

# EDM Webinar

## Building an External Data Hub

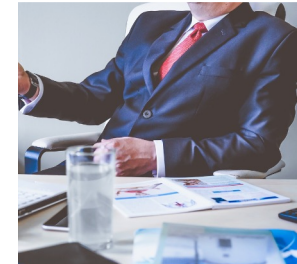
*A conversation with*



**Mark Hookey**  
Founder & CEO  
Demyst



**Jim Halcomb**  
Head of Product Management  
EDM Council



# Today's speakers

Moderator



**Jim Halcomb**

Head of Product Management  
EDM Council



Speaker



**Mark Hookey**

Founder & CEO  
Demyst





# External Data, Done Right

Enterprises are building powerful **Data Hubs**, however harnessing external data at scale is hard

The complex and often fractured dynamics of accessing and utilising external data leave **valuable gains unrealised**

Demyst is here to address this pain, providing the **trust and ease of use** that progressive enterprises demand

## External data is painful



Many sources & formats

Many use cases & end users

Ops friction & constraints

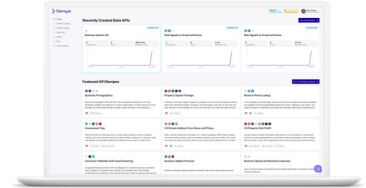
Technology limitations

Low trust

Inhibitive costs



*Drive your business with trusted external data, at scale*



### Streamline connectivity

ease discovery, testing and procurement cycles

### Generate trust in external data

address complex certification & governance challenges

### Centralise external data ops

Create visibility, implement controls and optimise

### Drive efficient utilisation

address deployment challenges, and reduce TCO

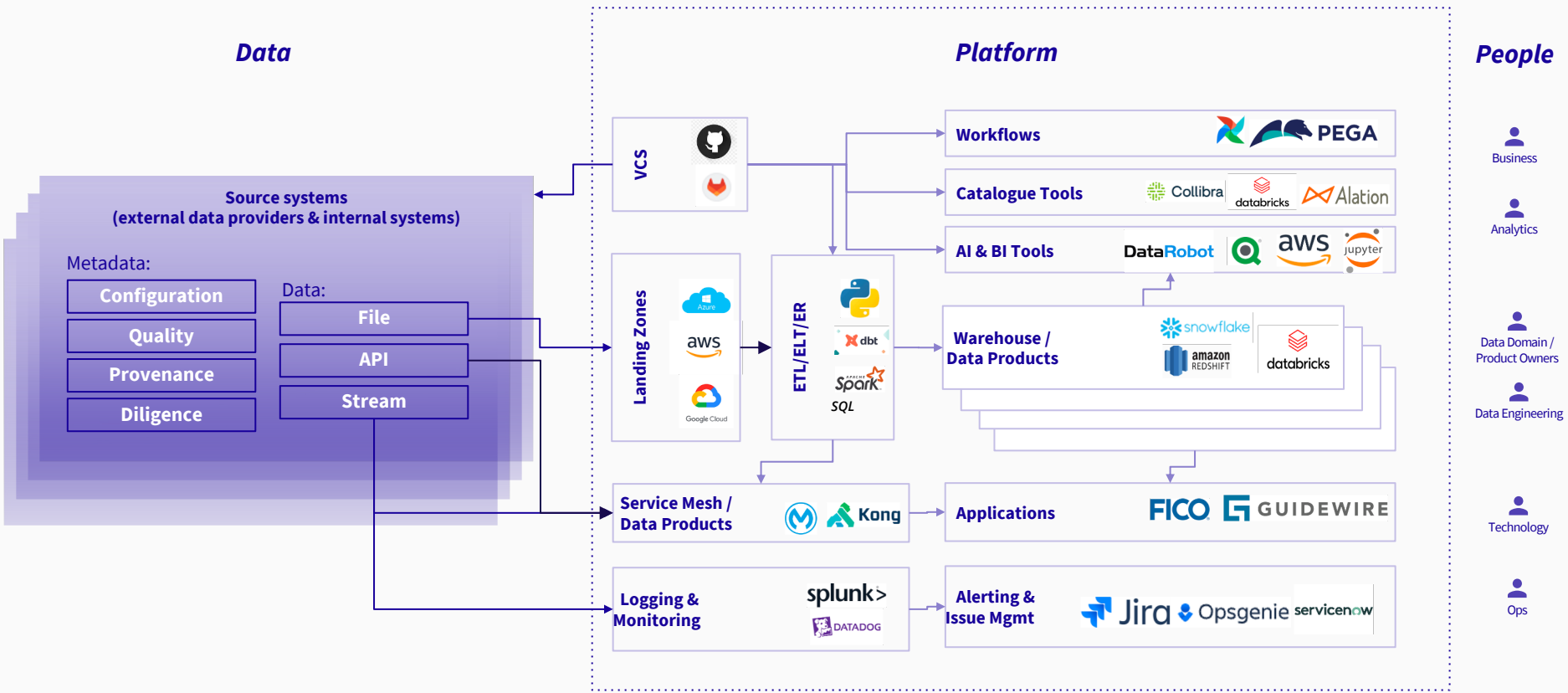
### Dramatically improve ROI

solve both operational and analytical use cases, at scale

## > Agenda

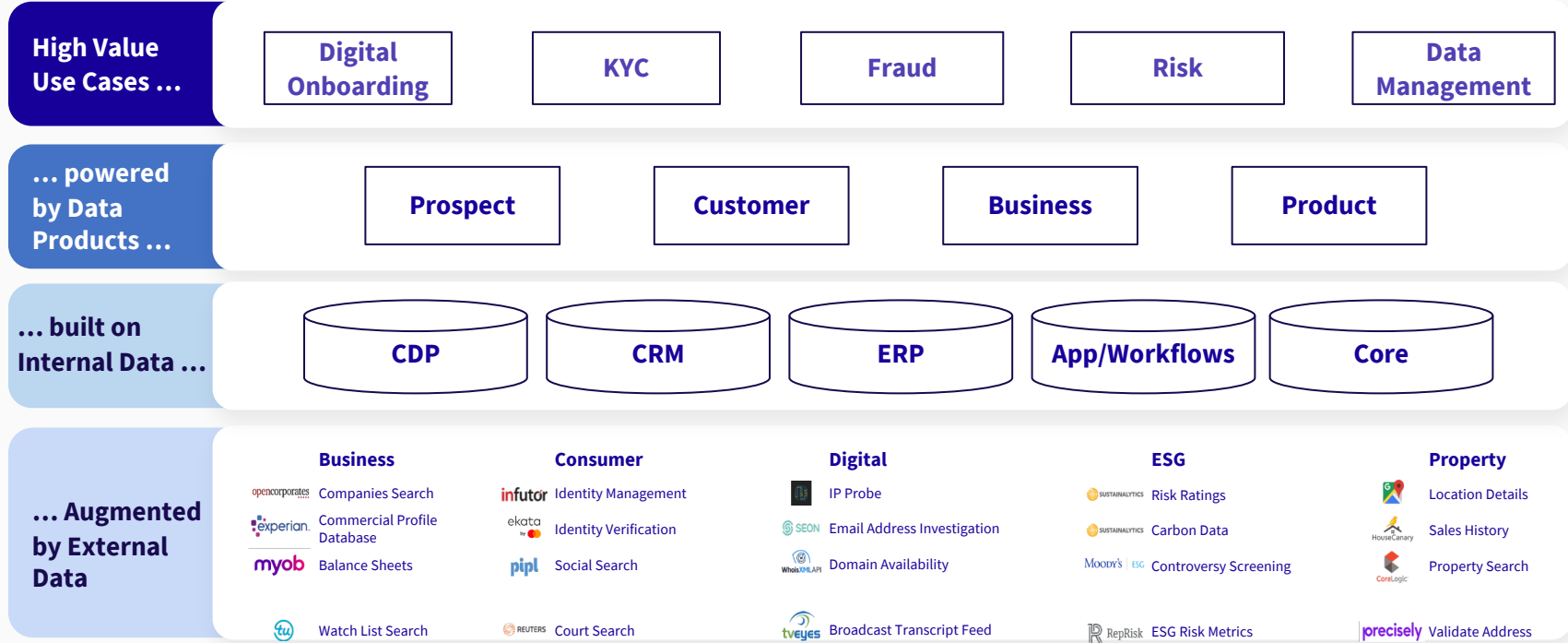
1. *Enterprise Data Hubs today*
2. *External Data Idiosyncrasies*
3. *How To*
4. *Q&A*

