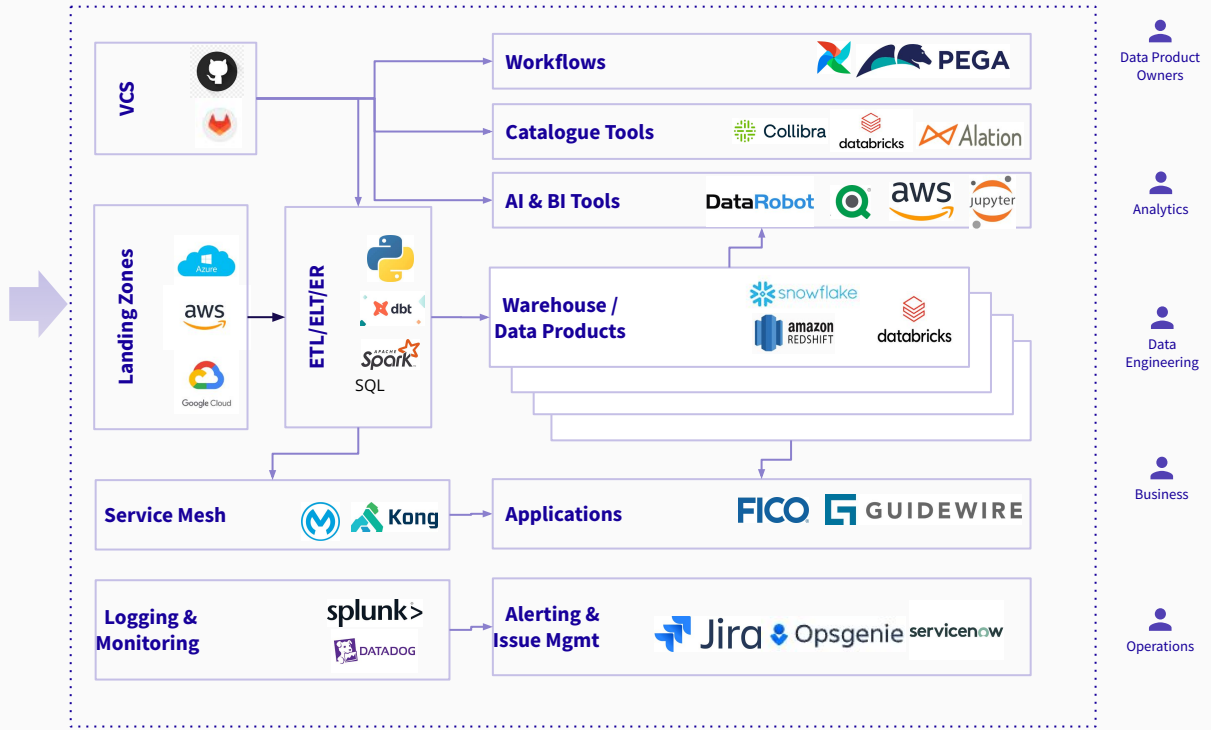
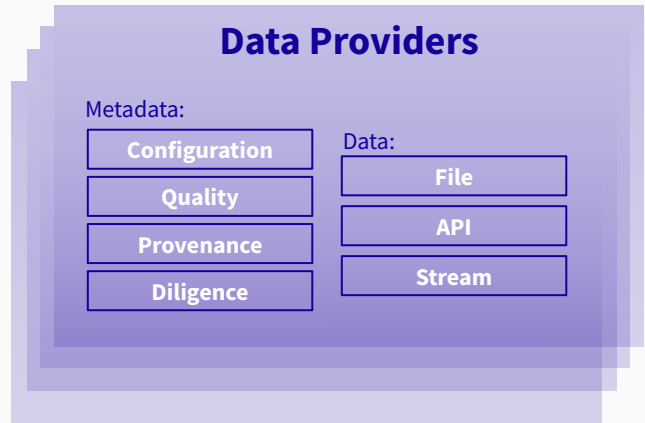
- ✓ Public scrutiny
- ✓ Consent
- ✓ Biases

Infosec

- ✓ Certifications
- ✓ Standards
- ✓ Uplift advice

**How does this data get deployed and
where is all this data going?**

» How does this compare to your technology environment?



