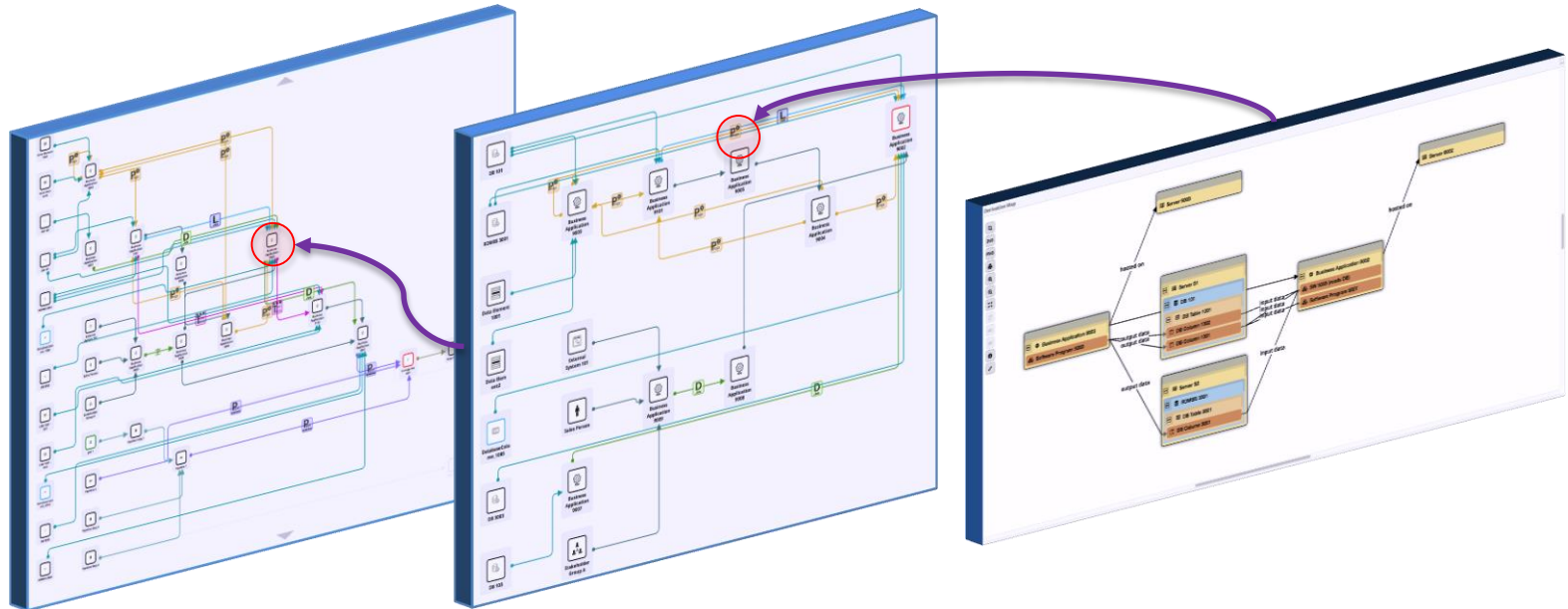# NAVIGATING DATA LINEAGE AND ITS IMPACT WITH INTERACTIVE VISUALIZATION MAPS

Data Lineage as Defined by Wikipedia

- **Data lineage** includes the data's origins, what happens to it and where it moves over time.

- **Enriched data lineage** information may include:
  - reference data values
  - data models
  - business vocabulary
  - enterprise information systems linked to the data points and transformations
  - data quality test results
  - data usage
  - stewardship and program management information

Data Lineage is about Network of Connections

In Understanding a Journey, a Variety of Information is Important

How much do we capture?

- The scope of the data lineage information is usually determined by data governance - based on:
  - Regulatory compliance needs
  - Enterprise data management strategy
  - Data impact and reporting needs
  - Critical data elements of the organization.

Operational Efficiencies?   Compliance?   Analytical Insights?

- The scope then determines the nature and volume of metadata required to represent lineage.