# » A typical target state enterprise data hub ...

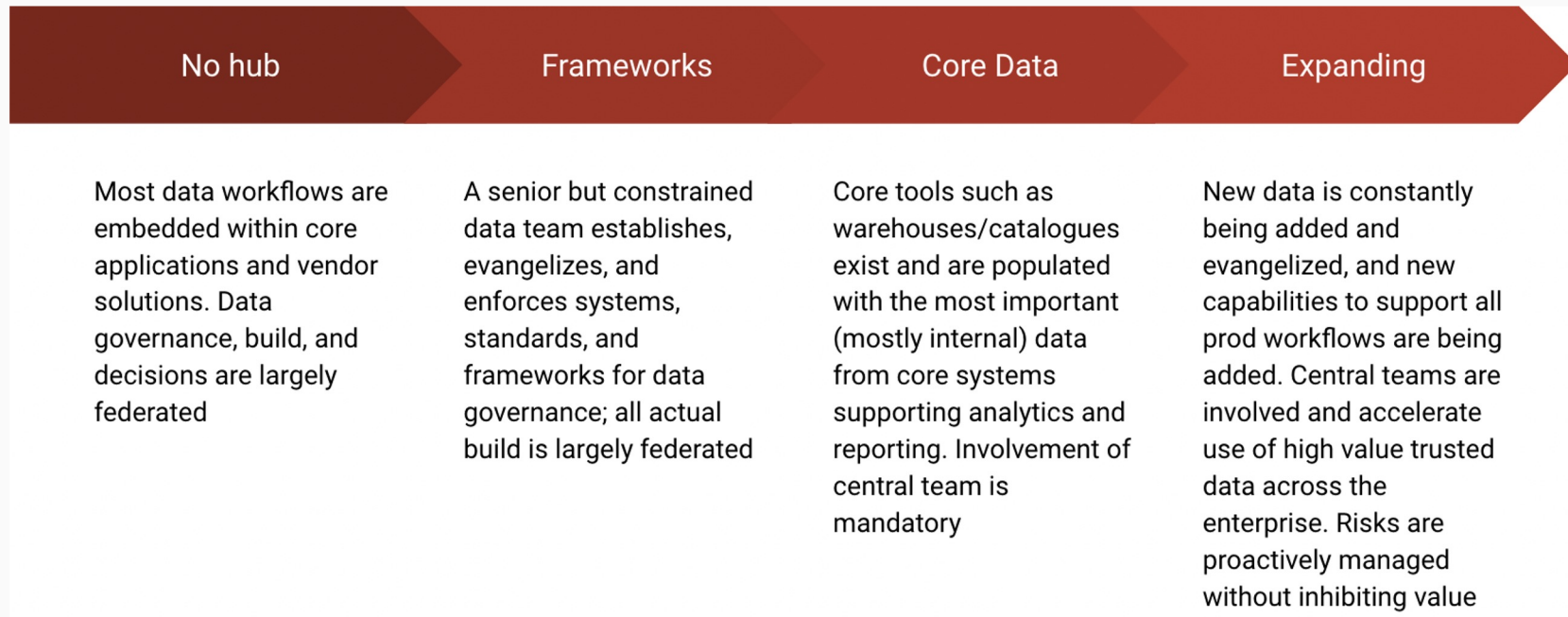


# » ... fueling high value use cases

## Financial services example



## Enterprises are at varied stages in developing their Data Hubs



# Poll: How mature is your Data Hub?



# Key challenges with external data\*

## Low trust in external data

Observability

Uncertain  
Performance

Controls

## Discovery is cumbersome

Simple  
Recommendations

Effective  
Testing

Cataloging