▶ A cottage industry of niche capabilities is springing up around external data, highlighting various needs

Discovery Tools

External data technology providers that sell their products mostly focused on discovery in narrow markets / verticals



- ✗ Limited transparency into data sourcing, data lineage and provenance of data
- ✗ Typically a single market / use case / vertical focus with limited global coverage
- ✗ Tech first and lacking domain expertise in relevant use cases
- ✗ Typically does not provide flexible resell arrangements to accelerate data delivery

Data Providers

Individual data providers that sell their data directly to enterprises for domain specific use cases



- ✗ Enterprises still required to solve the discovery, certification and deployment challenges themselves
- ✗ Incentivized to position their data as the best even if it's not the reality. Little incentive or capacity for incumbents to innovate

Consultants

Consultants provide staff augmentation and advisory services to help implement use cases



- ✗ Often focused on one-off, non-recurring professional services projects
- ✗ Consultants lack the platform & technology to deliver External Data in a cost-efficient or scalable way

Pure Tech / Marketplaces

Data ops, marketplaces and/or catalogs that provide partial solutions; typically focused on internal data assets



- ✗ BYO contract required, do not address governance or procurement frictions
- ✗ Limited to data providers 'on platform' only, cannot meet holistic enterprise demand
- ✗ Nascent data domain expertise to guide adoption of data

What should I think about in the external data lifecycle?

› Companies need to navigate the external data lifecycle, which is different from internal data

Govern

Connector security, secret management, connector certification, connector usage rights and tracking



Discover

Data cataloging
Evaluation data ingestion
Testing environments



Deploy

Scheduling and orchestration
Blending, matching & transformation
Flexible yet consistent integrations



Maintain

Multi-vendor incident management
Upstream changes and versioning
Alerting, monitoring and reporting



Companies need to navigate the external data lifecycle, which is different from internal data



External data requires a different level of diligence to adhere to modern regulatory and infosec standards



External data deployment requires careful consideration of provider fit, workflows and costs

Data Deployment Activities

1

Optimising what you already buy

Understand spend distribution and promote re-use by indexing and cataloging what and how data is used today

Evaluate utilisation for what is bought vs what is used to identify potential redundancies and promote full utilization

Evaluate contracts to understand spend distribution, renewals and shutdown of contracts for data products that aren't used **and avoid duplicate spend!**

20%

2

Evaluate and leverage substitutes in procurement

Rapidly evaluate potential substitutes against incumbent sources to rank-order by cost, accuracy, and availability

Initiate contracting with new and established data vendors based on results of data evaluation efforts

25%

3

Re-optimize workflows around new sources

Leverage data platforms to simplify data access and enable rapid switching to new data sources as they become available

Implement waterfall options to leverage multiple data sources where fill-rate and coverage must be high

Switch sources to lower cost or favourable substitute(s)

