EA Forum 18 May 2022

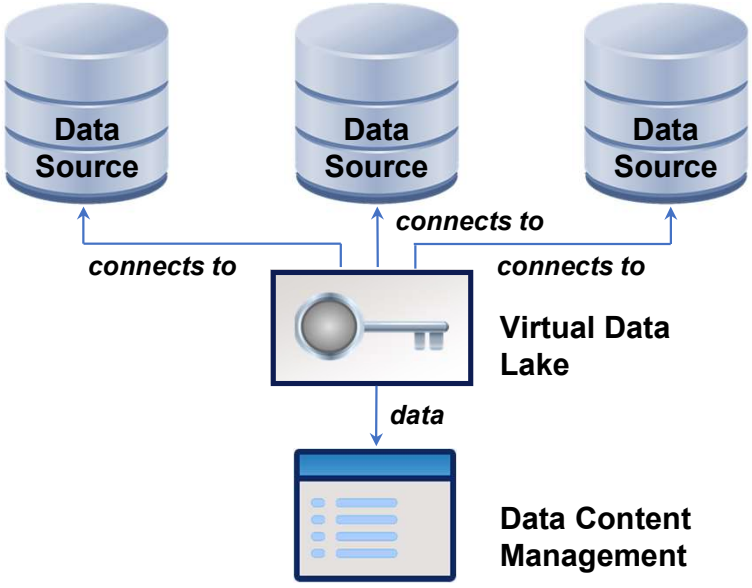
Why is Data Integration so Difficult?

Dr Chris Harding
Lacibus

LACIBUS

About Me

Chris Harding



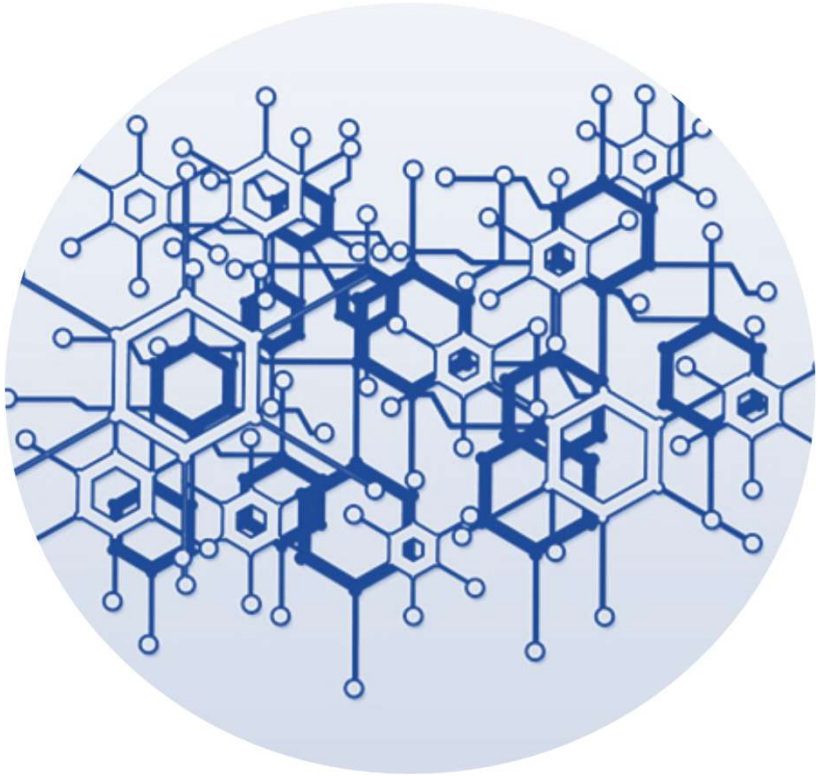
LACIBUS

Disclaimer

The views expressed in this webinar are those of the presenter. They are not necessarily the views of The Open Group.