## Compliant onboarding is complex

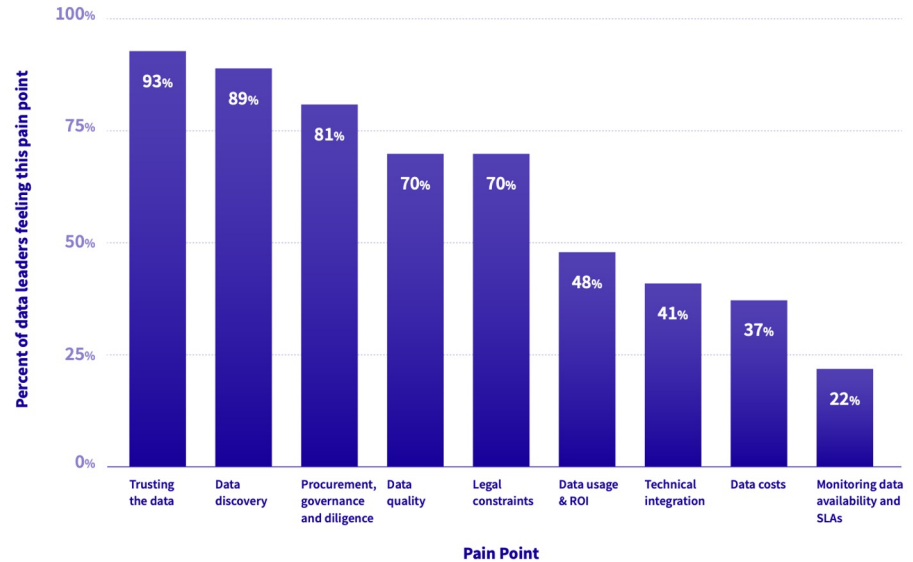
Compliant Testing

Best  
Practice

Ethics

Confidential

External Data Pain Points Distribution for Data Leaders



\* Src : Demyst market research from over 100 data leaders

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

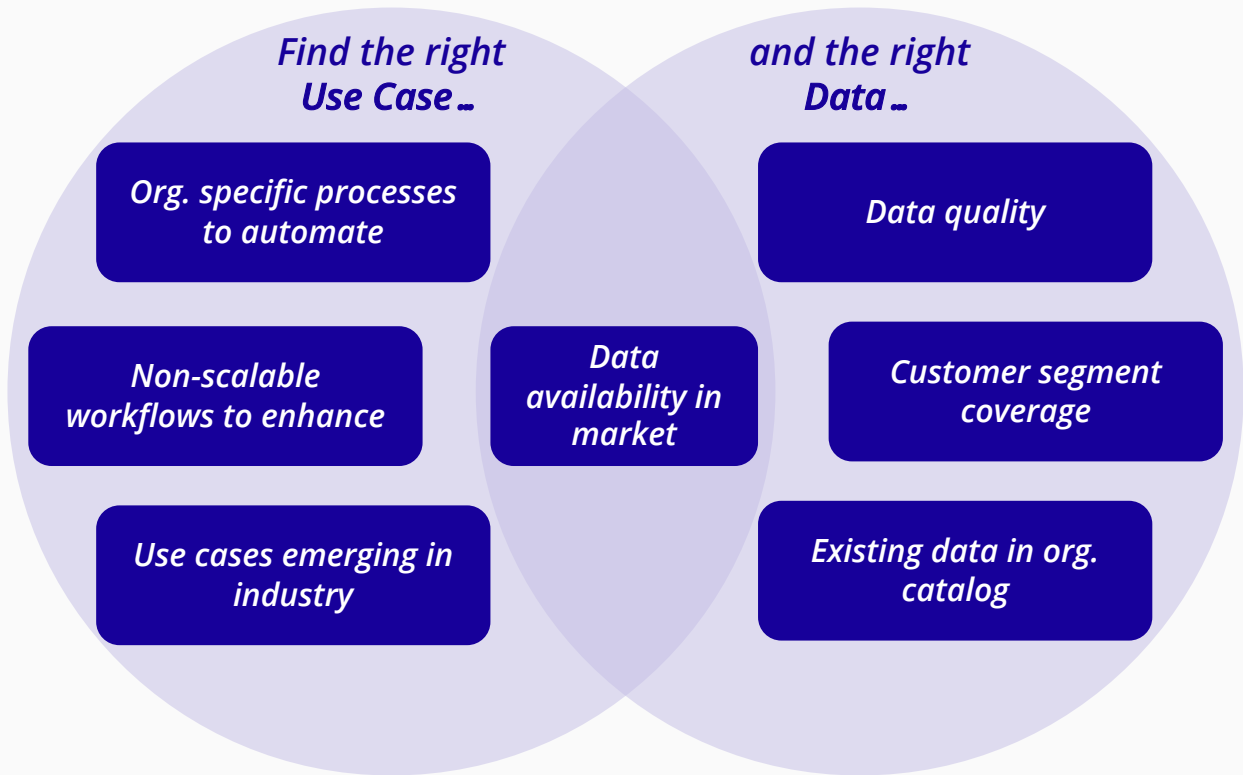
- ✓ *What is this data's est. ROI?*
- ✓ *Is this source accurate?*
- ✓ *What happens if it goes down?*
- ✓ *How can we optimize it?*