Become API driven and pay only for the data you need, when you need it

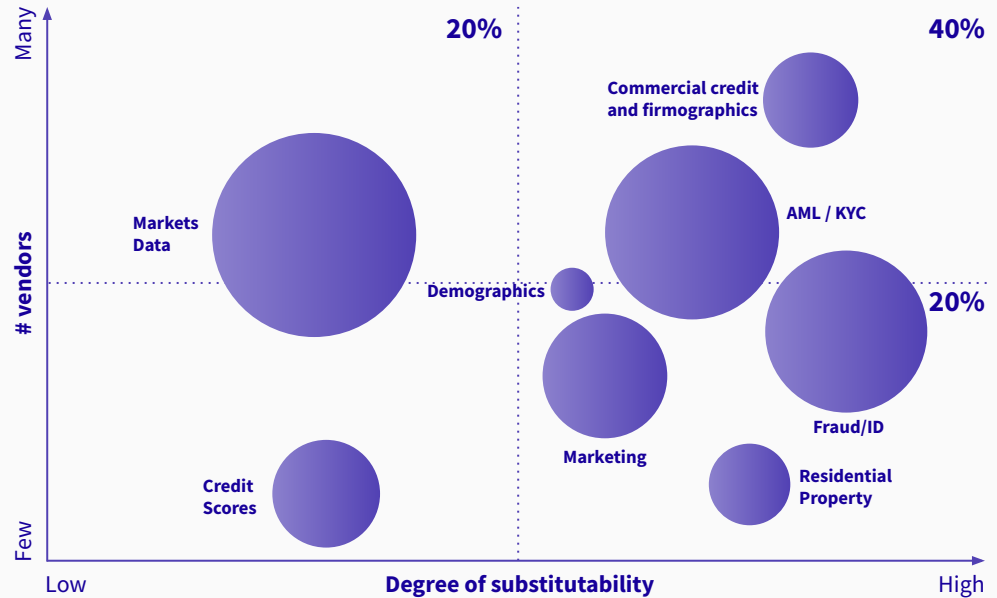
55%

Likely %
of savings

➤ The explosion in data providers (more than 2000) requires a constant optimisation to stay competitive

— Cost reduction targets based on potential substitutability, as well as current percent contribution to overall data spend by vendor / data type (e.g., may only need 5% reduction on data source or domain that accounts for 25% of total data spend)

● Size of bubble denotes \$ spend



Emerging substitutes in every data domain allow for far more data, often for lower total cost

Previously:
\$10MM spend



Attributes: <1,000

Today:
\$6MM spend on alternatives



Attributes: 10,000+

Demyst Client example

Quotes to operationalise substitutable data:

Vendor A (\$240k+, restricted use)

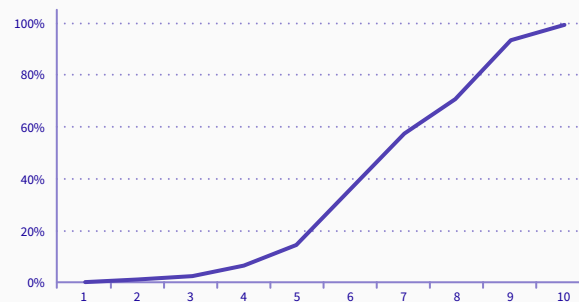
Vendor B (\$180k+, restricted use)

Vendor C (\$30k, restricted use)

Vendor D (\$30k, unlimited re-use)

Data source predictor efficacy

(near perfect substitution vs incumbent at 40% lower price)



Across 100s of integrations, these are examples of common challenges of maintaining external data

Common upstream issues:

- Upstream data vendor changes
 - File / schema changes
 - New file(s) / new attribute(s)
 - Changes to file or attribute naming conventions
 - Changes to data formats e.g., DDMMYY to DDMMYYYY
 - Changes to package mapping/ file nesting structure
 - Changes to transfer method e.g., IP address change for SFTP, or preference to use S3 rather than SFTP or Delta Sharing
- Upstream data vendor issues:
 - Special characters present that impact the data structure
 - Missing data e.g., null fields for a primary key
 - Files not available when expected

Common downstream issues:

- Downstream client system changes
 - Delivery to a new landing zone (including rewiring of existing flows if a system is decommissioned)
 - Ability to push a single feed to multiple endpoints / formats
 - Deliver data feed updates at a new cadence if required
- Downstream client system requirements
 - Consistent file name structures (across data providers)
 - Transformation of files to supported types e.g., excel sheets to .csv or PDFs to .csv or other common formats
 - Encryption of files and/or masking of attributes when required, e.g. removal of non-approved black box scores
 - Encoding requirements e.g., converting the encoding of the file from cp1252 to utf8 (or as required to other standards)

Poll:
**What is the most important next step
for your company to get more value
from external data?**

To recap...

