Webinar

DATA CROSSROADS

Consulting firm, focused on data management maturity assessment, data management, governance frameworks, and data lineage business cases. Learn more at datacrossroads.nl

Trigyan

Software firm, offering a patent-pending no-code knowledge graph solution which can be based on bespoke or public ontologies. Trigyan transforms “uncurated data” to “actionable data” which you can trust to make business decisions. Learn more at trigyan.com

knowledge graphs, metadata management, and data lineage

Irina Steenbeek

Poll 1

What industry are you from? (single choice)

- Commercial
- Government
- Education
- Non-Profit

Poll 2

What is the status of a KG/DL/MM initiative in your company? (single choice)

- No intention
- Planning
- Implementing
- Implemented
- I don’t know

A few words about me…

CIVIL ENGINEER
ACTIVE BLOGGER
INTERNATIONAL BANKS
DATA LINEAGE
DATA AND INFORMATION VALUE CHAIN
CONSULTANT
4 BOOKS
3 WHITEPAPERS
50 ARTICLES
SPEAKER AT INTERNATIONAL CONFERENCES
DATA MANAGEMENT FOR GLOBAL COMPANIES
OWNER DATA CROSSROADS
IMPLEMENTATION OF DATA MANAGEMENT
11 YEARS OF HANDS-ON EXPERIENCE
FINANCE & BUSINESS CONTROL
ERP IMPLEMENTATION
MANAGEMENT CONSULTANCY
DATA MANAGEMENT MATURITY ASSESSMENT (REVIEWS)

A three-headed serpent: knowledge graphs, metadata management, and data lineage

Why?

- Demonstrate similarities and profound relationships between these three concepts

What?

Describe each concept in terms of:

- Definitions and structure
- Business drivers
- Architecture and technology
- Use cases

How?

High-level approach of integrated implementation

A three-headed serpent: knowledge graphs, metadata management, and data lineage

Why?

- Demonstrate similarities and profound relationships between these three concepts

What?

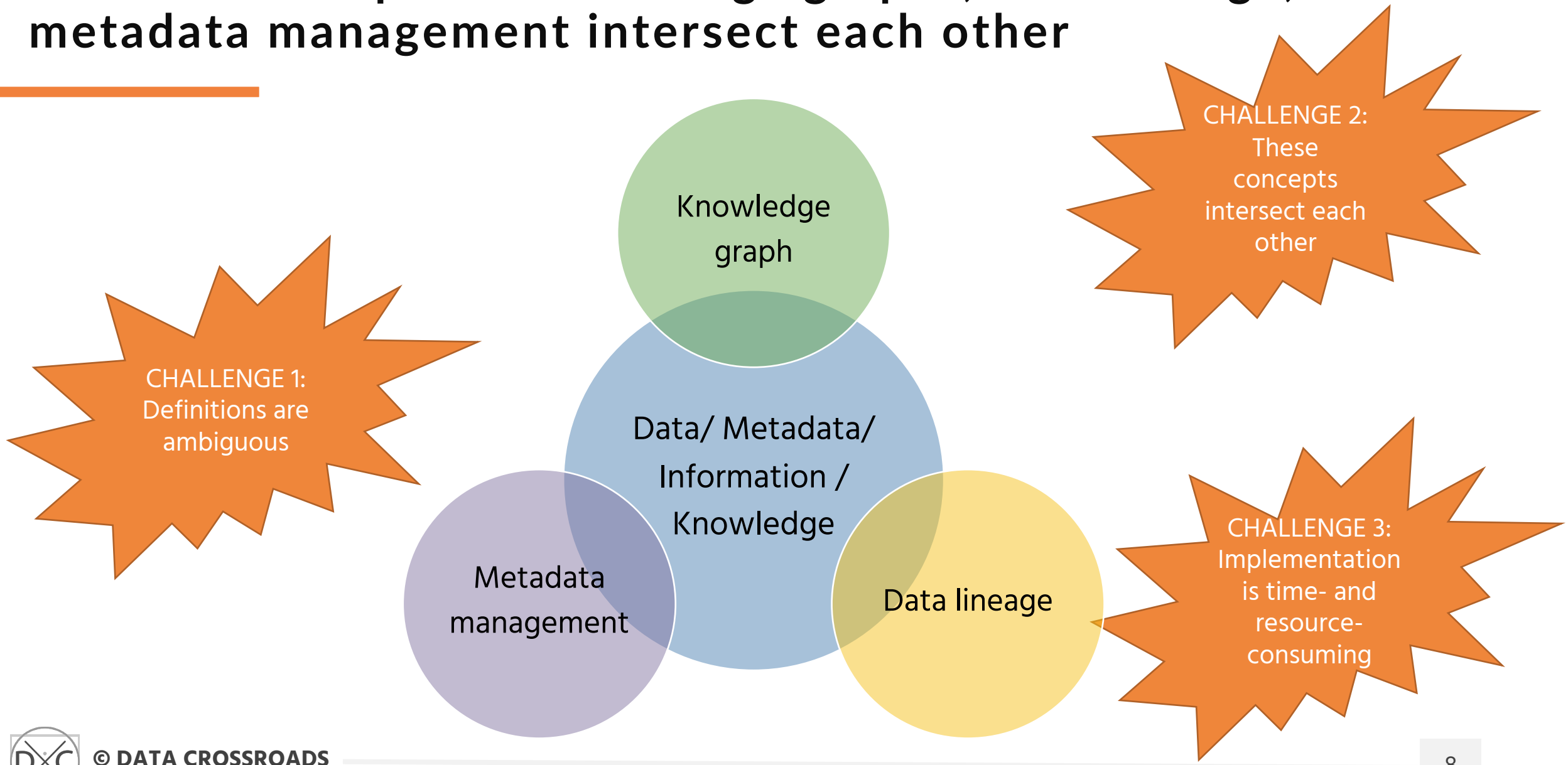
Describe each concept in terms of:

- Definitions and structure
- Business drivers
- Architecture and technology
- Use cases

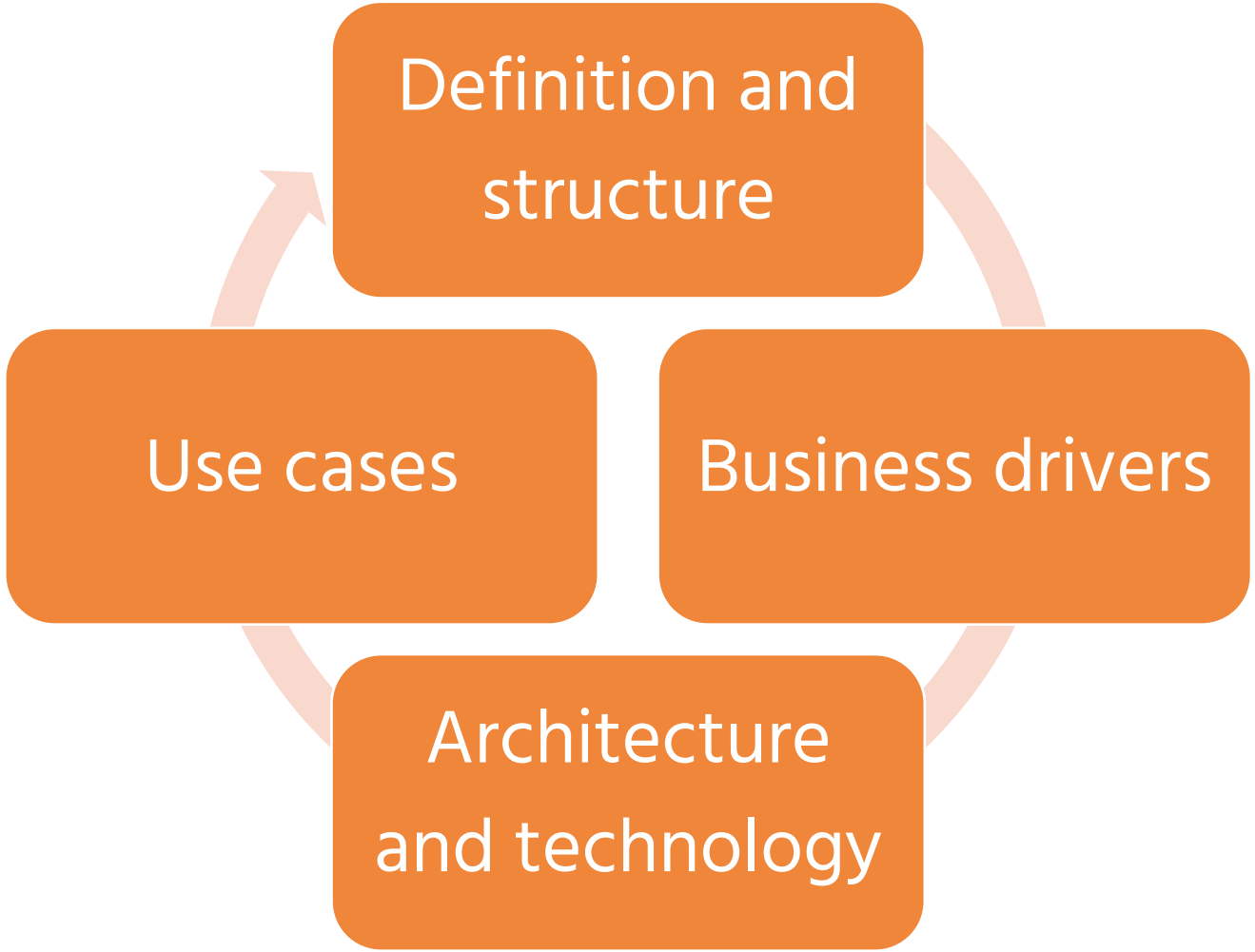
How?

High-level approach of integrated implementation

The DM concepts of knowledge graphs, data lineage, and metadata management intersect each other



We will compare these capabilities and the relationships between them by investigating their:



A three-headed serpent: knowledge graphs, metadata management, and data lineage

Why?

- Demonstrate similarities and profound relationships between these three concepts

What?