LACIBUS

The EA Forum Webinar

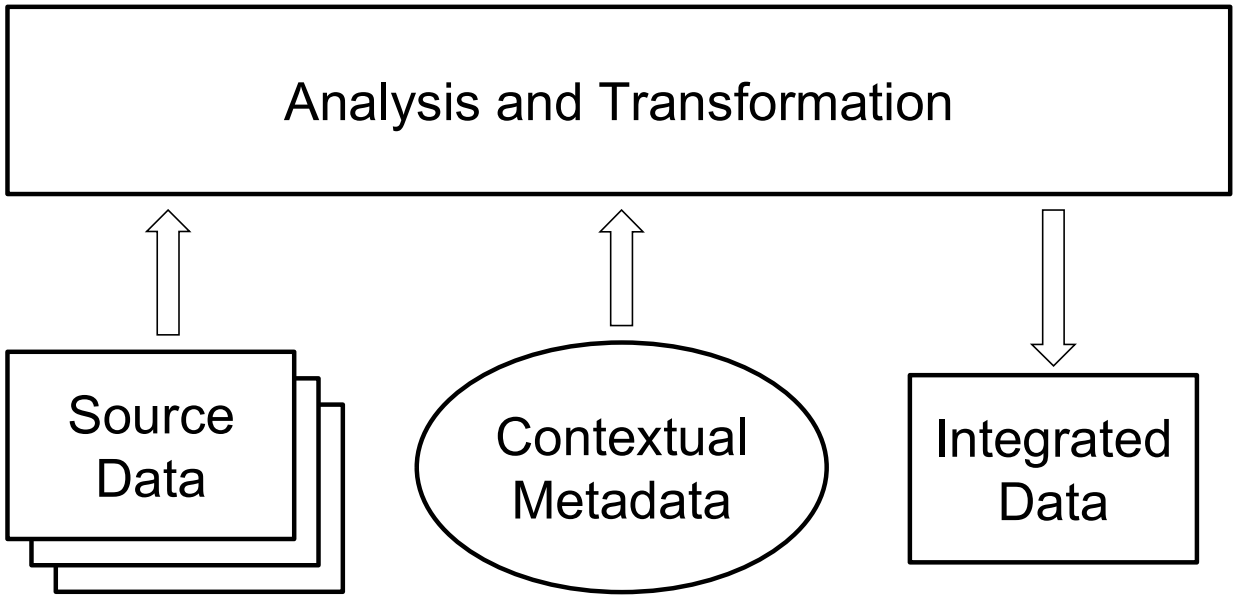


- Data Integration
- Relevant Work from The Open Group
- Future Trends

LACIBUS

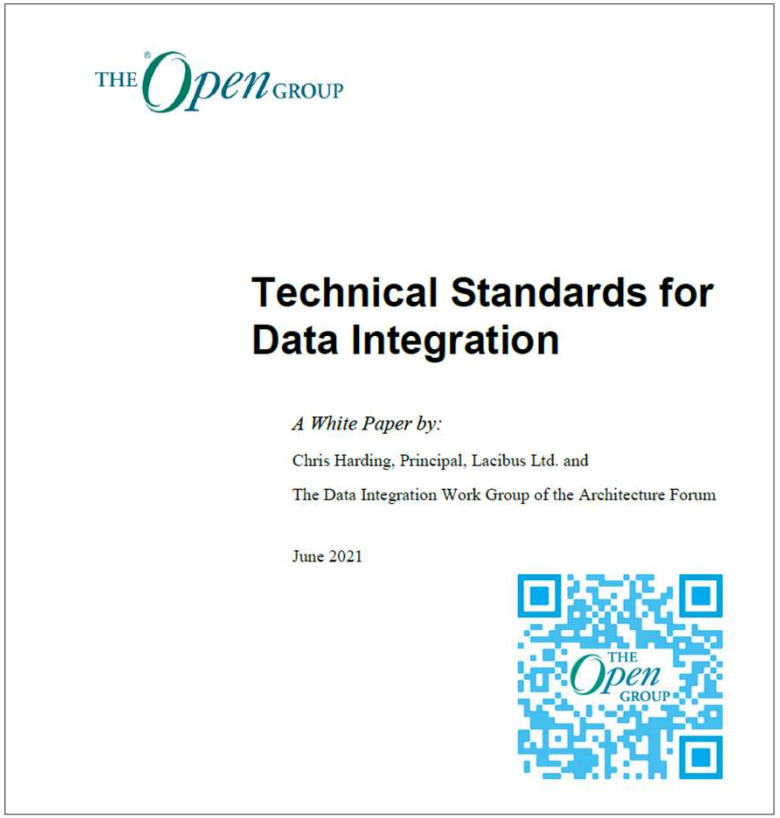
Data Integration

Data integration involves combining data from varying sources to provide users with a unified single view



LACIBUS

Technical Standards



THE Open GROUP

Technical Standards for Data Integration

A White Paper by:

Chris Harding, Principal, Lacibus Ltd. and

The Data Integration Work Group of the Architecture Forum

June 2021

LACIBUS

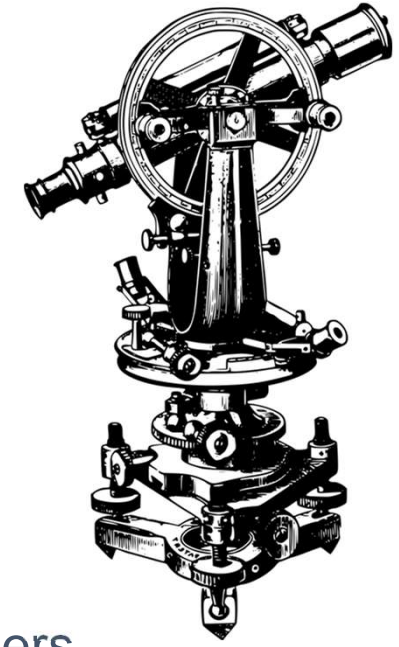
Data Integration Survey

23 responses from Architecture Forum members

- Mostly external consultant architects
- Accelerating speed of discovery and delivery of data (e.g. DataOps) biggest gap for improving data use

592 responses from AEA members

- Mostly internal architects
- Lack of governance and stewardship biggest gap for improving data use
- Other results quite similar to results from Forum members



LACIBUS

Data in the Enterprise

- Business leaders mostly view data as a strategic corporate asset
- Data use is often localized by business unit
- Some data is in the Cloud, some on-premise
- Overall data quality is mixed: some excellent, some terrible, most somewhere in between
- There are often islands of “quality” data with differing management regimes

LACIBUS

Data Integration

- Analytics and decision support are important reasons for integrating data, but operations (transactional) most important
- Data to be integrated most often from databases, but surprising amount from electronic documents, and some from IoT and social media
- Data to be integrated mostly from within the enterprise, but often with some external data

LACIBUS

Pain Points

Lack of commitment from business units

- Departments don't want to share their data
- No understanding of the business value
- Difficult to find the data
- SMEs don’t explain the data

LACIBUS

Pain Points

Lack of commitment at corporate level

- Enterprise data integration not seen as a business initiative that justifies investment

LACIBUS

Pain Points

Heterogeneous sources and tool-stacks

- Different formats with different processing needs
- Different data platforms
- Web services with different languages and operating systems
- SaaS providers with different interfaces

LACIBUS

Pain Points

Conflicting data models

- No enterprise data model
- The data is not standardized
- It includes data from legacy and open systems
- It incorporates external data that is ontologically and taxonomically un-normalized and/or at odds with internal data

LACIBUS

Pain Points

No culture of data management

- No data governance task force
- No policies that speak directly to data
- Data quality issues, including
  - Inconsistent data from different sources
  - Duplicate records

LACIBUS

Why is Data Integration so Difficult?

- Integrations involving multiple departments must have department and corporate backing
- Business case must be clearly explained to all stakeholders
- Difficulties of differing data formats and interfaces must be addressed
- Poor data quality must be addressed
- Access control must be addressed, especially for PII