Today's Demo

Part 1: Lineage in the financial domain

- Demonstrates a Lineage Model (LM) with Derivation Maps (DMs) for Federal Reserve Y-9C compliance reporting of Assets-For-Sales (AFS) and Held-To-Maturity Assets (HTM).

Part 2: Examples of specific capabilities / mechanisms for managing lineage models, display & visualization

- Essential to support actual use and exploration of lineage given the complexity and (very) large scope of many real world data lineage ecosystems

Part 1: An Example of Data Lineage Tracking for Compliance Reporting

TopQuadrant™

Let's look at an example from finance … FR Y-9C

✓ Federal Reserve Board Form

✓ Used to monitor financial institutions

✓ Requires data to be reported to assess the financial condition of holding companies.

FR Y-9C
OMB Number 7100-0128
Approval expires July 31, 2018
Page 1 of 65

Board of Governors of the Federal Reserve System

Consolidated Financial Statements for Holding Companies—FR Y-9C

Report at the close of business as of the last calendar day of the quarter

This Report is required by law: Section 5(c) of the Bank Holding Company Act (12 U.S.C. § 1844) and Section 225.5(b) of Regulation Y (12 C.F.R. § 225.5(b)) and Section 10 of the Home Owners Loan Act (12 U.S.C. § 1467a(b)).

This report form is to be filed by holding companies with total consolidated assets of $1 billion or more. In addition, holding companies meeting certain criteria must file this report (FR Y-9C).

regardless of size. See page 1 of the general instructions for further information. However, when such holding companies own or control, or are owned or controlled by, other holding companies, only the top-tier holding company must file this report for the consolidated holding company organization. The Federal Reserve may not conduct or sponsor, and an organization (or a person) is not required to respond to, a collection of information unless it displays a currently valid OMB control number.

NOTE: Each holding company's board of directors and senior management are responsible for establishing and maintaining an effective system of internal control, including controls over the Consolidated Financial Statements for Holding Companies. The Consolidated Financial Statements for Holding Companies is to be prepared in accordance with instructions provided by the Federal Reserve System. The Consolidated Financial Statements for Holding Companies must be signed and attested by the Chief Financial Officer (CFO) of the reporting holding company (or by the individual performing this equivalent function).

I, the undersigned CFO (or equivalent) of the named holding company, attest that the Consolidated Financial Statements for Holding Companies (including the supporting schedules) for this report date have been prepared in conformance with the instructions issued by the Federal Reserve System and are true and correct to the best of my knowledge and belief.

Date of Report: **June 30, 2016**
Month / Day / Year (BHCK 9999)

Printed Name of Chief Financial Officer (or Equivalent) (BHCK C490)

Legal Title of Holding Company (TEXT 9010)

Signature of Chief Financial Officer (or Equivalent) (BHCK H321)

(Mailing Address of the Holding Company) Street / PO Box (TEXT 9110)