Describe each concept in terms of:

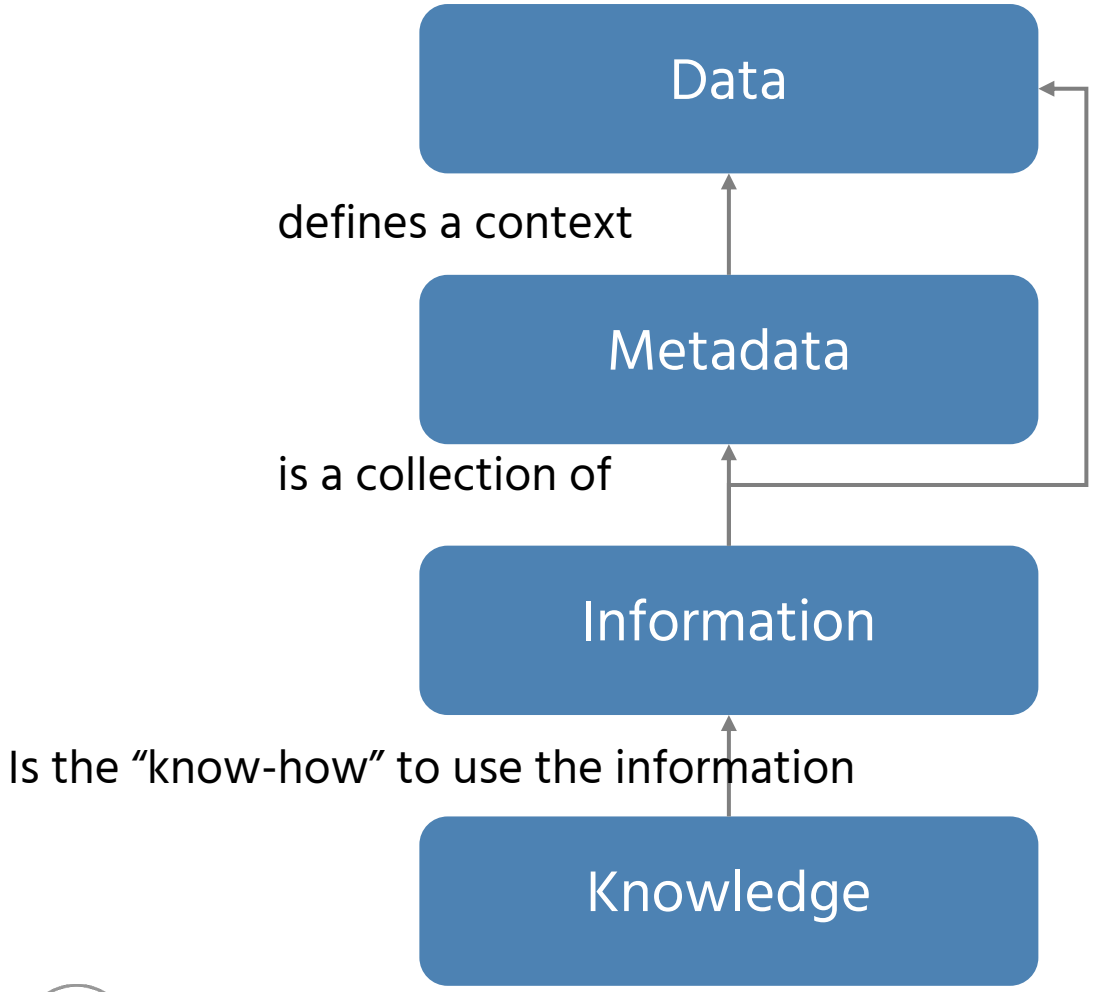
- Definitions and structure
- Business drivers
- Architecture and technology
- Use cases

How?

High-level approach of integrated implementation

DATA, METADATA, INFORMATION, AND KNOWLEDGE

(Meta)data, information, and knowledge are all subjects of data management



50

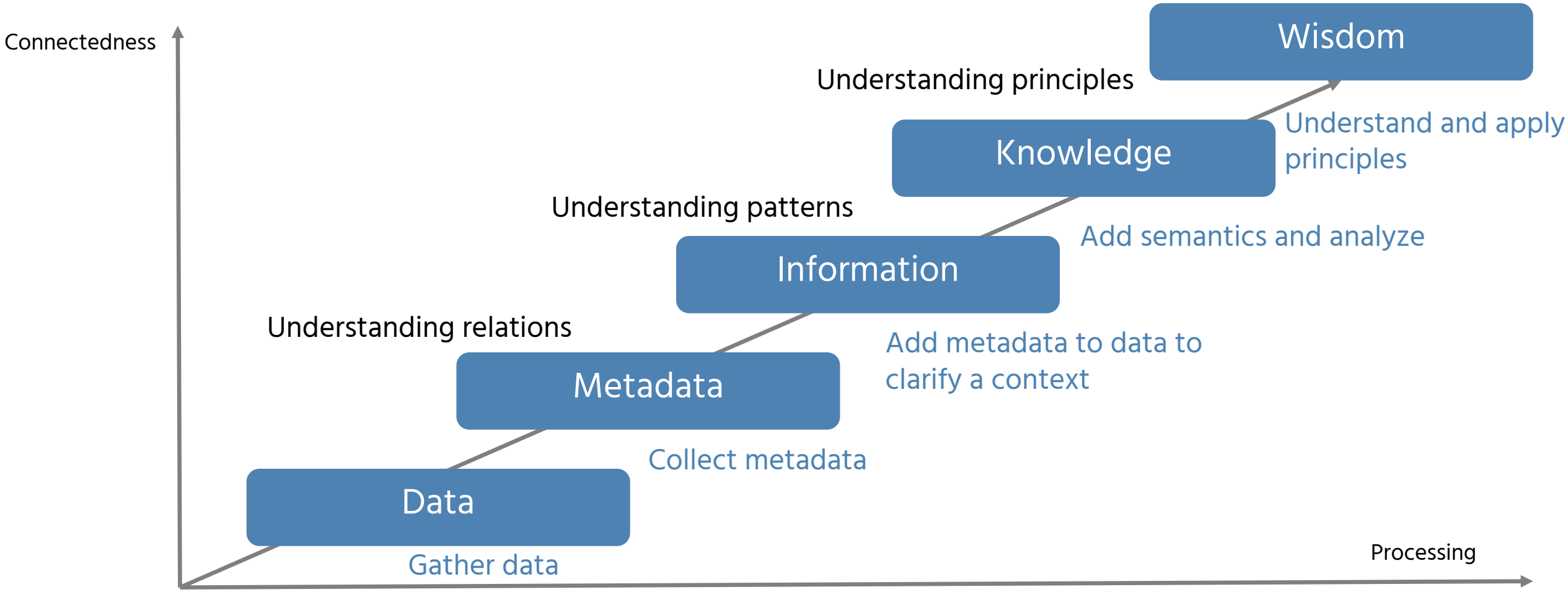
50 km/h

50 C/F

Historical weather information per region

Weather forecast per region

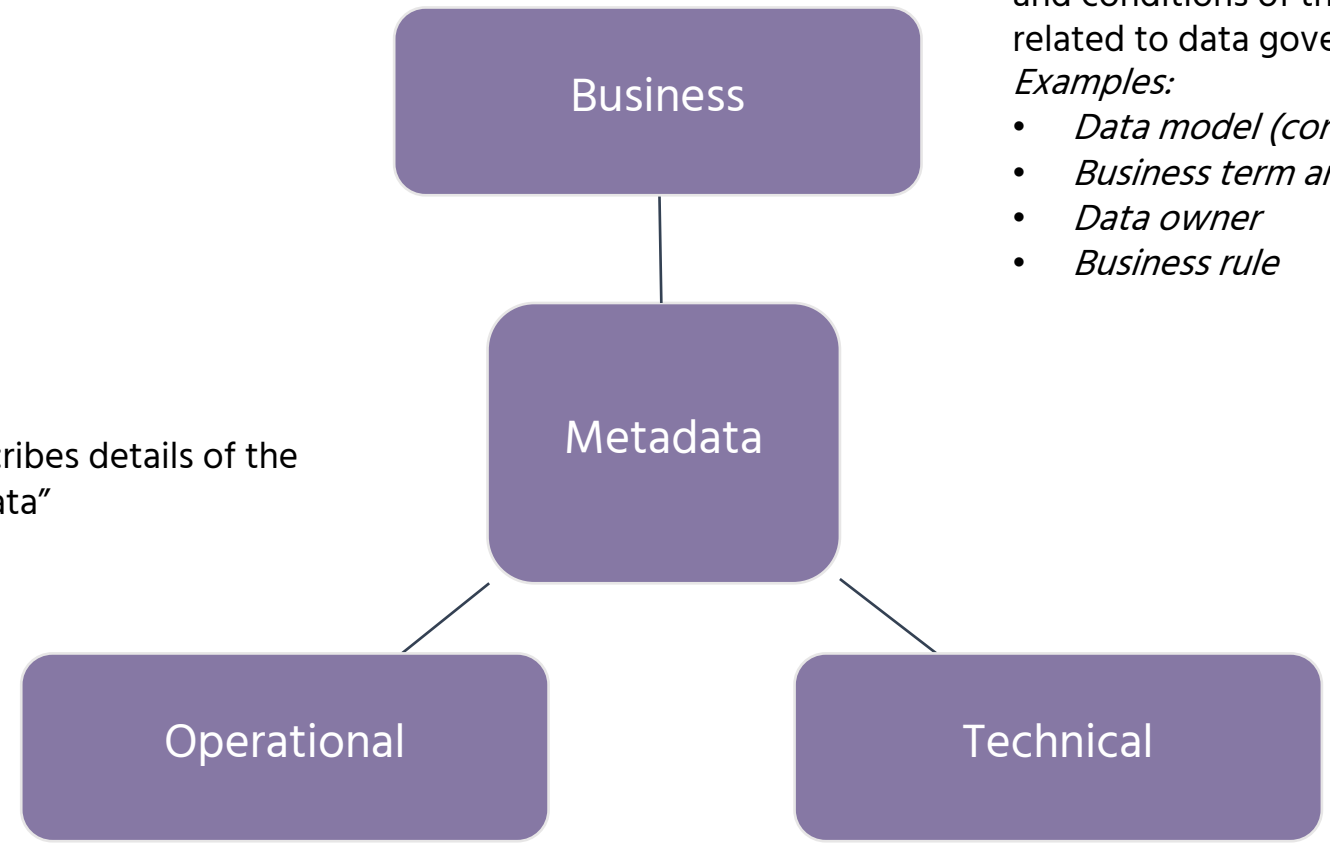
Data, information, and knowledge in the contexts of a human's mindset and data processing



METADATA MANAGEMENT, KNOWLEDGE GRAPHS, AND DATA LINEAGE:

DEFINITIONS AND STRUCTURE

Metadata has various classifications



"Business metadata focuses largely on the content and conditions of the data and includes details related to data governance"

Examples:

- *Data model (conceptual/semantic, logical)*
- *Business term and definition*
- *Data owner*
- *Business rule*

"Operational Metadata describes details of the processing and accessing data"

Examples:

- *Logs of jobs execution*
- *Error logs*

"Technical Metadata provides information about technical details of data, the systems that store data, and the processes that move it within and between systems"

Examples:

- *Physical data models*
- *Table and column properties*
- *ETL jobs*

Source: DAMA International. DAMA-DMBOK: Data Management Body of Knowledge, Second Edition. Bradley Beach, N.J.: Technics Publications, 2017, p.423.