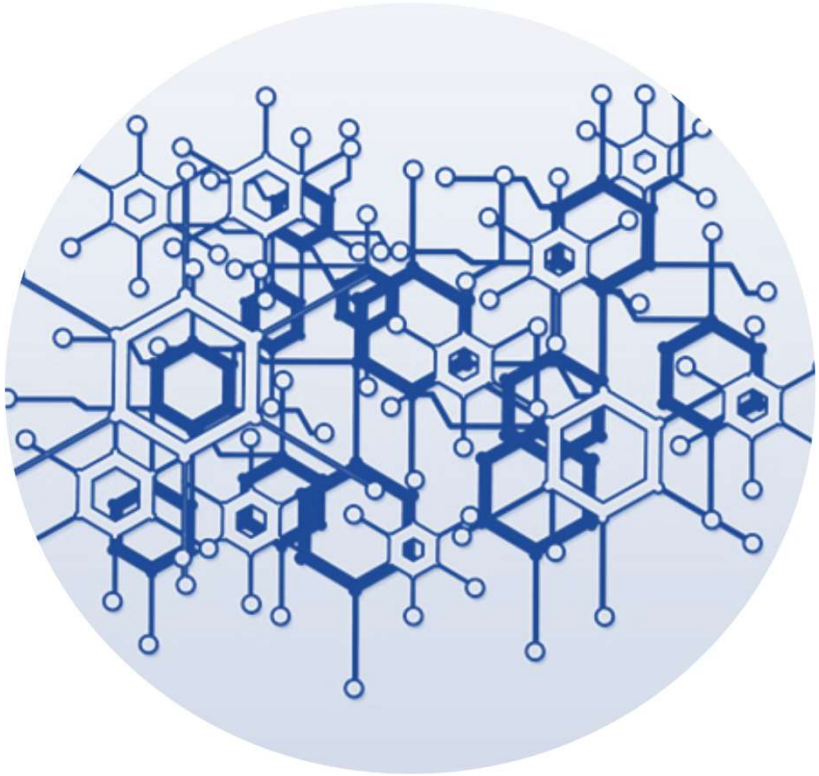
LACIBUS

How Can We Make Data Integration Easier?

- By applying architecture methods and techniques

LACIBUS

Relevant Work from The Open Group



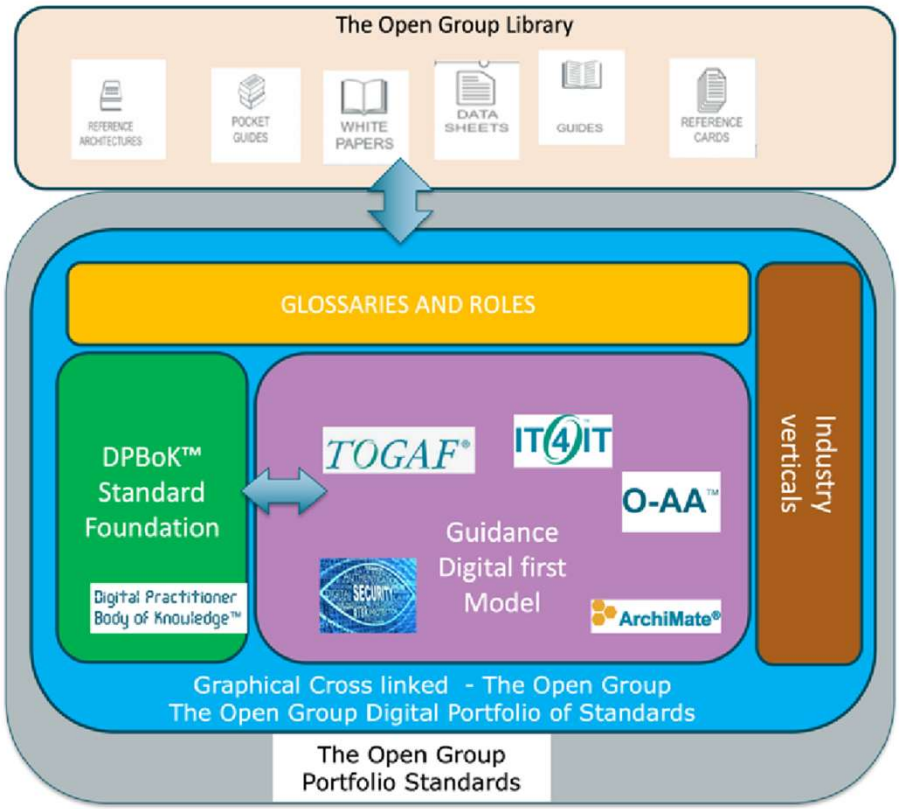
- Method standards
- Business Domains
- Platforms
- Other Publications
- Data Integration Work Group

LACIBUS

Method Standards

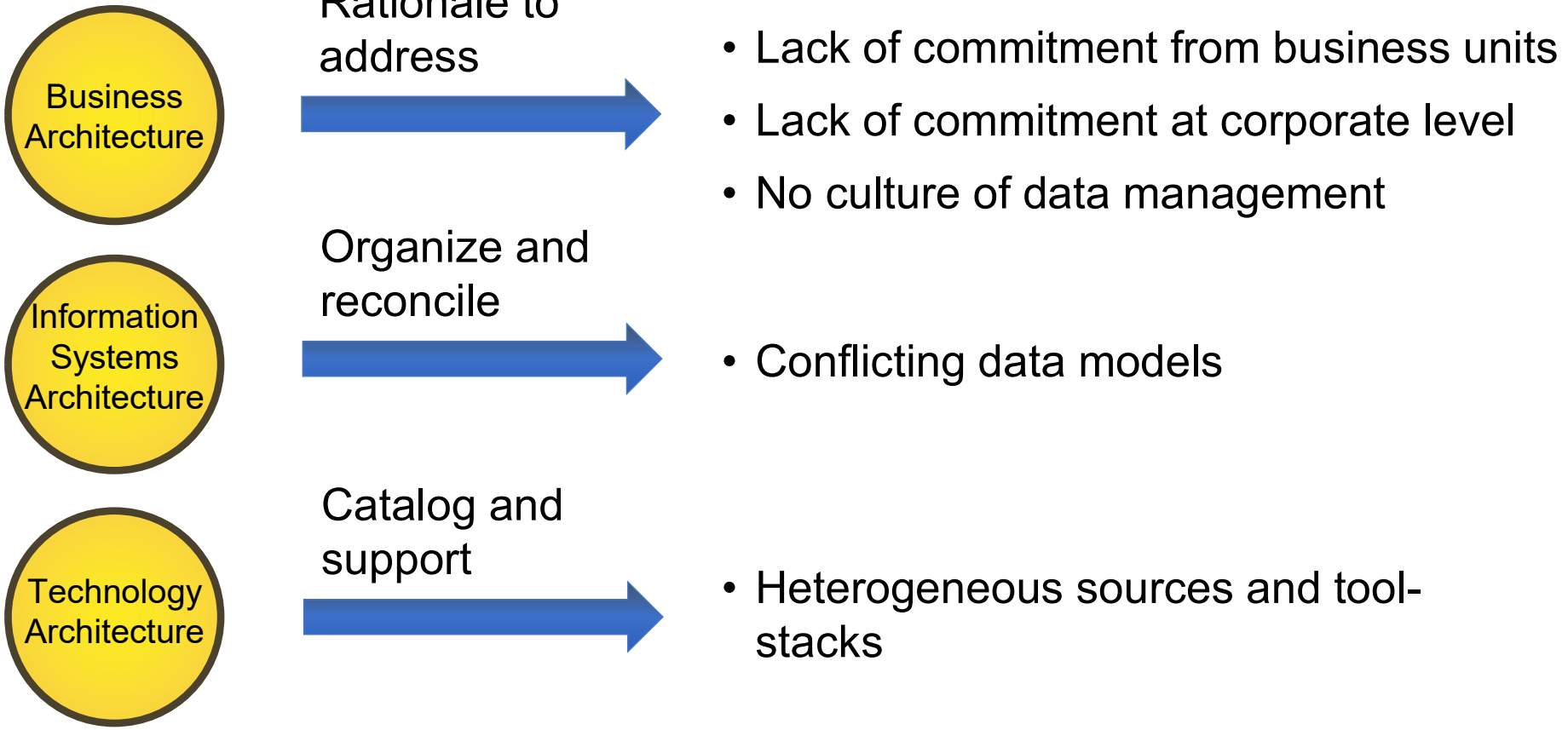
- The TOGAF® Standard
- The Open Agile Architecture™ standard (O-AA)
- The Digital Practitioner Body of Knowledge™ Standard (DPBoK)