### Auditability and controls

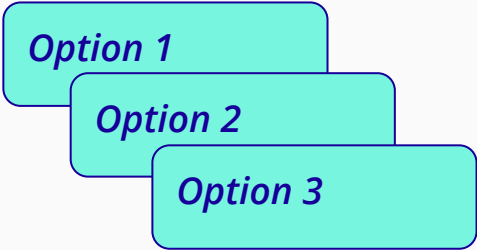
- ✓ *Who is accessing each source?*
- ✓ *Is access in line with policy & budget?*
- ✓ *Have you implemented granular user access?*
- ✓ *Was the data actually deleted?*

# Discovery is cumbersome

89%  
of Data Leaders



to create simple  
recommendations..



with clear  
trade-offs..

- ✓ Cost
- ✓ Performance
- ✓ Availability

# Compliant onboarding is complex

81%  
of Data Leaders

*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**

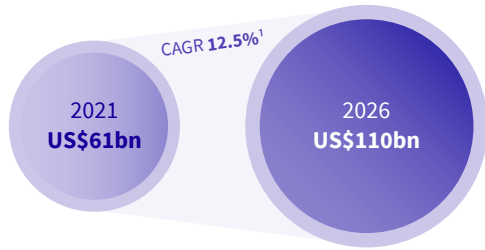
- ✓ Public scrutiny
- ✓ Consent
- ✓ Biases

**Infosec**

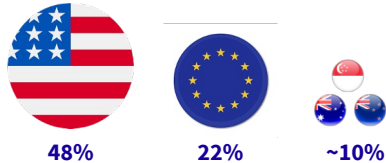
- ✓ Certifications
- ✓ Standards
- ✓ Uplift advice

... in spite of this, the external data market will represent >\$110b in spend by 2026

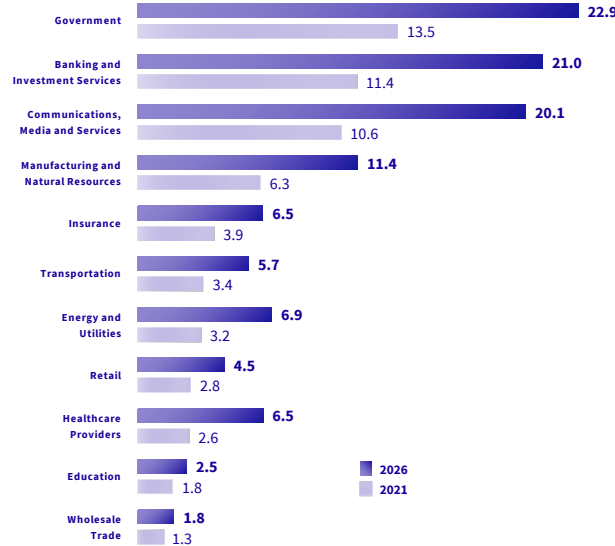
### Size and industry distribution of the market for External Data