Date of Signature (MM/DD/YYYY) (BHTX J196)
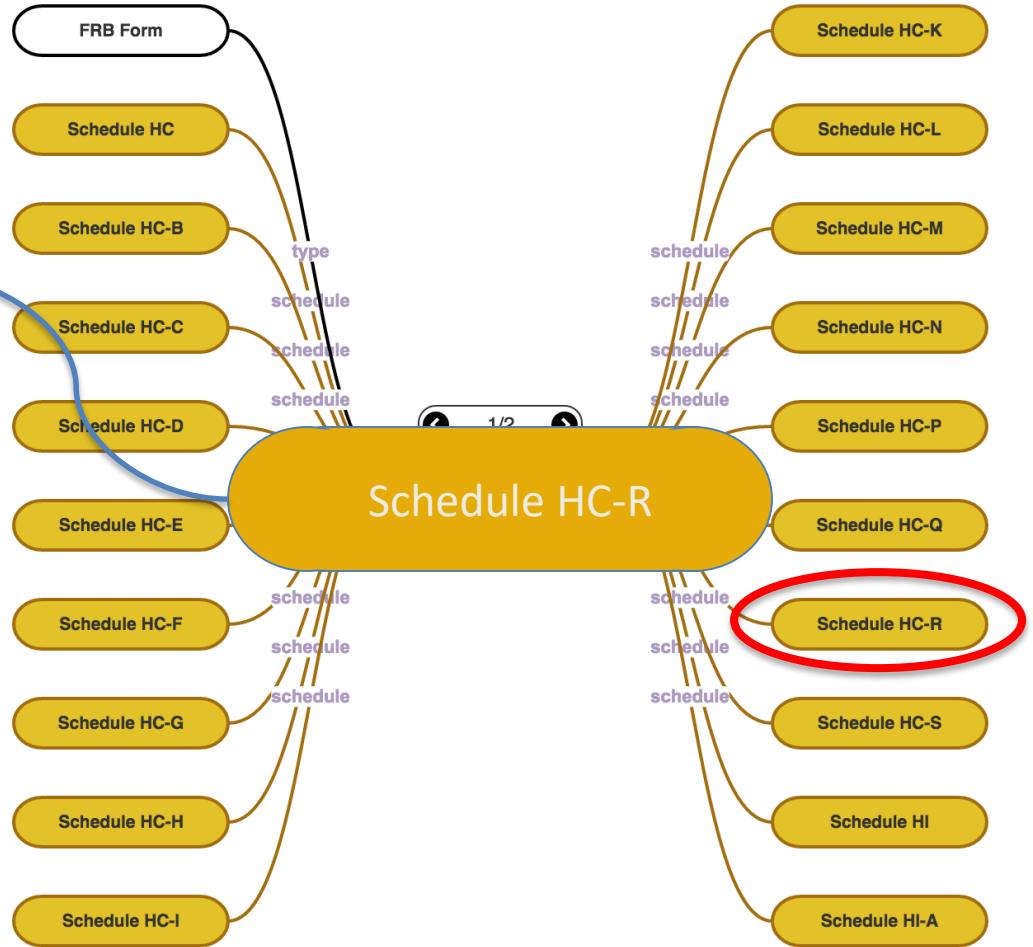
City (TEXT 9130)   State (TEXT 9200)   Zip Code (TEXT 9220)

Person to whom questions about this report should be directed:

FR Y-9C: HC-R Securitization Exposures

Lots of Schedules

Form FR Y-9C

✓ Many calculations…
How to know it's right?

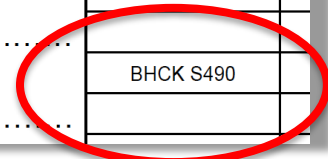# Schedule HC-R Data Field:
Total On-balance Sheet Securitization Exposures

TopQuadrant™

Data Lineage Tracking (From FR Y-9C, page 53 of 65)

9. On-balance sheet securitization exposures:
   a. Held-to-maturity securities .... | BHCK S475
   b. Available-for-sale securities .... | BHCK S480
   c. Trading assets .... | BHCK S485
   d. All other on-balance sheet securitization exposures .... | BHCK S490

Individual data elements and forms are represented in EDG

BHCK S490:
Total Securitization

BHCKS490 (MDRM Data Item)

Identifiers Metadata

type:	MDRM Data Item
label:	BHCKS490
identifier:	BHCKS490
asset name:	TOTAL ON-BALANCE SHEET SECURITIZATION EXPOSURES, ALL OTHER ON-BALANCE SHEET SECURITIZATION EXPOSURES

Status Metadata

effective start date:	2015-03-31
effective end date:	1999-12-31

Documentation

reference link:	http://www.federalreserve.gov/apps/mdrm/data-dictionary/search/item?keyword=S490&show_short_title=False&show_conf=False&rep_status=All&rep_state=Opened&rep_period=Before&date_st

Other Properties

confidential:	No
used in form:	FR Y-9C

TopQuadrant™

FR Y-9C

HC-R

BHCK
S490

APP

Form ➡️ **Schedule** ➡️ **MDRM
Data Item** • • • ➡️ **Apps & data
elements**

Where Does BHCK-S490 Come From?

Supporting both Business (Top-down) Lineage and Technical (Bottom-up) Lineage is Essential

High-level View: for example at the business level, managers just want to see that the FR Y-9C was completed and delivered in the appropriate report.