Part of The Open Group ***Digital Portfolio***



LACIBUS

TOGAF®



LACIBUS

O-AA™

- Presentation of integrated data

- Definition of Data-Ops

- Specification of Data Platforms

LACIBUS

DPBoK™

- The Digital Practitioner Body of Knowledge™ Standard
- Information Management chapter contains section on *Data Integration and the System of Record*
- Use of master data to maintain integration between data in multiple – perhaps hundreds of - databases
- More detail on Customer Master Data Management in the C-MDM TOGAF® Guide

LACIBUS

Business Domain Standards

Business Definitions

- Exploration and Mining Business Reference Model and Capabilities Map
- Government Reference Model

Data Models

- The Federated Health Information Model (FHIM)
- Commercial Aviation Reference Model

LACIBUS

Data Platforms

Open Subsurface Data Universe™ (OSDU)

- Open source standards-based technology agnostic data platform
- For the energy industry (oil and gas exploration)

Open Footprint™ Forum

- Common model for footprint-related data covering all types of emissions
- Base calculations to normalize and aggregate data

LACIBUS

Other Publications

Search results for:
'data integration'

Items 1-10 of 338

. . . .

Related search terms

- Data Architecture
- Data standards
- Integration of SABSA Security Architecture Approaches with TOGAF ADM
- Data lake reference architecture
- Data center reference architecture

LACIBUS

Data Integration Work Group

- Part of The Open Group Architecture Forum
- Objectives
  - Create a body of architecture artefacts for data integration
  - And an overall framework to stitch them together
- Current State
  - First White Paper published
  - Survey of Enterprise Architects conducted
- Forward work program to be defined

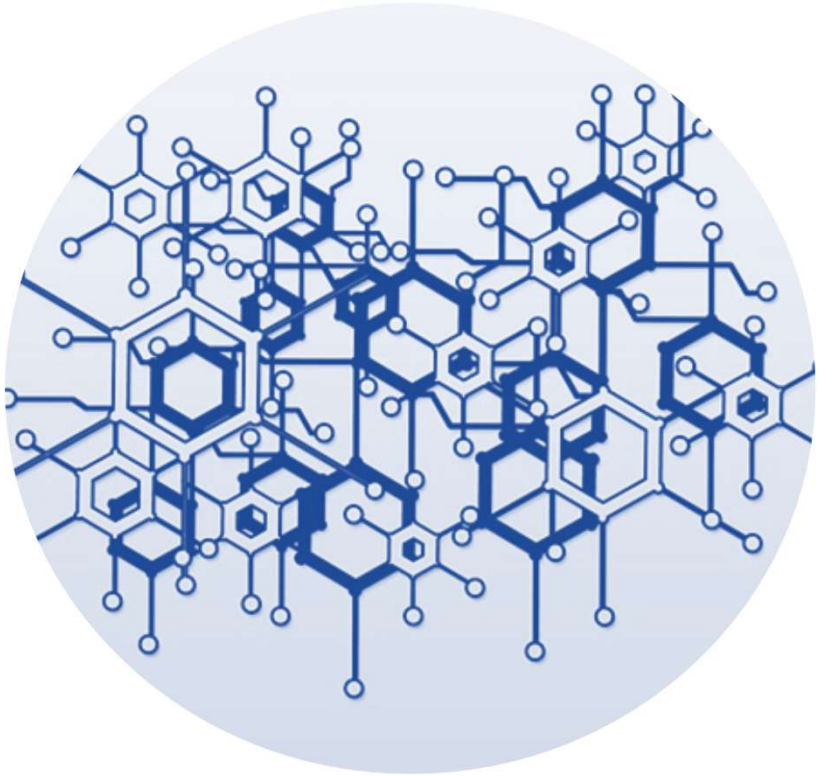
LACIBUS

Guide to Data Integration Architecture

- Possible Work Group deliverable
- How to use The Open Group architecture standards for data integration
- Based on research on data integration use cases such as mergers, rationalisation, federated enterprise interworking, and analytics
- Using TOGAF®, O-AA™ and DPBoK™