DEFINITION

Metadata management

"PLANNING, IMPLEMENTATION, AND CONTROL ACTIVITIES TO ENABLE ACCESS TO HIGH QUALITY, INTEGRATED METADATA."

SOURCE: DAMA INTERNATIONAL. DAMA-DMBOK: DATA MANAGEMENT BODY OF KNOWLEDGE, SECOND EDITION. BRADLEY BEACH, N.J.: TECHNICS PUBLICATIONS, 2017, P.419.

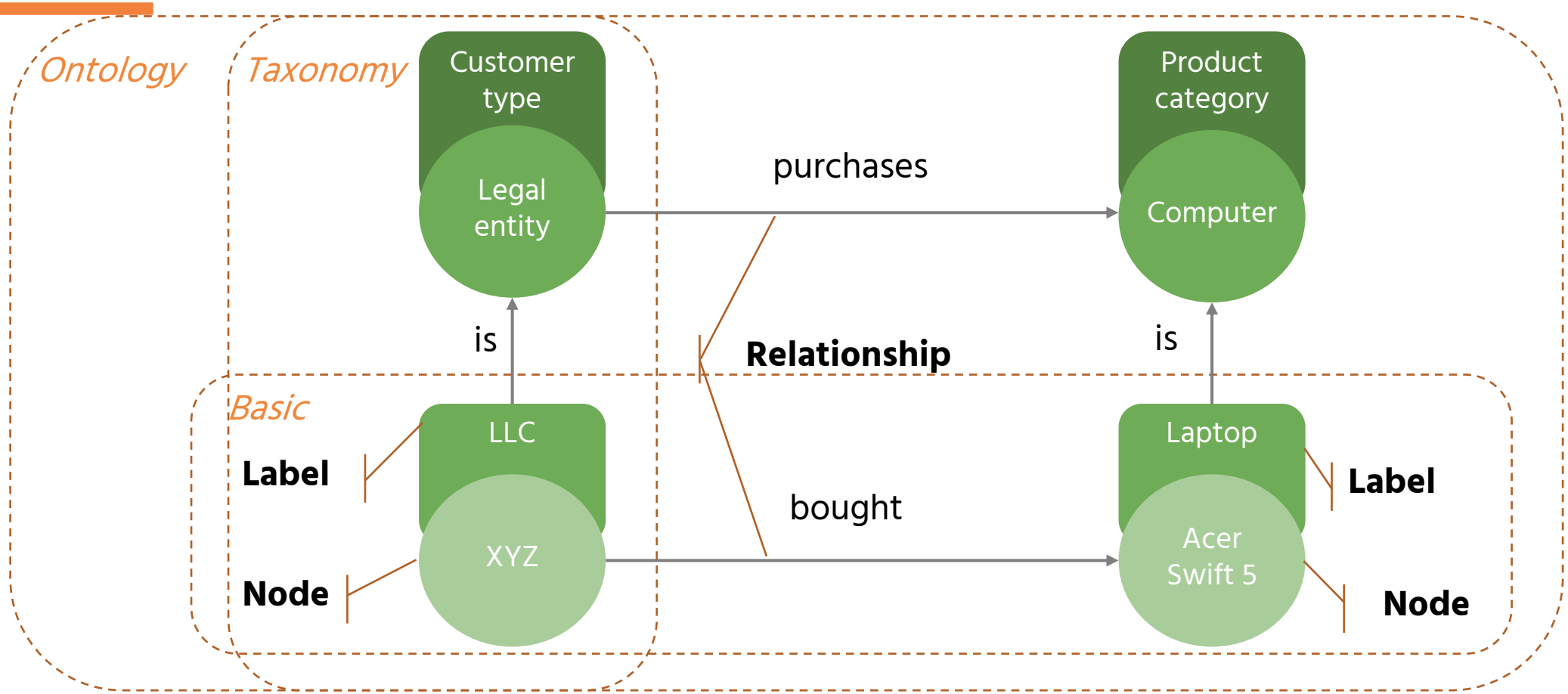
DEFINITION

Knowledge graph

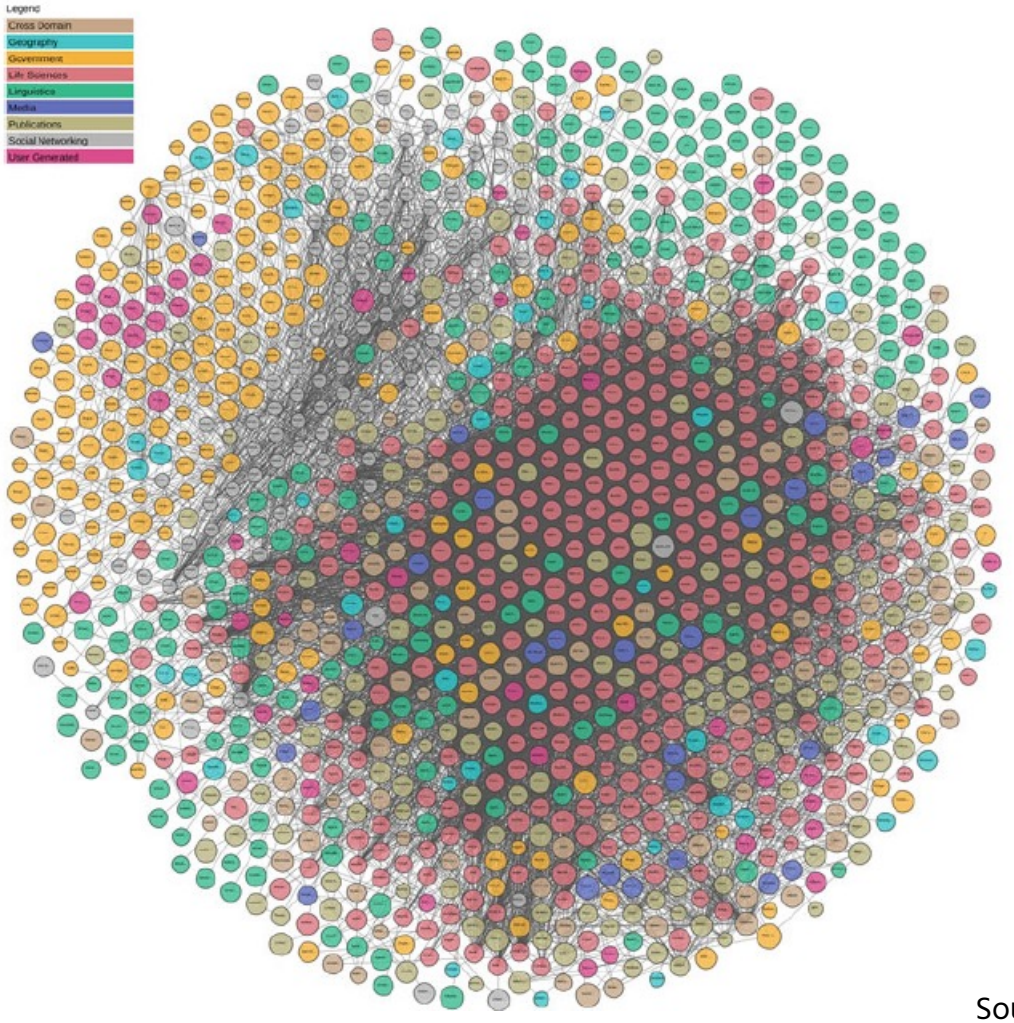
KNOWLEDGE GRAPHS ARE INTERLINKED SETS OF FACTS THAT DESCRIBE REAL-WORLD ENTITIES, EVENTS, OR THINGS AND THEIR INTERRELATIONS IN A HUMAN- AND MACHINE-UNDERSTANDABLE FORMAT.

Source: Knowledge Graphs: Data in Context for Responsive Businesses. O'Reilly Media, Inc., 2021.