Enterprises face significant obstacles to realize value from thousands of fragmented External Data providers



Data Discovery

- Finding fit for purpose External Data is frustrating and costly
- Testing data providers can take over 6 months and requires sensitive data sharing and specialized skills
- Many costs are born upfront but most POCs fail to proceed



Governance

- Enterprises need to consider increasing obligations for External Data such as compliance, provenance, data ethics, privacy regulations and more
- Responsibility sits with the enterprise to evaluate and manage data compliance with a tapestry of regulations but it's a moving target



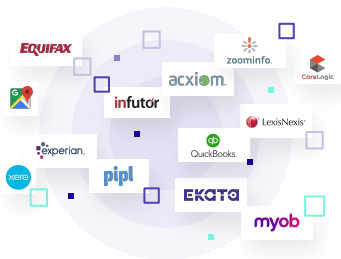
Deployment

- Typically takes 6-12 months to implement a single data source
- Complexity and resource requirements increase as new data sources are added, each with their own architecture and requirements
- Data engineering capacity required to develop resilient processes are scarce



Maintenance and Support

- Enterprises need to maintain data solutions (risk monitoring, uptime management, data drift, changes to data schemas, regulations and quality)
- Maintenance necessitates constant sourcing for improvements, but many enterprises are not adequately resourced
- Data providers are often brittle, requiring failover and redundancy




Highly fragmented data needs

Costly and complex data governance risks



Follow-up research you will receive

 Demyst

External Data Management - Applying the CDMC Framework

Executive summary: This document is a response to the exceptional [CDMC framework](#) and provides suggestions on key aspects that are relevant to external data use cases. Broadly, the key takeaway is that data leaders need to consider the risks, responsibilities and opportunities associated with harnessing external data and, across all aspects of the framework, how external data is necessarily a centrally managed function. Leading enterprises are, with the right investments and processes, establishing external data hubs and teams that finally start to deliver on the promise of value from external data.


By: Mark Hookey, CEO of Demyst

Data Management


What is "Data Management"? Most of my conversations are with data leaders in enterprises and in every conversation I learn about a different slant. It's a broad concept and can mean many things to different people.

Some talk about data management in the context of having clear ownership and accountability for data assets. Others speak tactically about the mechanics of how to get value from data. Others mention information security concerns and risk management. It is an emerging field and, in reality, data management means all of the above and more.

The EDM Council produces an [exceptional framework](#) that covers the end to end of managing data.



CONFIDENTIAL | 1

 Demyst

Proprietary and Confidential

External Data Ops - Top 20 Best Practices

Many Demyst clients are embarking on a strategic program of change, involving many vendors and data products where the need for a central data ops platform is clear. However, many clients are solving something much more tactical which involves deploying data from a single source. If they know the source already, it's not unusual that these clients will try to integrate directly, or use a general purpose API gateway for this integration. On the surface, it makes sense, why add an aggregator like Demyst in the middle? Why add an extra hop and potentially cost when the individual data vendor has a clean API?

It's typically not that simple. Many enterprises face unexpected delays integrating with and managing external data connectors, even if it's just one. Our mission ultimately is to increase the compliant impactful use of external data in the ecosystem, whether or not via us. To that end, this document seeks to lay out the best practise of what enterprises need to manage for external data connectivity.

Requirement	Reason
1. Define and confirm data suitability and quality	<ul style="list-style-type: none"> - Confirm suitability for business use case - Test integration approach and address teething issues
2. Defined types for input and output fields (automatic detection)	<ul style="list-style-type: none"> - Accelerate integration - Allow immediate detection and capture of data quality issues - Allow for data errors to be handled in a consistent manner - Support data governance requirements
3. Optimized matching (entity resolution to maximize match rate and quality for entity lookup)	<ul style="list-style-type: none"> - Maximize match rate - Maximize quality - Tackle fuzzy matching
4. Allow for batch, streaming, and realtime operations within same interface	<ul style="list-style-type: none"> - Rapid deployment and use - Process batches for testing and reconciliation - Take receipt of data for eventual copy and analytics within internal data lake - Allow for retry / failover / replay - Accelerated deployment between test / production for agile release schedules

 Demyst

Extracting Business Value from

External Data in 2023

A global survey of 100+ Data, Technology and Business Leaders



Questions?