#### Top 3 regions<sup>2</sup>



#### External Data Market by Industry (\$bn)

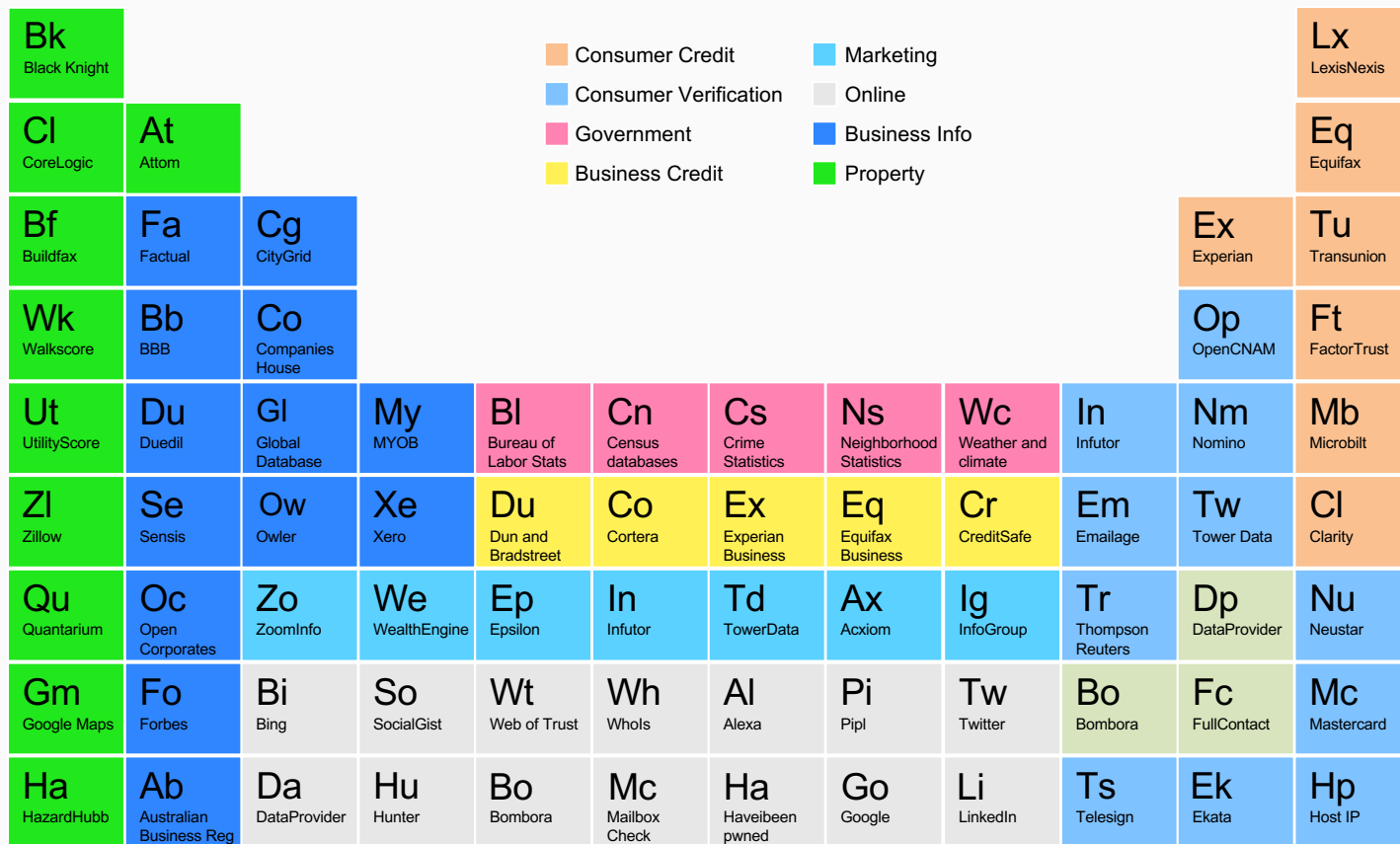


### Familiar industry themes

- 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

Poll: What types of external data are centrally governed?

The explosion in data providers (thousands, with tens of thousands of products) requires a constant optimisation to stay competitive

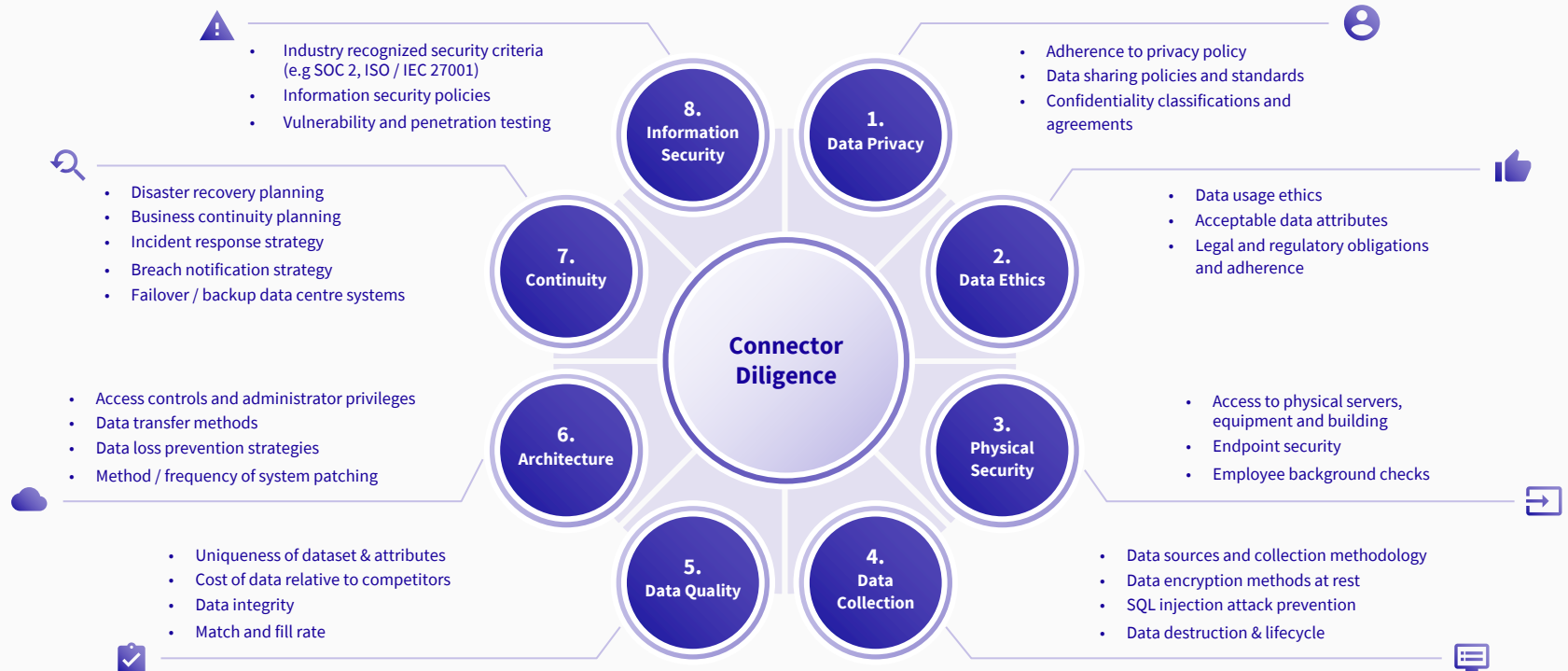


## › Aligning external data scope requires ...

- 1. The right people :*
  - 1. Data stewards / hunters owning data domains*
  - 2. Tight partnership with “the business” and “use cases”*
- 2. Clear priorities :*
  - Education, awareness, brainstorming, and vendor / SME engagement*
  - Internal awareness and alignment enabled by a robust catalogue*
  - Clear top down strategic priorities*
  - ... leading to a connector backlog*