Knowledge graphs add value by collecting information (data and metadata) and adding semantics to it



This is how a real knowledge graph looks like:



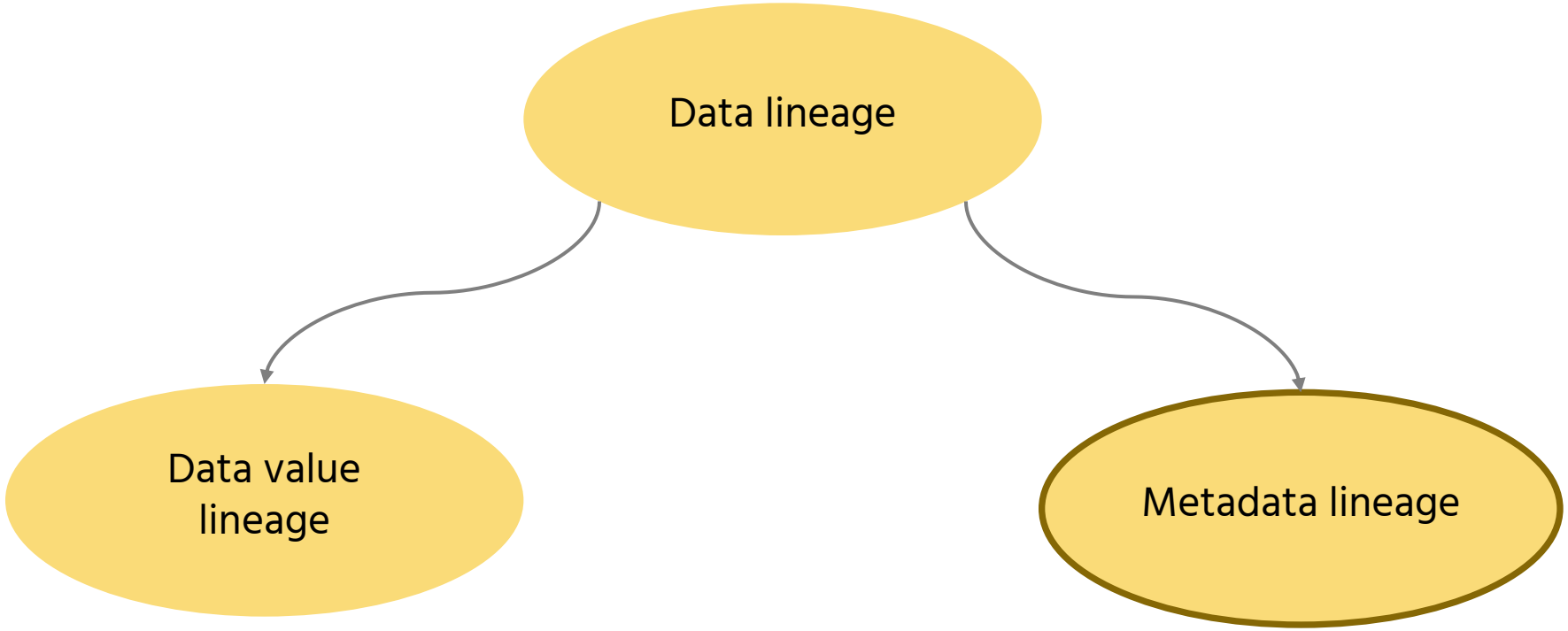
Source: https://lod-cloud.net/

DEFINITION

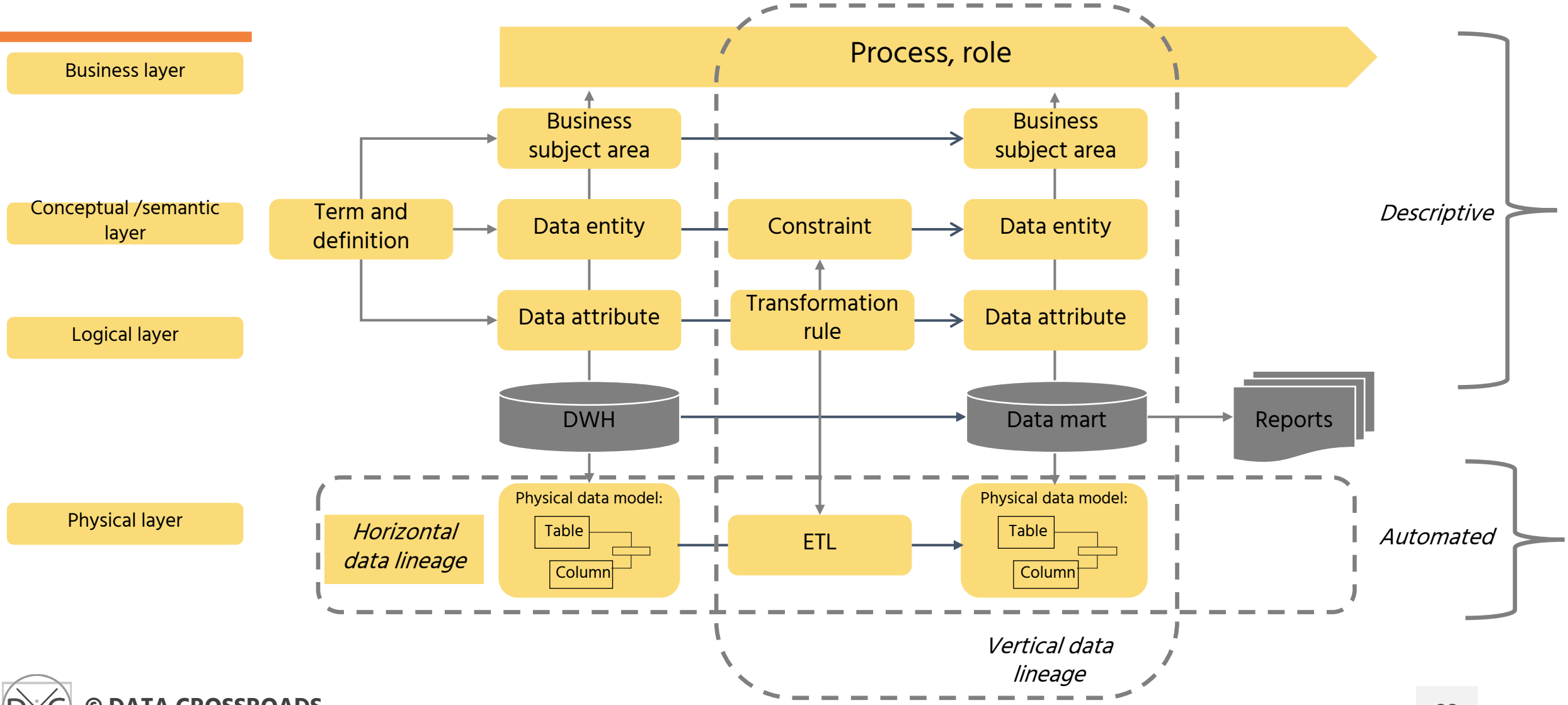
Data lineage

A DESCRIPTION OF DATA MOVEMENTS AND TRANSFORMATIONS AT VARIOUS ABSTRACTION LEVELS ALONG DATA CHAINS, AND RELATIONSHIPS BETWEEN DATA AT THESE LEVELS

Data lineage describes the movement of data by a means of metadata

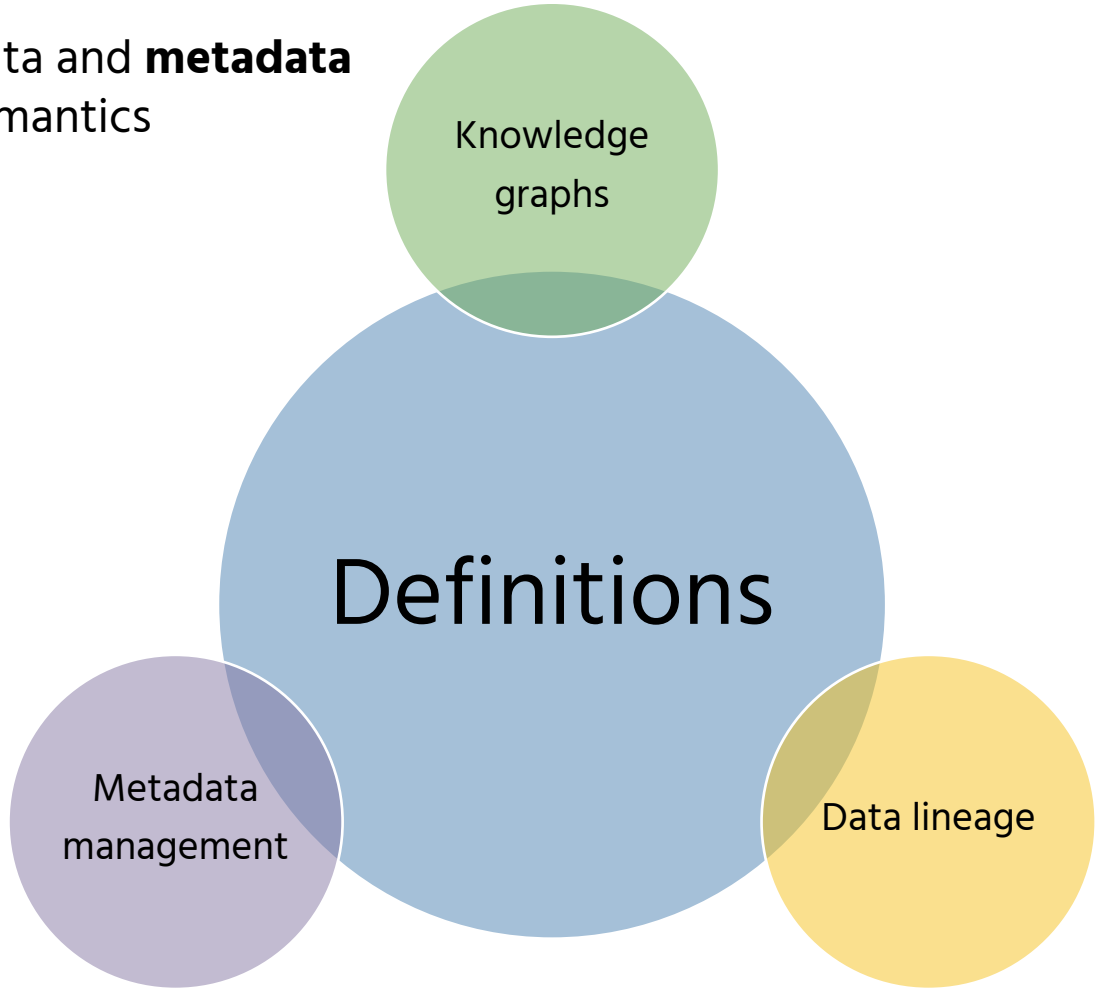


Data lineage can be documented at different abstraction levels



All concepts have similar components in their definitions:

Interlinked data and **metadata** with added semantics



Planning, implementation, control, and **integration of metadata**

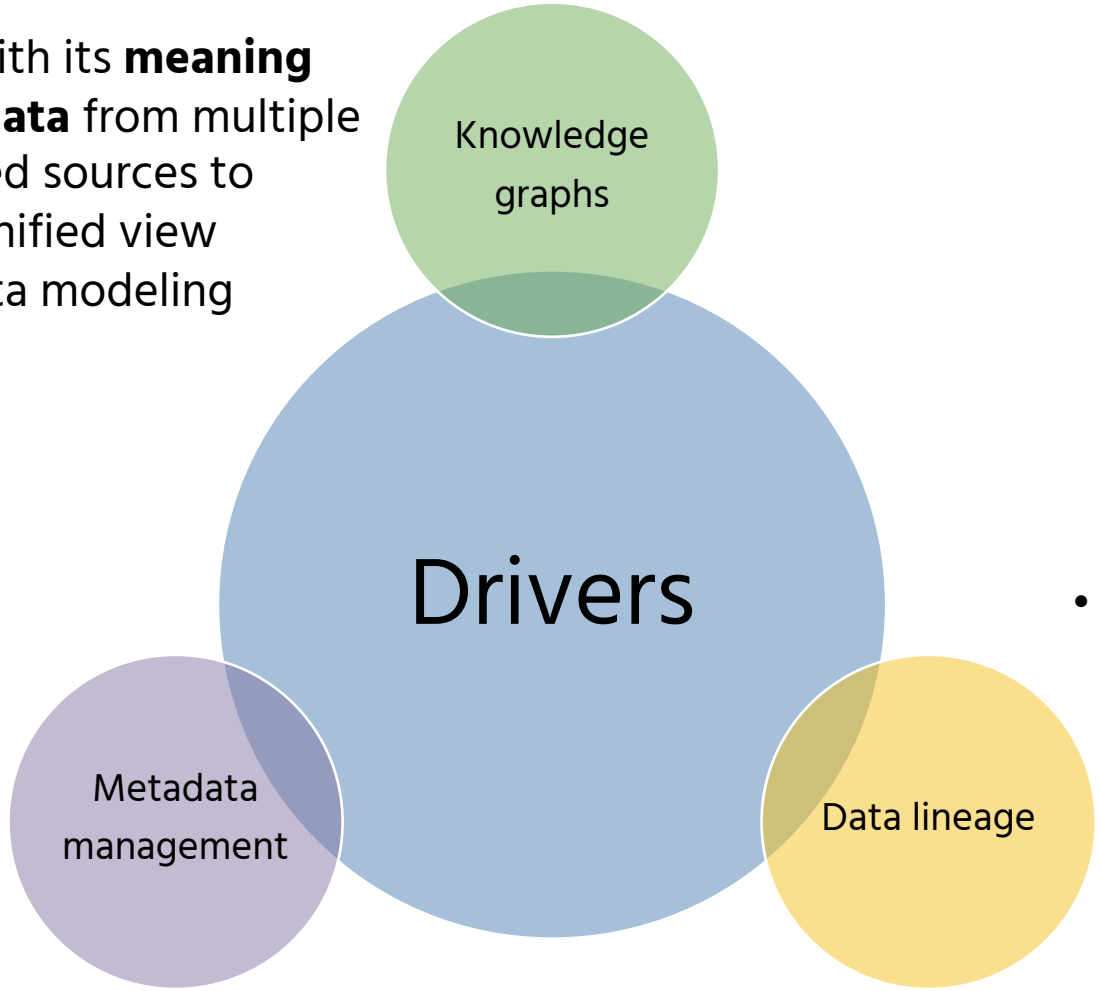
Description of data movements, transformation, and **relationships** at various abstraction levels by a means of **metadata**

METADATA MANAGEMENT, KNOWLEDGE GRAPHS, AND DATA LINEAGE:

BUSINESS DRIVERS

All concepts have similar drivers for their realization

- Link data with its **meaning**
- **Integrate data** from multiple unconnected sources to provide a unified view
- Simplify data modeling



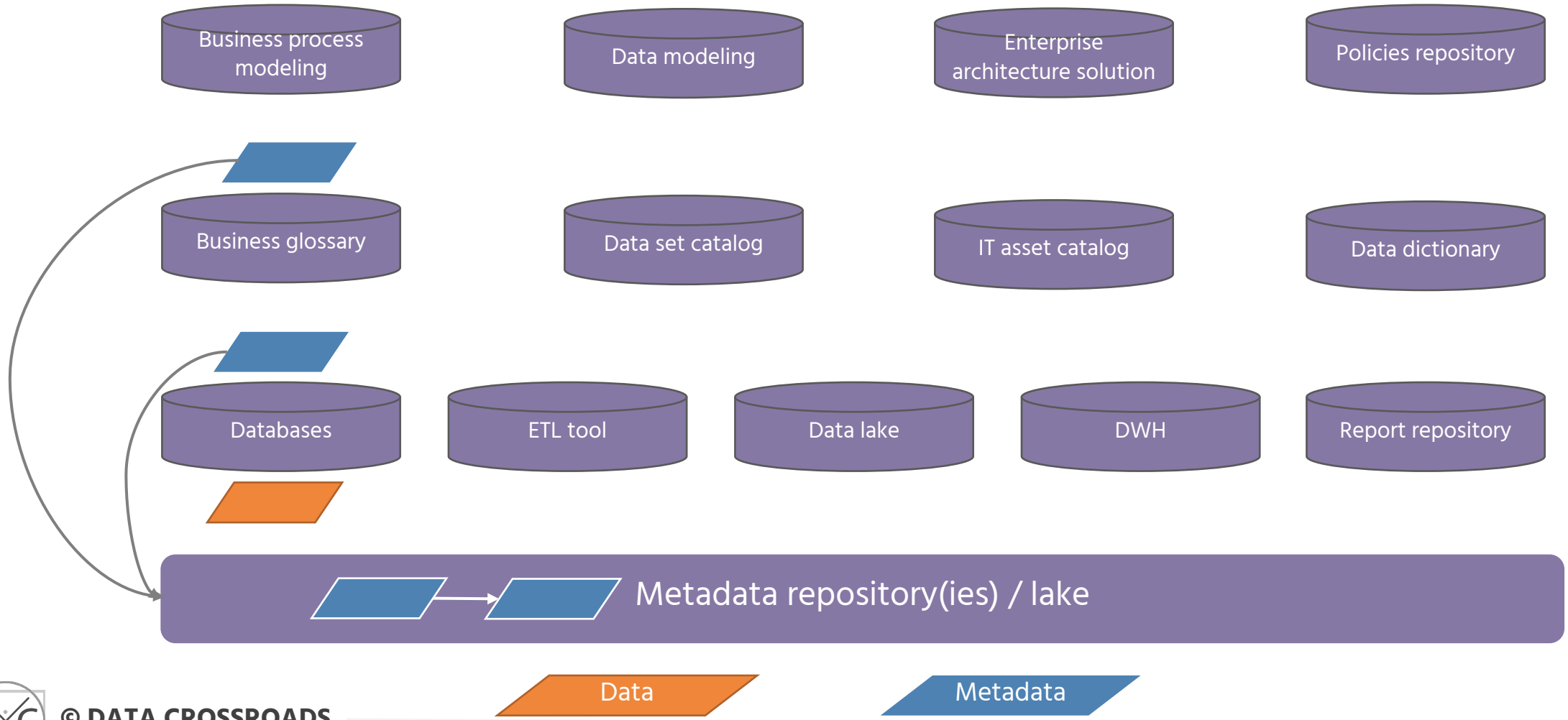
- Increase confidence in data by **providing context**
- **Integrate data**
- **Improve efficiency by identifying redundant data**

- Traceability and transparency of data processing, transformation, and **integration** due to:
  - Legislative and audit requirements
  - **Necessity to reduce IT and DevOps costs**

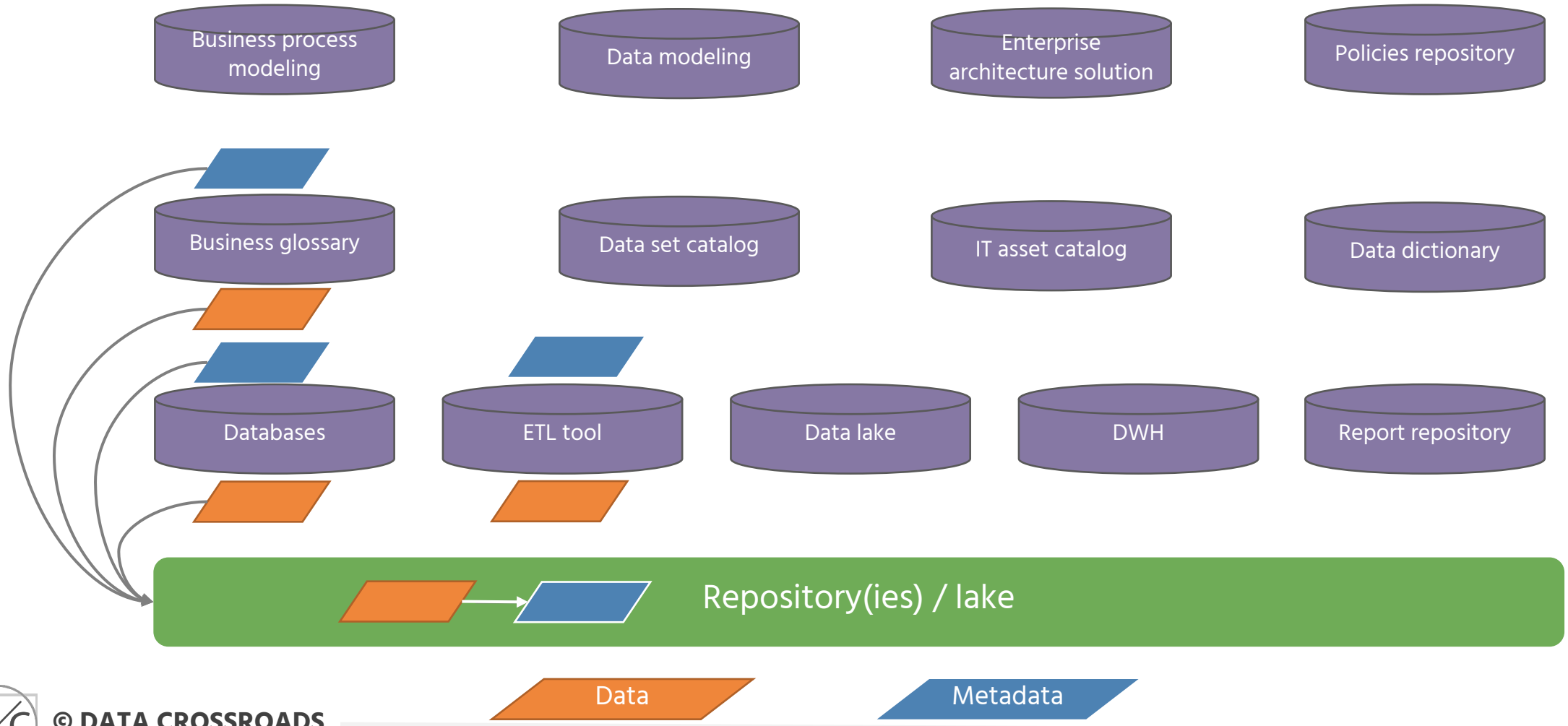
METADATA MANAGEMENT, KNOWLEDGE GRAPHS, AND DATA LINEAGE:

ARCHITECTURE AND TECHNOLOGY

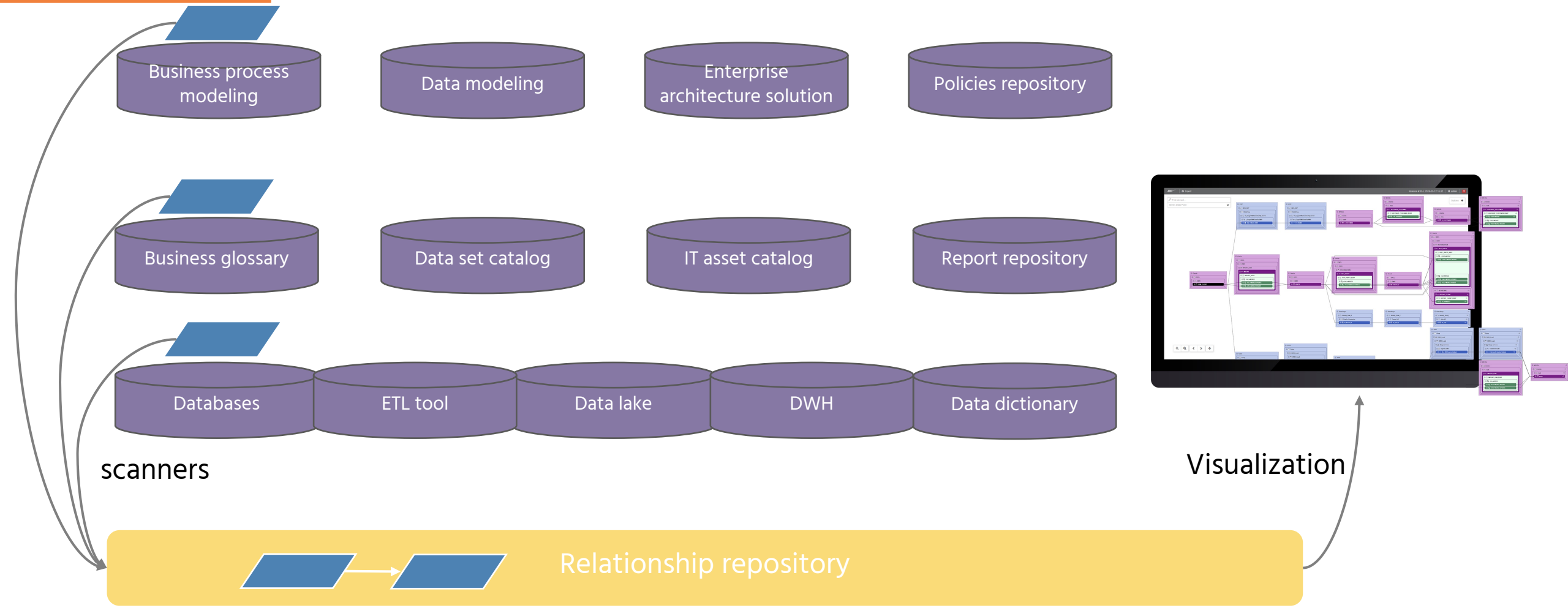
Metadata repository(ies) or lake collect metadata stored in different repositories



Knowledge graphs link both data and metadata

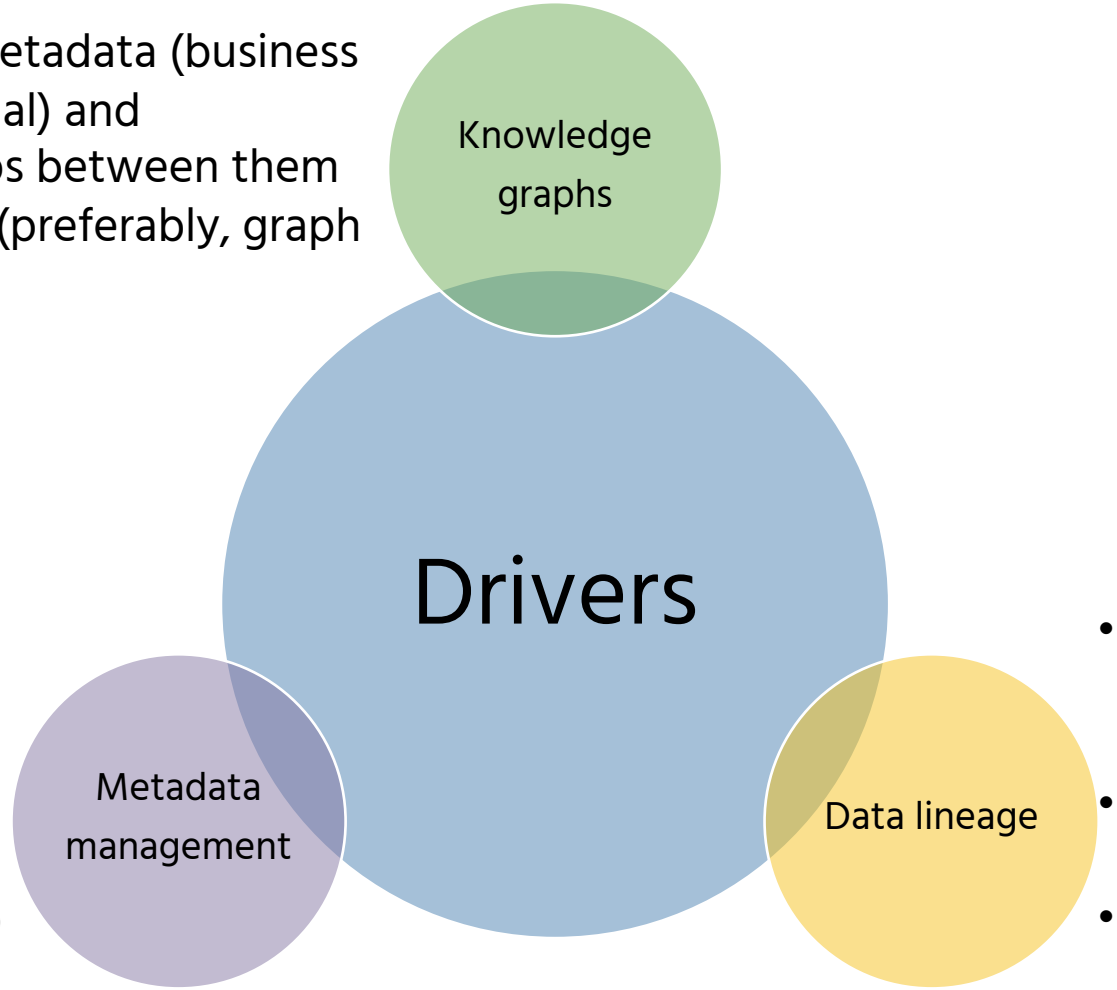


Various repositories with different database types store data and metadata



All capabilities require similar metadata objects and their relationships from the same data repositories.

- Data and metadata (business and technical) and relationships between them
- Repository (preferably, graph databases)

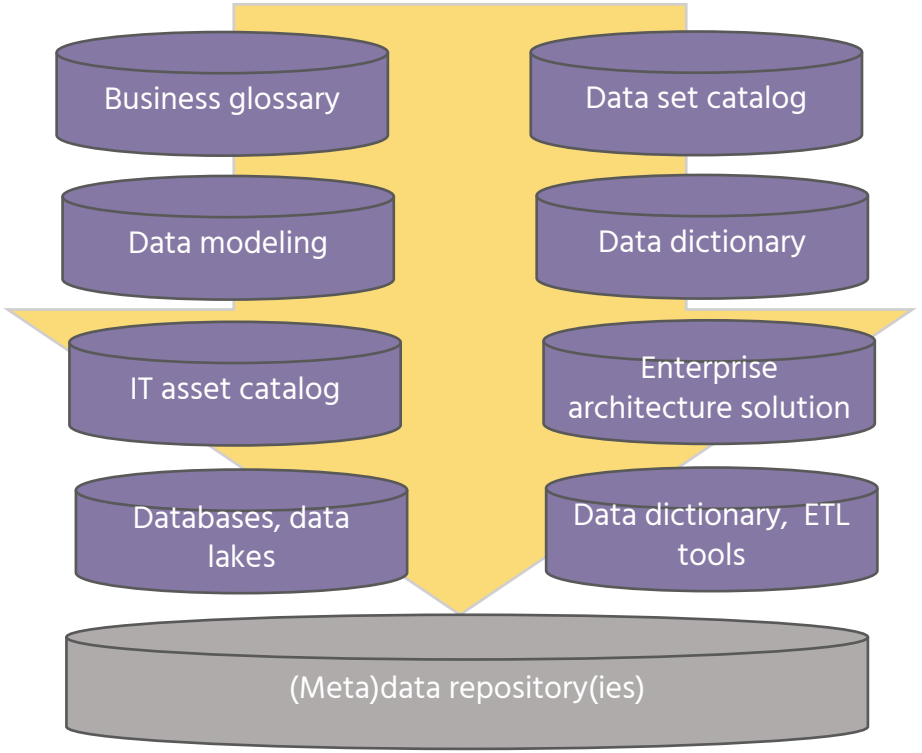


- Business, technical, and operational metadata
- Metadata repository (relational or graph databases)

- Metadata (business and technical) and relationships between
- Business rules that describe transformations
- Repository (relational or graph databases)

All concepts have their roles in managing data and metadata at different abstraction levels

THE METADATA MANAGEMENT CAPABILITY PLANS, DESIGNS, IMPLEMENTS, AND MAINTAINS METADATA OF VARIOUS TYPES IN CORRESPONDING (META) DATA REPOSITORIES.