(illustrated metaphorically as the high-level map from a bank in NYC to the FRB in DC.)
• Was all of the data needed compiled?
• Was the report completed and filed on time?
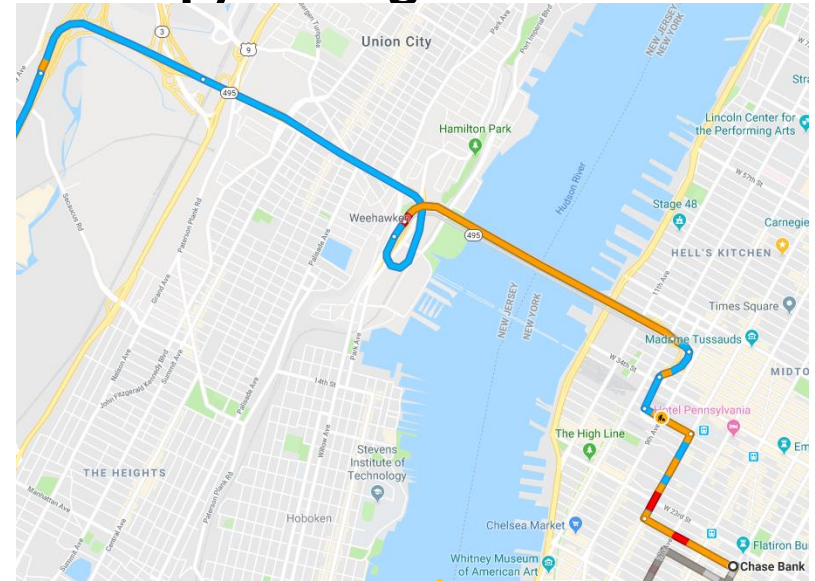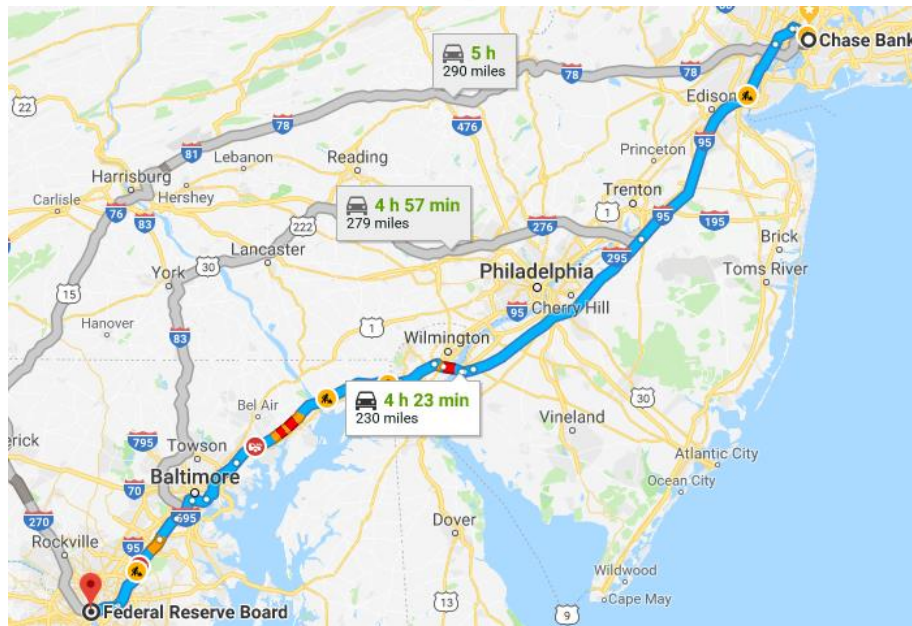
Drill-down View: at the technical level, CIOs/CDO/IT Managers et al need to be able to know, track and verify every step and turn the data took in compiling the FR Y-9C report– in case of audit by FRB etc.

(metaphorically, e.g., understanding the nitty-gritty details (complex path, possibly including detours) necessary just to get the out of NYC and onto the highway to DC)
• Where did that data come from?
• How did it get there?