LACIBUS

Future Trends

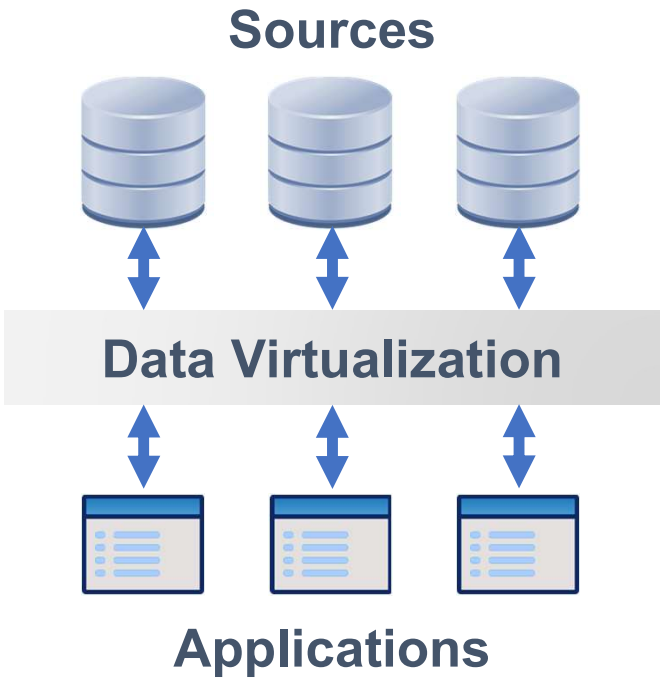


- Data Virtualization
- Data Fabric
- Data Mesh
- Data Platforms
- Natural Language Data
- Knowledge Graphs

LACIBUS

Data Virtualization

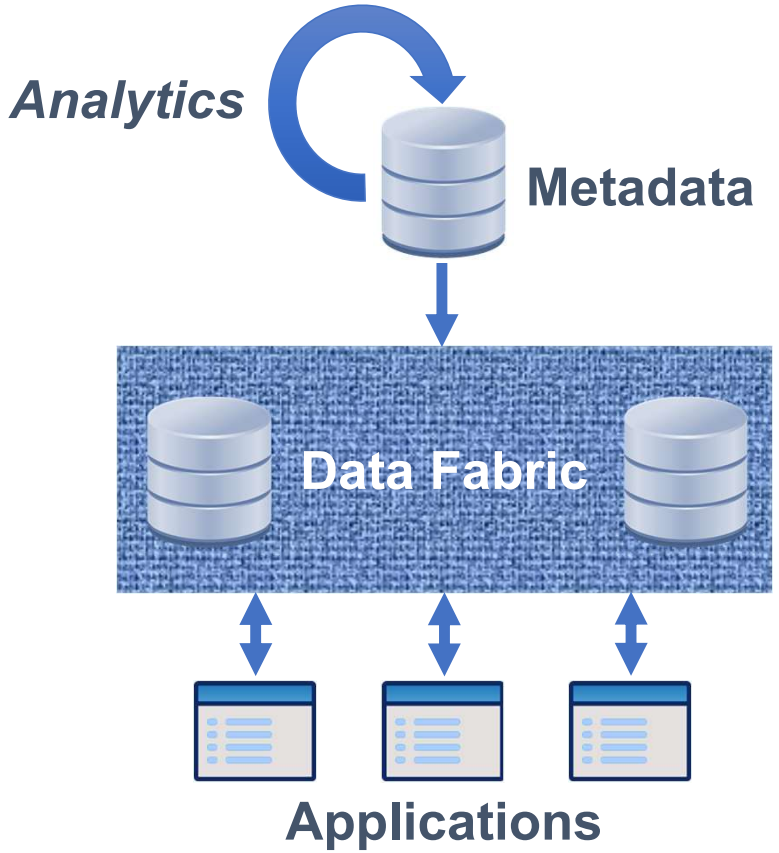
- Any approach to data management that allows an application to retrieve and manipulate data without requiring technical details
- Can provide a single customer view (or single view of any other entity) of the overall data
- Mostly for access rather than update
- Often used with relational data
- Often used by Data Fabric and Data Mesh platforms “under the hood”



LACIBUS

Data Fabric

- Unified access to data of different kinds
- SQL, NoSQL, IoT, cloud, on-premise, edge . . .
- Can use continuous automated analytics over metadata to organize the data



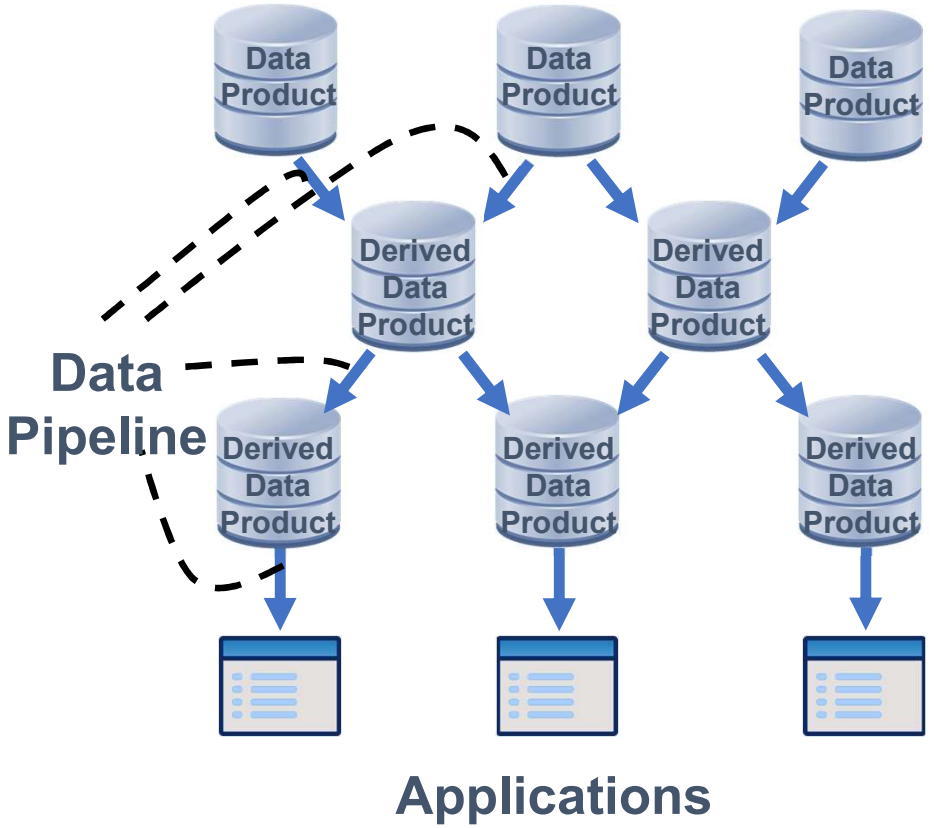
LACIBUS

Data Mesh

- Data is treated as a product
- Owned by teams that most intimately know and consume it
- Self-serve data platforms
- Federated Governance