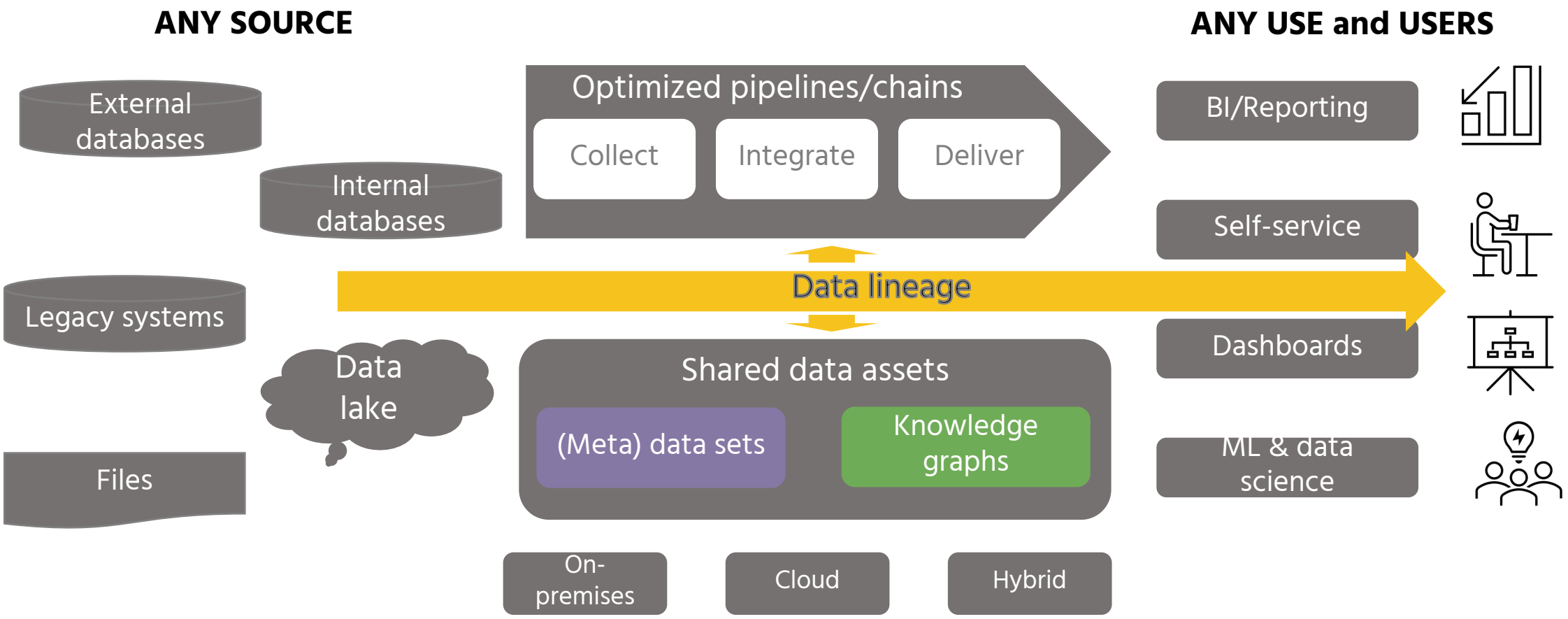


THE KNOWLEDGE GRAPH CAPABILITY PLANS, DESIGNS, IMPLEMENTS, AND MAINTAINS INTEGRATION OF RELATIONSHIPS BETWEEN DATA AND METADATA FROM VARIOUS (META)DATA REPOSITORIES

THE DATA LINEAGE CAPABILITY PLANS, DESIGNS, IMPLEMENTS, AND MAINTAINS THE INTEGRATION AND VISUALIZATION OF METADATA AND ITS RELATIONSHIPS STORED VARIOUS (META)DATA REPOSITORIES.

.

“A data fabric’s job is to connect any kind of data to anywhere and anyone (or anything).”*



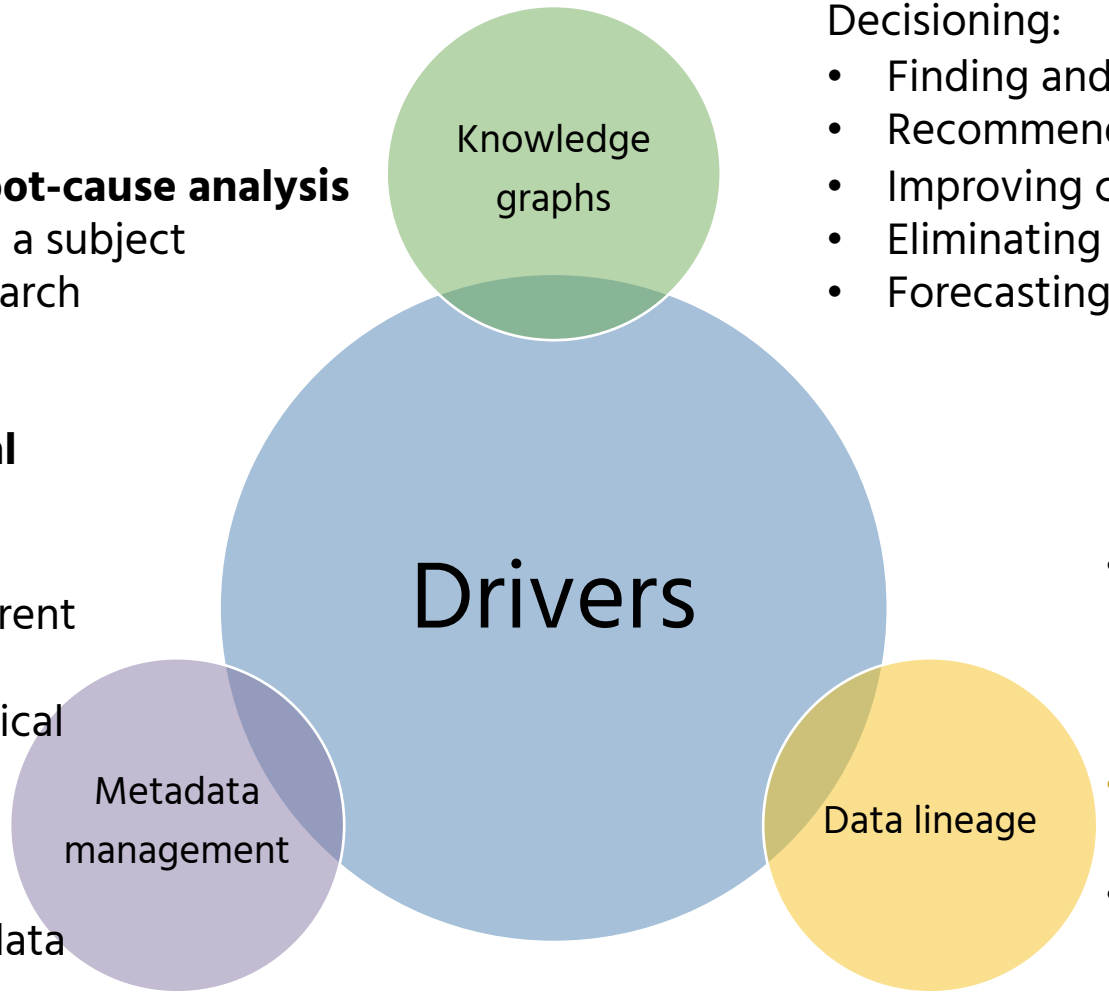
*LaPlante, Alice, and Safari. *Data Fabric as Modern Data Architecture*. Van Duuren Media, 2021.

USE CASES

Use cases of all concepts intersect each other

Actioning:
- **Data lineage**
- Data catalog
- **Impact and root-cause analysis**
- Single view on a subject
- Information search

Decisioning:
- Finding and preventing
- Recommending products
- Improving customer experience
- Eliminating duplicates
- Forecasting business needs