Business Terms, Data/Technical/Enterprise Assets

- ## What you will learn is:

  1. How a SEC Business Term is mapped to Data/Technical Assets
 2. How Business Activities can also be a starting point for lineage
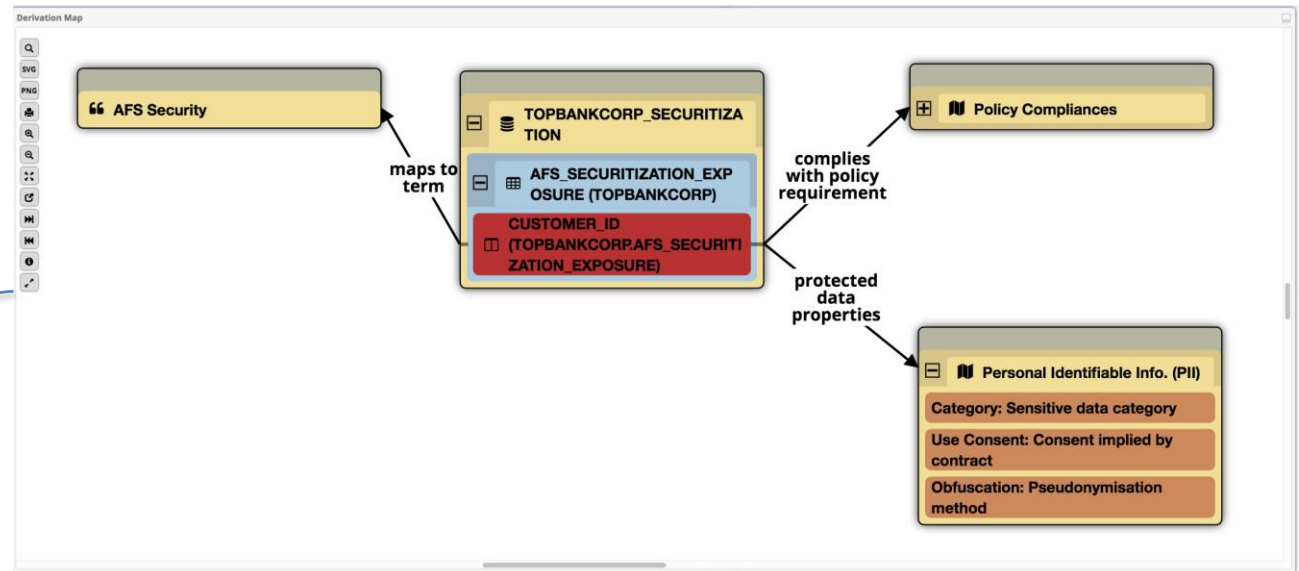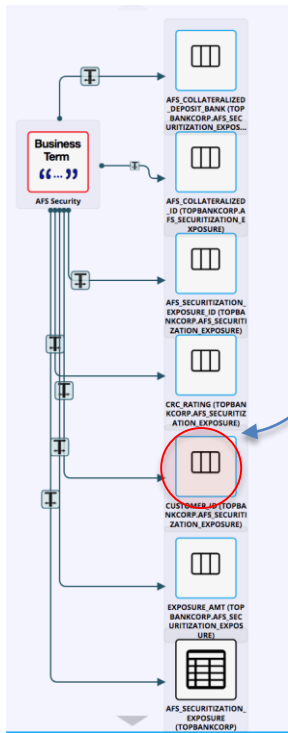
  3. Navigation within and from the LineageGram
  4. Drilldown to then Derivation Maps (DMs)
  5. Navigation within and from a DM

Demo – Part 1

TopQuadrant™

- **Fan-in/Fan-out** on a node in the LineageGram

- **More Upstream/More Downstream**

- Data Flow Dependencies

- Function Level Dependencies

- Logical Flow Dependencies

- Program Level Dependencies

- Recursive Dependencies

Display and Navigation of Complex Lineage Models

- ■ What you will learn is:

  1. How a LineageGram manages the display/visualization of a long chain of upstream dependencies (Reach)

  2. How the number of incoming and outgoing connections of any given node can be constrained (Breadth)

TopQuadrant™
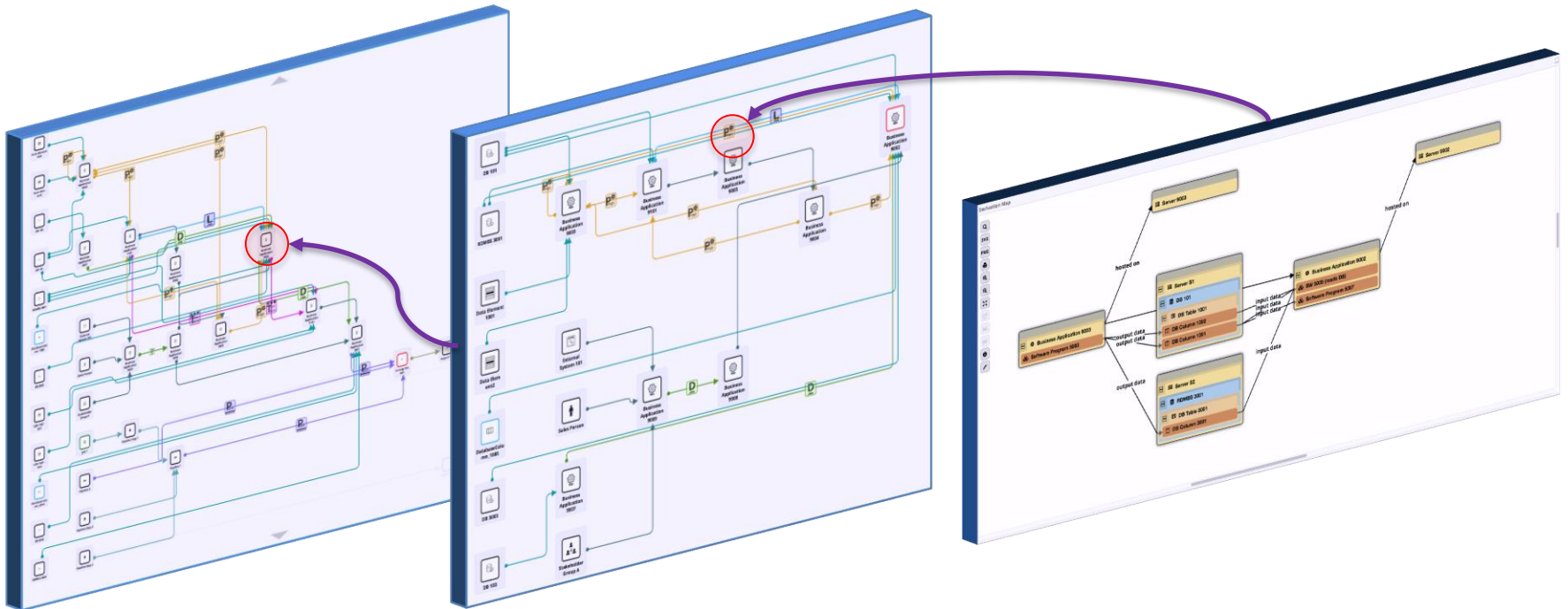
Demo – Part 2

What you saw in the demos – How did we display lineage?

TopQuadrant™

- Presentation of the lineage can be done at various zoom levels to simplify the vast information

- At a very high level, data lineage shows what systems the data interacts with

- As the granularity increases it goes into the data point level where it can provide the details of the data point

- All of which has controlled navigation based on "reach and breadth" and filtering