# Enterprises need to agree clear standards tailored for external data that drive metadata requirements



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

### *New connector onboarding (example)*

1. Find the right data from the broad and complex landscape
2. Assess the data for compliance and governance
3. Acquire sample data for evaluation purposes
4. Evaluate the data for fit and quality
5. Determine approach for useability in a Production solution
6. Approve solution for Production usage
7. Conduct compliance activities and procure the data
8. Populate data catalogue with required metadata
9. Engage data provider and conduct technical requirement gathering
10. Construct pipelines (either API or Flat File) for repeatable ingestion to data hub landing zone
11. Transform and prepare data for consumption by target applications or to create a "data product"
12. Connect the external data source to realtime microservices or directly into applications

### *Connector monitoring (examples)*

13. Monitor and manage data provider performance
14. Monitor and manage data quality and effectiveness
15. Monitor for problems, generate relevant alerts, and ensure resolution of external data related issues
16. Ongoing engineering support, maintenance, and uplift for external data.
17. Monitor and manage ongoing legal and regulatory compliance issues related to external data
18. Manage procurement, spend, and billing for all external data providers
19. Ensure external data remains available and optimize usage of external data
20. Scout for options to improve with new complementary connectors

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

## Example 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
    - Missing data, missing files, special characters
    - 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
  - SLA / uptime issues
  - Information security events
  - Data drift

## Example downstream issues:

- Downstream client system changes
  - New workflow creation (use cases (e.g. decision engine) or data systems (e.g. landing zones))
  - Expose datasets to multiple endpoints / formats
  - Changed client system requirements :
    - i. Frequency changes
    - ii. File schema changes
    - iii. New conversion requirements
    - iv. Encryption / masking / encoding requirements
    - v. New workflow creation (use cases (e.g. decision engine) or data systems (e.g. landing zones))
  - Updated POCs / analytics to continue to improve lift

# ➤ How should I develop my external data hub (EDH)?

## Discover

Which upstream connectors are used today? What's the roadmap? How are business analytics teams supported?



## Govern

How are data governance objectives applied to external data? How are upstream connectors certified and maintained?



## Deploy

What are the downstream deployment architectures supported? How does the team manage the deployment lifecycle?




## Maintain

How are production data workflows maintained? Proactive optimization and reactive incident management?



# Follow-up research you will receive



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



## Extracting Business Value from External Data in 2023

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



# Questions?



**EDM** Webinar 

---

# Join EDM Council and our membership community of companies...



The screenshot shows the EDM Council website homepage. At the top, there is a navigation bar with links for Membership, Frameworks, Training, Engage, Innovation, About, Sign in, and a prominent 'Join now' button. The main header features the EDM Council logo and a large headline: 'Global Advocates for Data & Analytics Management'. Below the headline, a sub-headline reads: '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 visible. To the right, a 'TODAY'S HIGHLIGHTS' section lists three new members: Bank of Valletta, Webber Wentzel, and Lion Group. At the bottom of the screenshot, a banner states: 'Join a vibrant community of 25,000+ business leaders, CDOs, and data and analytics professionals across all industries.' Below this banner are logos for Bank of England, NOVARTIS, HSBC, AWS, Schneider Electric, Microsoft, Google, and Egon. 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](https://edmcouncil.org)



**EDM** Webinar 

Thank you!

**FOR MORE INFORMATION:**

Mark Hookey  
Founder & CEO  
Demyst  
[mhookey@demystdata.com](mailto:mhookey@demystdata.com)