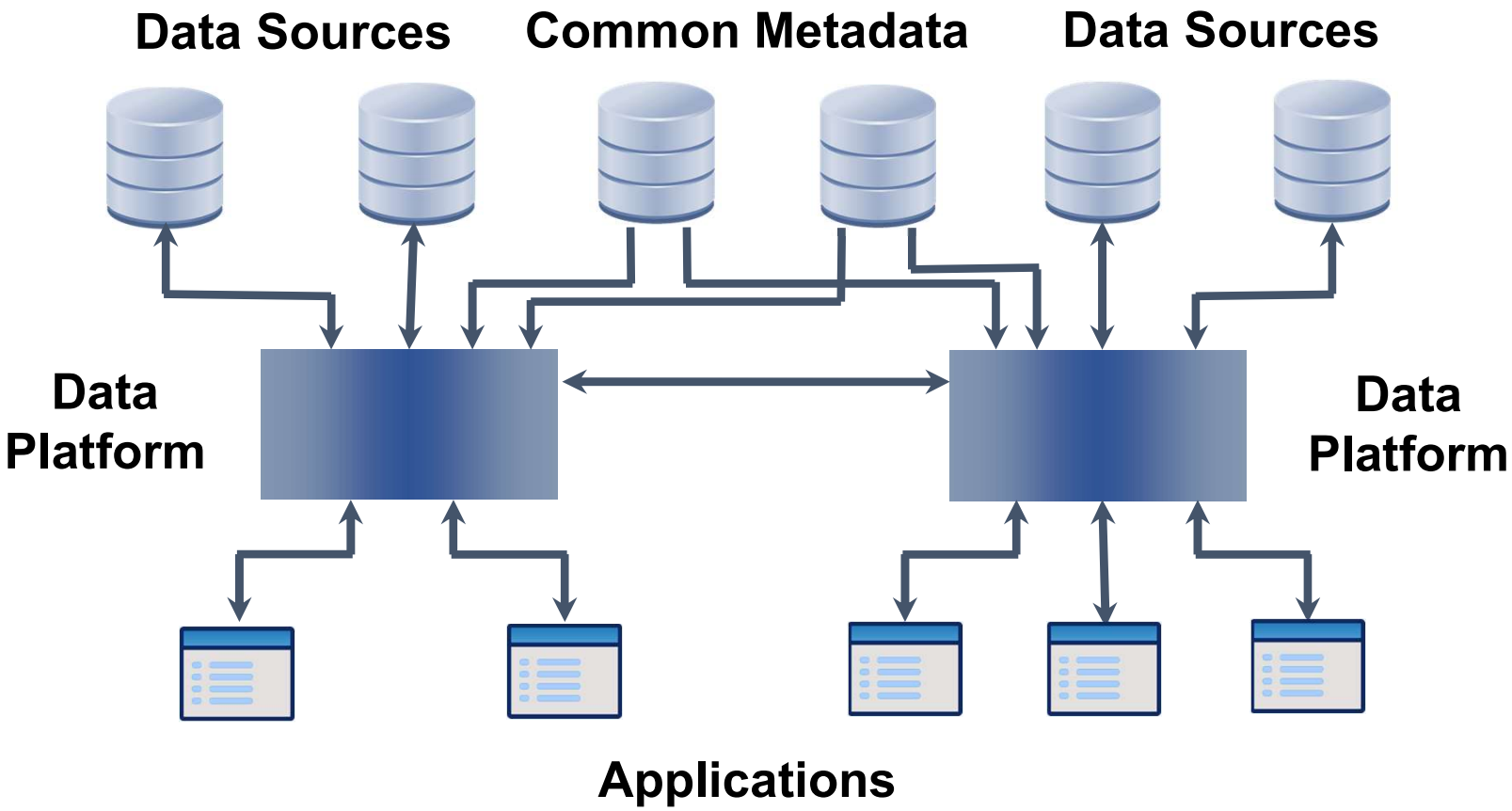
Data Mesh

- Data is treated as a product
- Owned by teams that most intimately know and consume it
- Self-serve data platforms
- Federated Governance