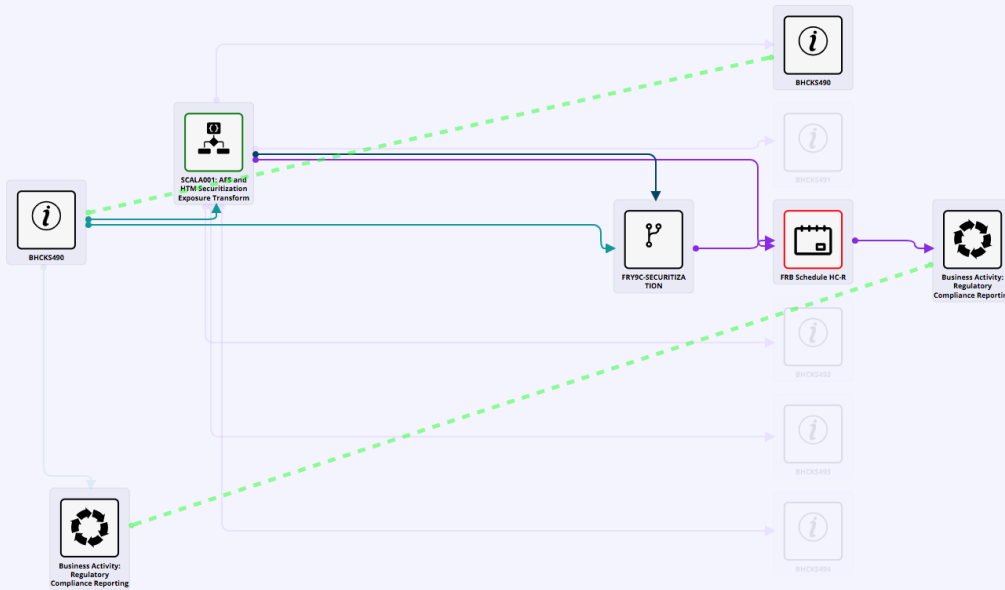# Beyond Visualization to Knowledge-Based Navigation

A visualization is a User Interface for exploring the Data Governance Knowledge Graph at different levels of details.
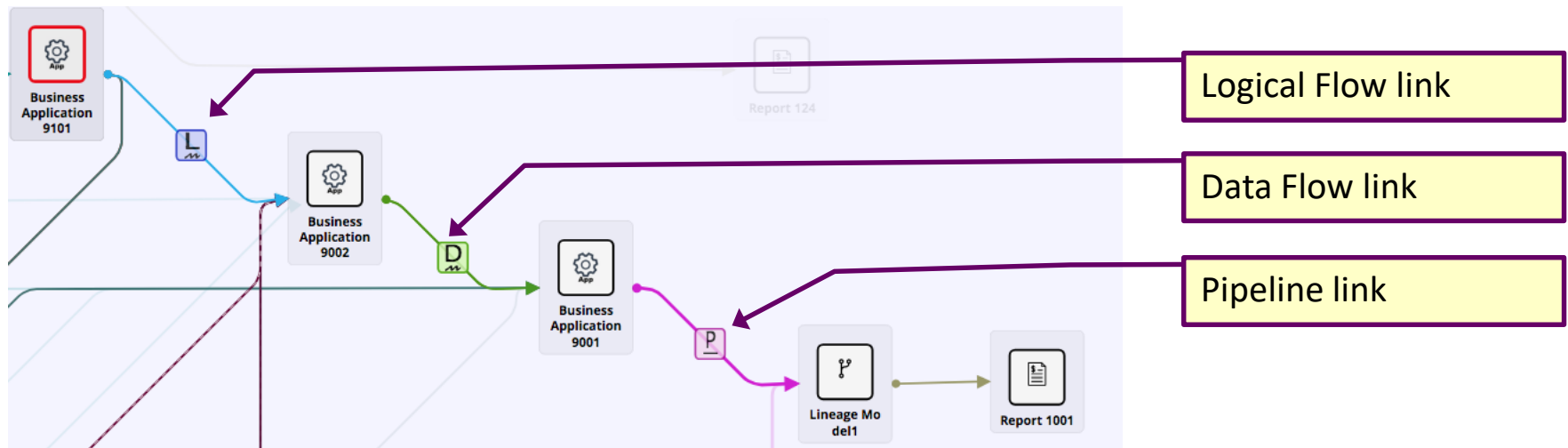
Visualization uses "Graph Operators" for enhanced usability and scalability

Nodes "*know their place*" in a lineage map. Selecting a node causes edges not in the selected upstream flow to "grey-out":