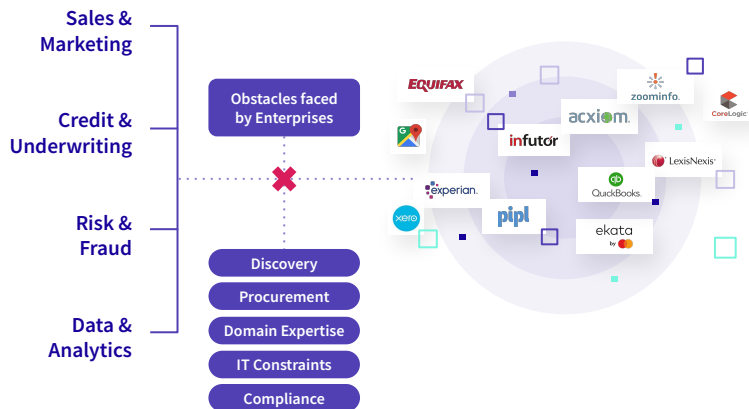
EDM Webinar 

External Data, Done Right



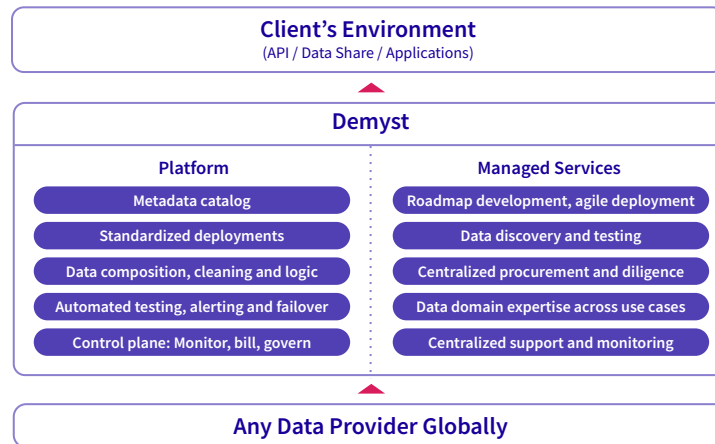
Enterprises are stuck

with growing external data costs and complexity, but limited forward progress



Demyst is the solution

Leading enterprises trust Demyst to centralize and manage all their External Data needs



Why Us?

- ✓ Increased success rates for use cases, with enterprise external data discovery
- ✓ Improved risk outcomes with centralized ingestion and management
- ✓ Reduced TCO and increased usage with centralized procurement and deployment

Speak to a Demyst expert today at

demyst.com/contact

Thank You

Scott Albin

GM, International & CSO
salbin@demystdata.com

Join our membership community of companies...



The screenshot shows the EDM Council website homepage. At the top, there is a navigation bar with links for Calendar, DataVision, Knowledge Portal, Guest Registration, and Contact. Below this is a secondary navigation bar with links for Membership, Frameworks, Training, Engage, Innovation, About, Sign in, and a prominent "Join now" button. The main content area features a large blue banner with the headline "Global Advocates for Data & Analytics Management" and a sub-headline "The leading global trade association providing best practices, standards and education to data and business professionals in our data-driven world." A "What we do" button is located below the banner. To the right, there is a "TODAY'S HIGHLIGHTS" section with three items: "Bank of Valletta becomes the newest member to join the EDM Council", "EDM Council welcomes Webber Wentzel as its newest member", and "Lion Group joins EDM Council as its newest member". At the bottom of the page, there is a section titled "Join a vibrant community of 25,000+ business leaders, CDOs, and data and analytics professionals across all industries." followed by logos for Bank of England, NOVARTIS, HSBC, AWS, Schneider Electric, Microsoft, Google, and AEGON. Two buttons are present: "Explore membership" and "See all 350+ member organizations".



350+ Member Firms
Cross-industry,
including Regulators



25,000+
Professionals



Worldwide
Americas, Europe,
Africa, Asia, Australia

edmcouncil.org





EDM Webinar 

Thank you!

FOR MORE INFORMATION:

Scott Albin
GM, International & CSO
Demyst
salbin@demystdata.com

 Demyst

 EDM Council