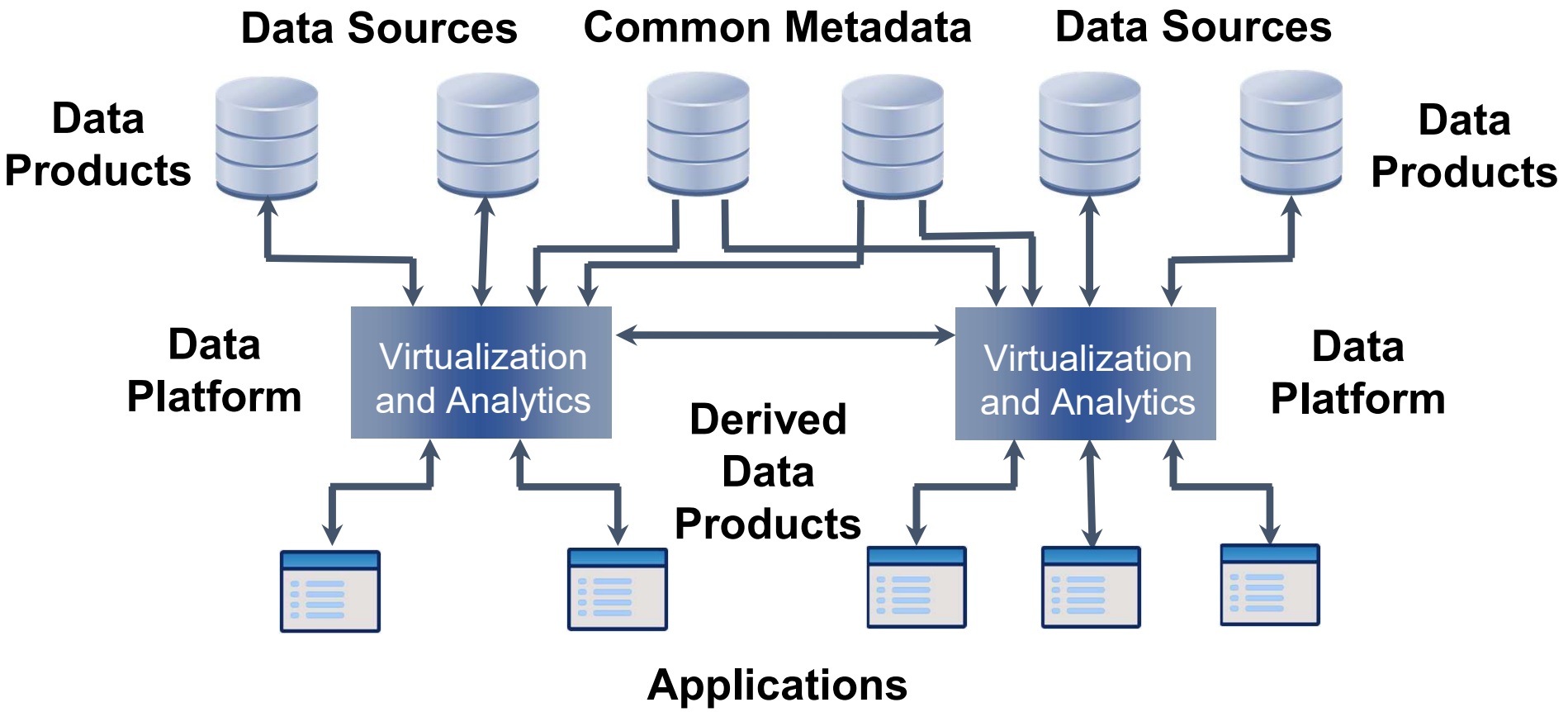


LACIBUS

Data Platforms



Data Platforms



LACIBUS

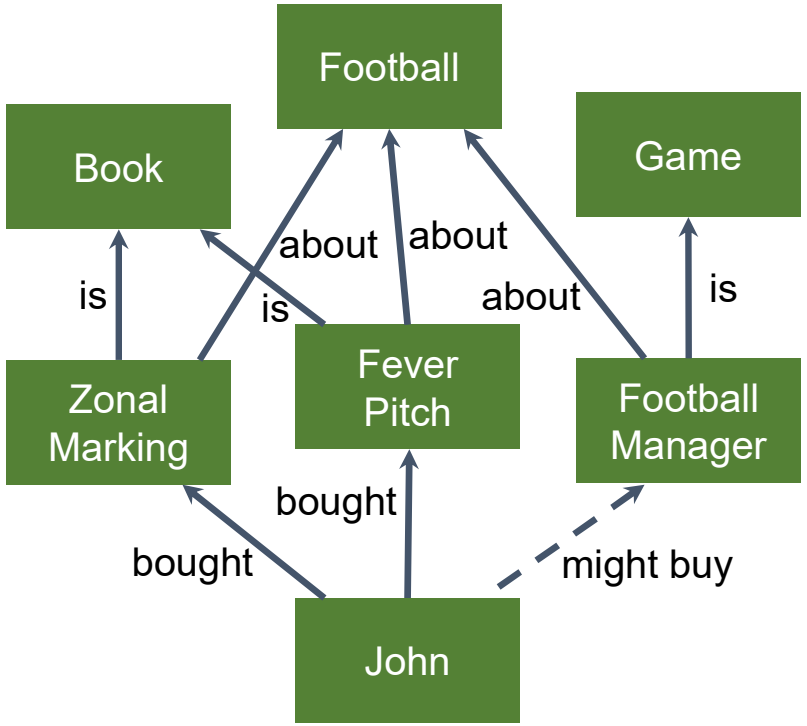
Natural Language Data

- Natural Language Processing (NLP) has improved enormously in recent years
- Sentence analysis and named object recognition
- Word vectors and sentence embeddings
- Enables integration of text and voice recordings

LACIBUS

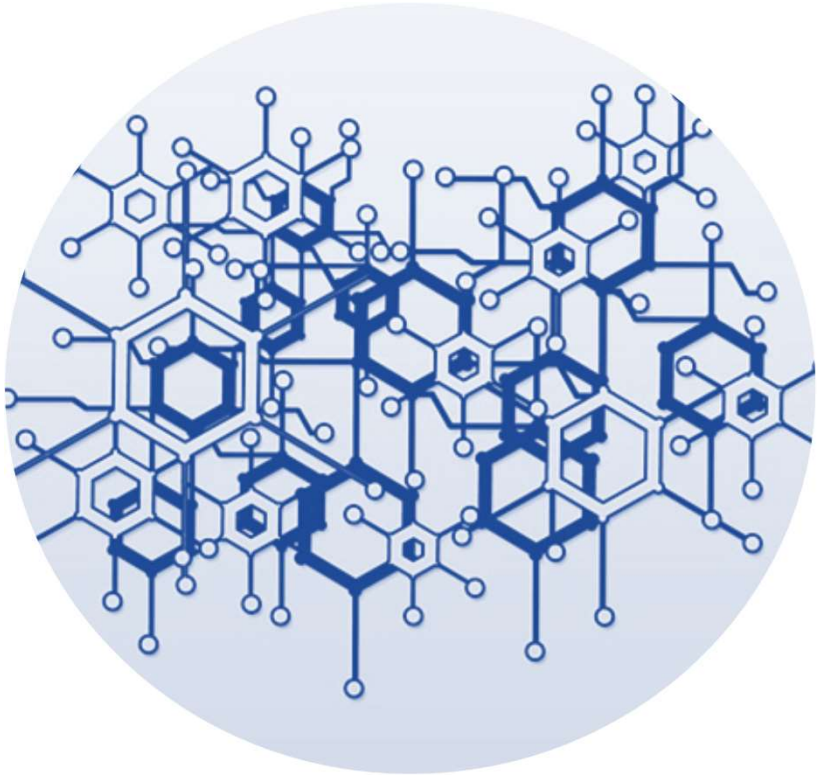
Knowledge Graphs

- A knowledge graph is a network of concepts and things, linked by the relations between them
- They are routinely used by marketing applications and search engines
- They can easily be merged and are a powerful tool in data integration



LACIBUS

Conclusions



- Data Integration
- Relevant Work from The Open Group
- Future Trends

LACIBUS

Data Integration *is* Difficult

- Integrations involving multiple departments must have department and corporate backing
- Business case must be clearly explained to all stakeholders
- Difficulties of differing data formats and interfaces must be addressed
- Poor data quality must be addressed
- Access control must be addressed, especially for PII