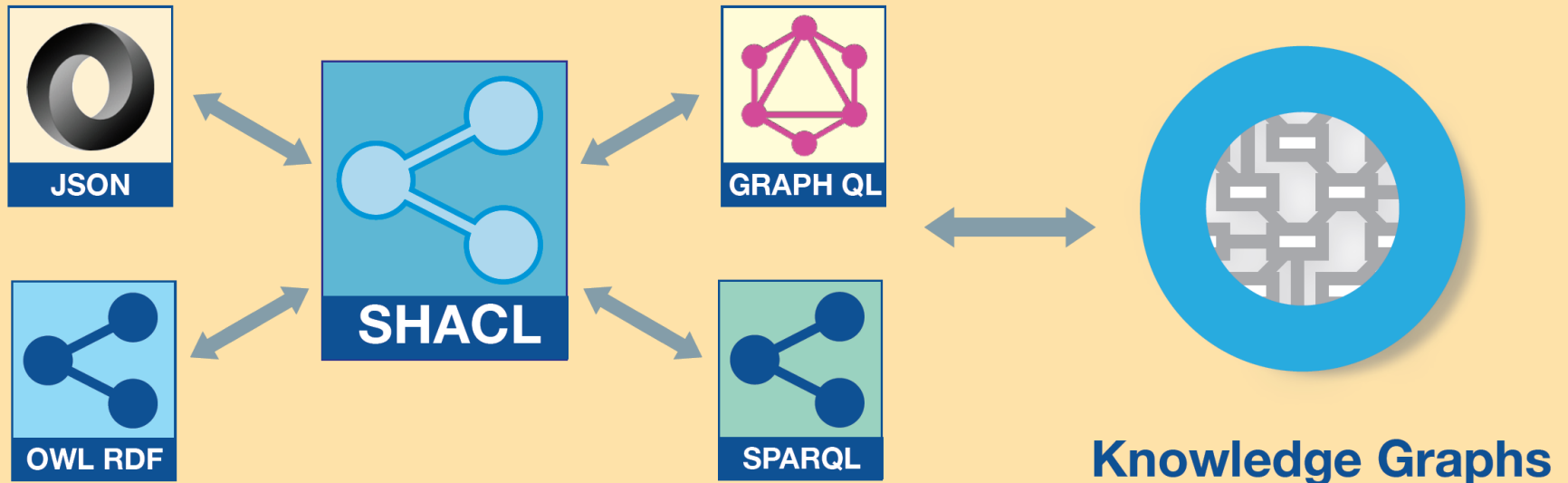Gather operation brings immediate nodes together in a cluster arrangement:
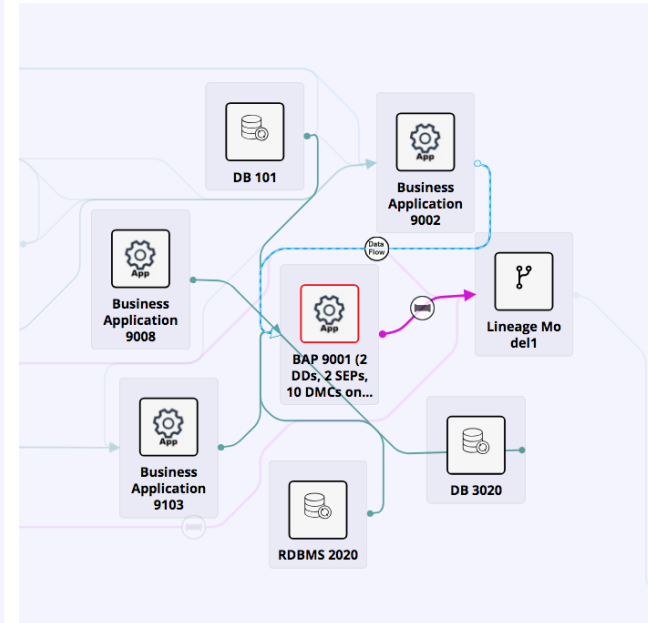
To Learn More …

EDG Product Info:

- http://www.topquadrant.com/products/topbraid-edg/
- http://www.topquadrant.com/products/topbraid-edg-gov-packs/
- https://www.topquadrant.com/knowledge-assets/whitepapers/#BUILDINGBRIDGES

Contact us: at info@topquadrant.com to:

- Discuss data governance solutions
- Request a more targeted demo of TopBraid EDG
- Ask for a free EDG evaluation account