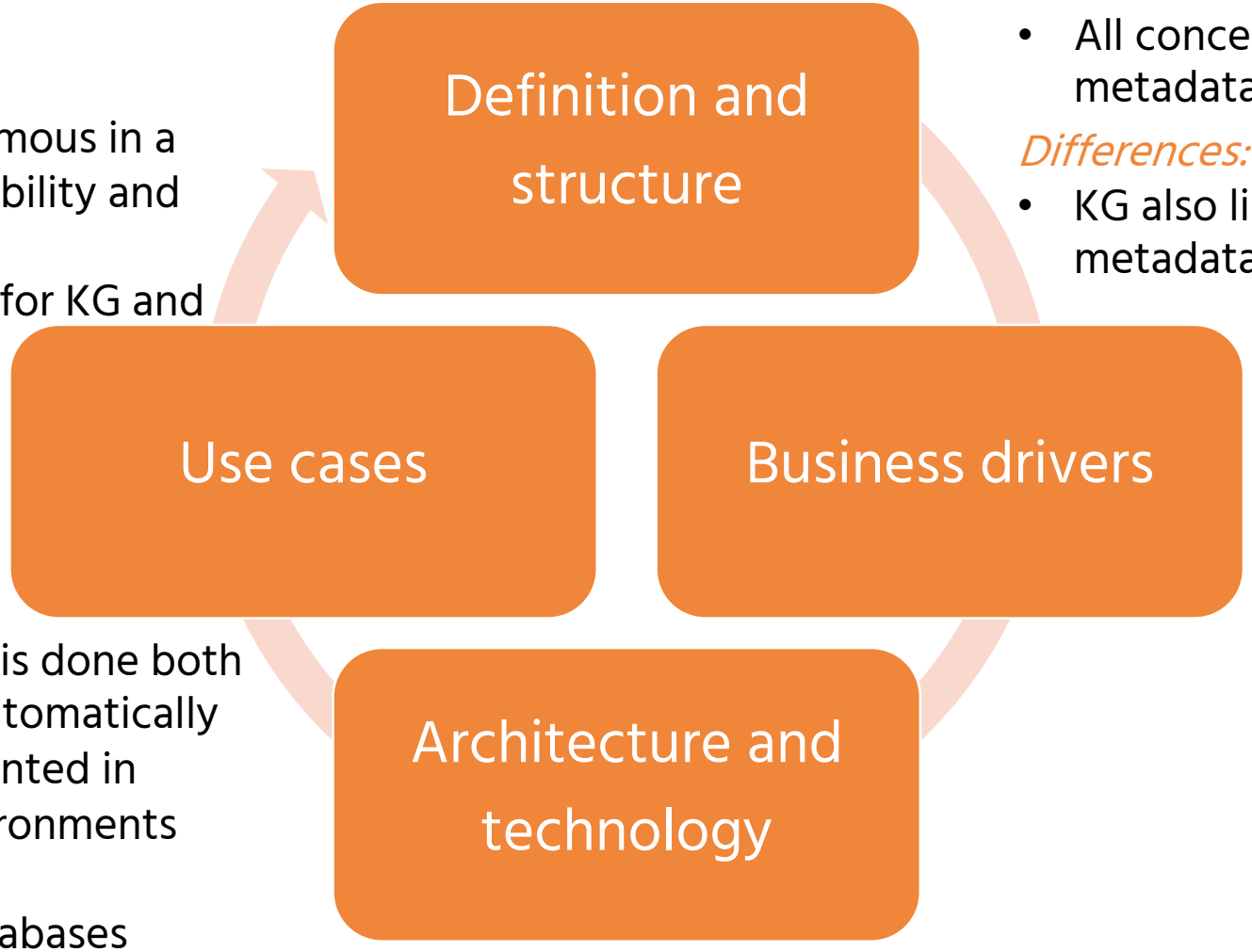


- Document and manage **organizational knowledge of data-related business terminology**
- **Collect and integrate data** from different sources
- Establish and enforce the use of technical metadata standards to enable data exchange
- Ensure metadata quality
- Provide standard ways to make metadata accessible to metadata consumers

- **Impact and root-cause analysis for multiple data** management initiatives
- Explain data origin and data transformation
- **Integrate metadata and data**

SIMILARITIES AND DIFFERENCES

Metadata management (MM), knowledge graphs (KG), and data lineage (DL) capabilities have similarities and differences



Similarities:
- All concepts deal with metadata

Differences:
- KG also links data and metadata

Similarities:
- Data integration
- DL and KG are synonymous in a context of data traceability and connections
- MM is the foundation for KG and DL

Differences:
- Some use cases vary

Similarities:
- Data integration, providing context, business change, and IT costs reductions are common drivers

Differences:
- Regulatory requirements are one of the key drivers for data lineage

Similarities:
- Documentation is done both manually and automatically
- Can be implemented in different IT environments

Differences:
- Repository's databases

A three-headed serpent: knowledge graphs, metadata management, and data lineage

Why?

- Demonstrate similarities and profound relationships between these three concepts

What?

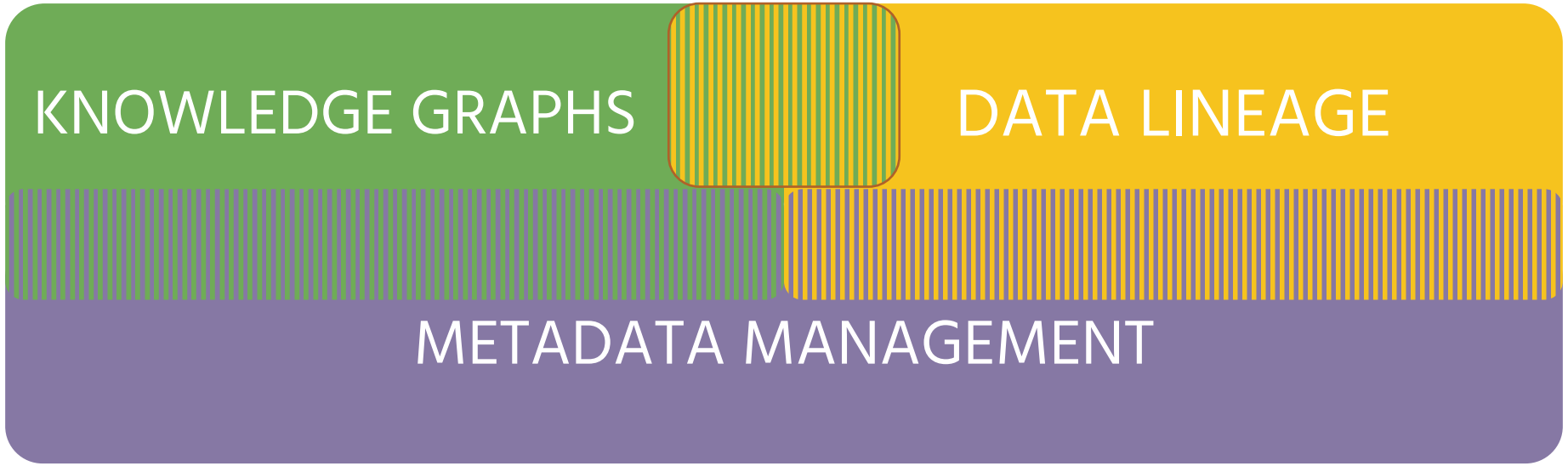
Describe each concept in terms of:

- Definitions and structure
- Business drivers
- Architecture and technology
- Use cases

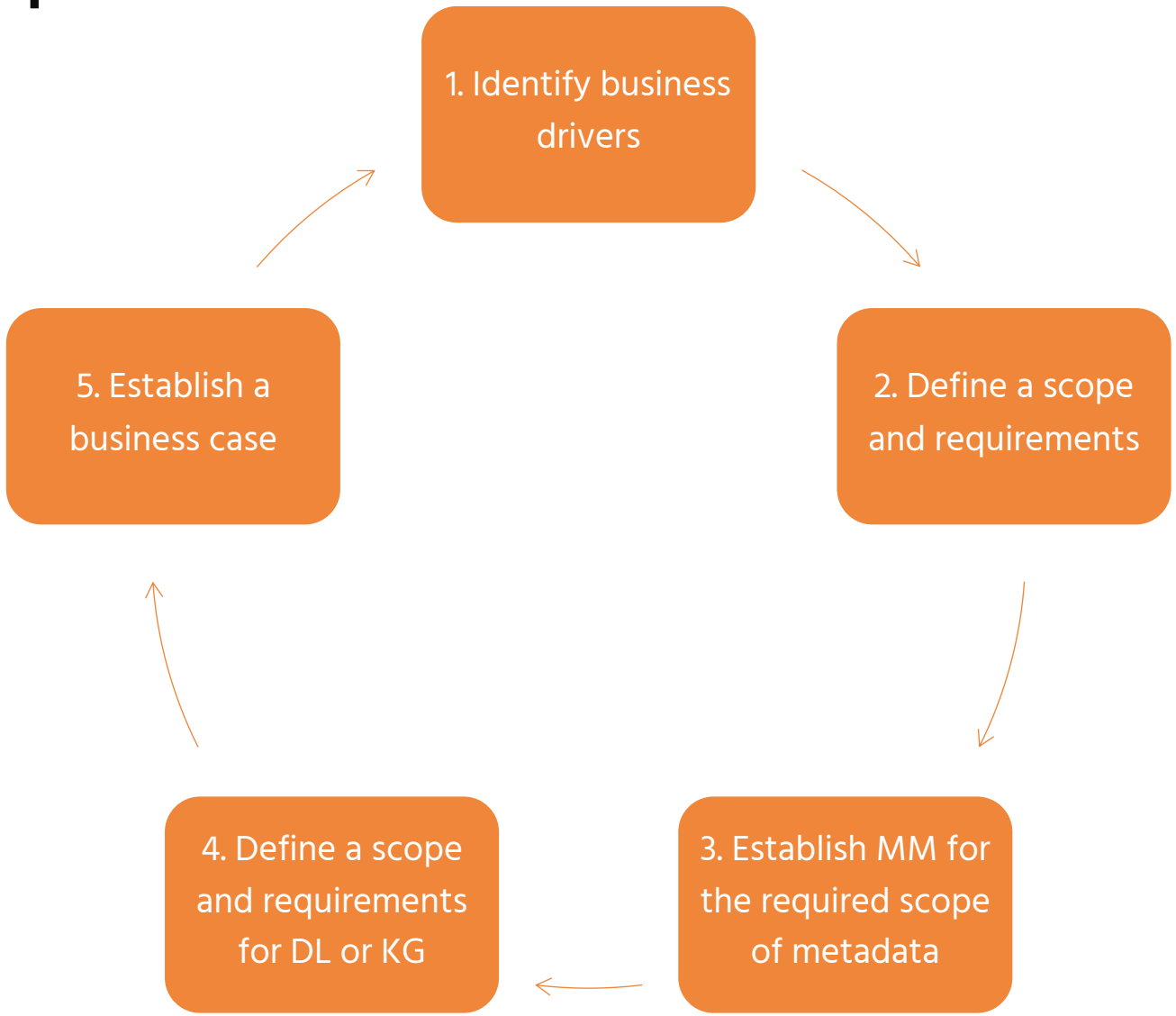
How?

High-level approach of integrated implementation

Metadata management creates a foundation for KG and DL. All three concepts intersect each other in some areas.



The relationships between MM, KG, and DL concepts define the logic of their implementation



Poll 3

Follow-up (multiple choice)

- Request a free strategy session with Data Crossroads
- Request a Trigyan demo
- Request a meeting to learn more about Trigyan

THANK YOU!

Do you have any questions?
Get in touch with us at

DATACROSSROADS.NL

https://datacrossroads.nl/free-strategy-session/

Or let's connect on LinkedIn:

www.linkedin.com/in/irina-steenbeek

TRANSFORM “UNCURATED DATA” TO “ACTIONABLE DATA”

To learn more and request a demo

Call: +1 (732) 516-1111
Email: info@ trigyan.com
Web: trigyan.com

Q & A