LACIBUS

The Open Group Work Can Help

Methods

- TOGAF, OAA, DPBoK, C-MDM

Business Domains

- Exploration and Mining, Government, Healthcare, Civil Aviation

Platforms

- OSDU, Open Footprint

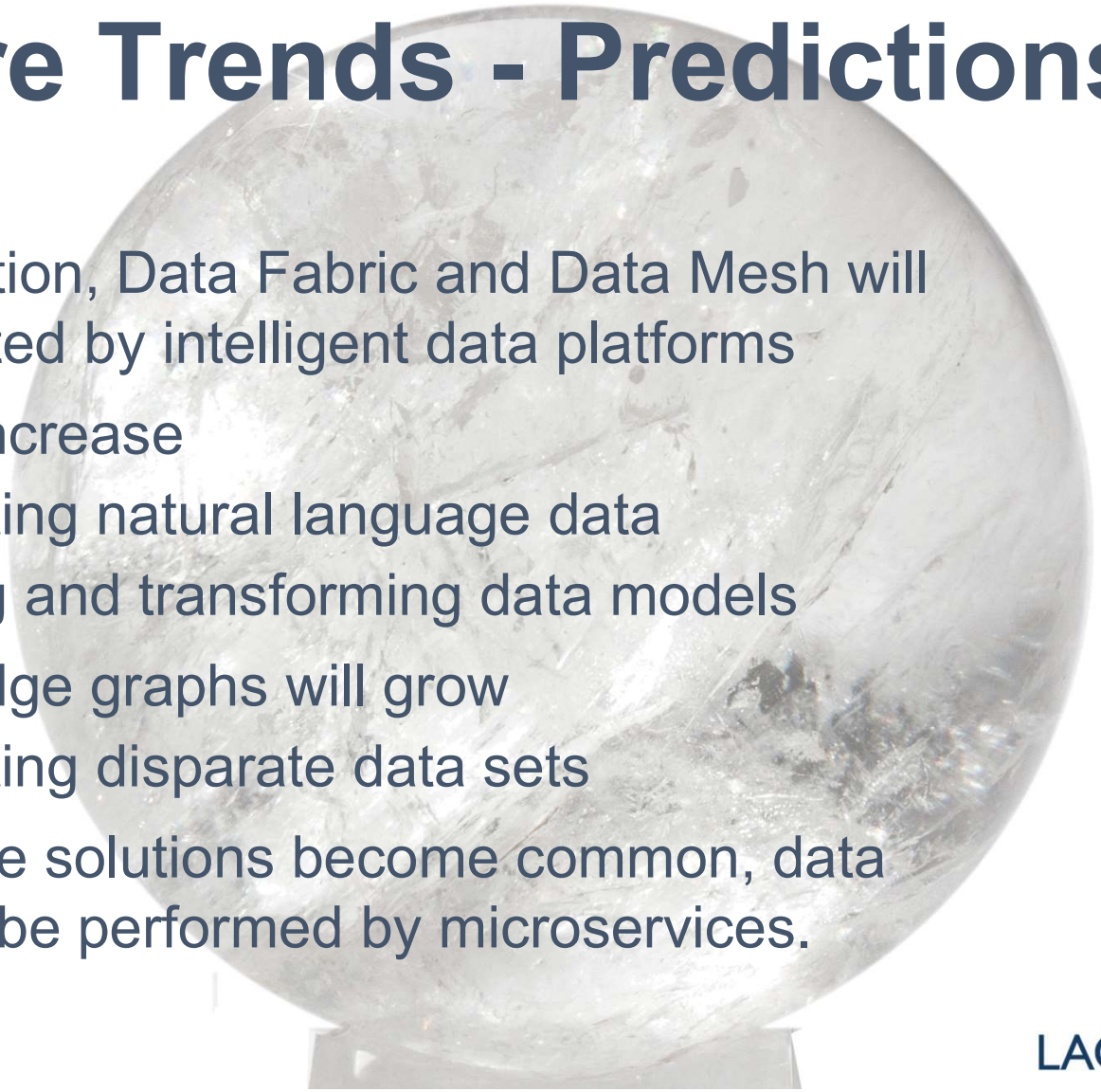
Other Publications

. . . .

The Data Integration Work Group

LACIBUS

Future Trends - Predictions



- Data Virtualization, Data Fabric and Data Mesh will merge, supported by intelligent data platforms
- Use of AI will increase
  - for integrating natural language data
  - for building and transforming data models
- Use of knowledge graphs will grow
  - for integrating disparate data sets
- As Cloud Native solutions become common, data integration will be performed by microservices.

LACIBUS

Questions?

LACIBUS

Thank You!

Chris Harding

- chris@lacibus.net
- linkedin.com/in/chris-harding-87b112
- @chrisjharding

LACIBUS

Website References

- Lacibus Virtual Data Lake and other open source material http://www.lacibus.net
- Lacibus Data Content Management System http://www.lacibus.com
- Data Integration Work Group https://collaboration.opengroup.org/projects/archmain/data_integration/
- The TOGAF Standard https://www.opengroup.org/togaf/
- The TOGAF Standard Digital Edition https://www.opengroup.org/togaf/10thedition
- The DPBoK Library https://publications.opengroup.org/dpbok-library
- The Open Agile Architecture™ website. https://www.opengroup.org/agilearchitecture
- The Federated Health Information Model (FHIM). https://www.fhim.org/
- The Open Group OSDU™ Forum. https://osduforum.org/
- The Open Group Open Footprint Forum. https://www.opengroup.org/openfootprint-forum
- The Open Group Architecture Forum. https://www.opengroup.org/architecture-forum

LACIBUS

Open Group Publications

Open Group Publications website https://publications.opengroup.org/

Search for content by keyword or document reference.

- Technical Standards for Data Integration. Open Group document reference w211
- Digital Practitioner Body of Knowledge™ Standard. Open Group document reference c196
- TOGAF® Series Guide: Information Architecture: Customer Master Data Management (C-MDM) Open Group document reference G218
- The Open Agile Architecture™ Standard. Open Group document reference c208
- The Exploration & Mining Business Reference Model. Open Group document reference c135
- The Exploration & Mining Business Capability Reference Map. Open Group document reference c143
- TOGAF® Series Guide: Government Reference Model (GRM). Open Group document reference G210
- The Open Group Commercial Aviation Reference Architecture. Open Group document reference P180

LACIBUS