DATA LINEAGE DIAGRAMS

It's time to rethink the paradigm for data lineage

Tom Dejonghe & Erwin Dral
Product Managers @ Collibra

May 2\textsuperscript{nd}, 2017
Data Lineage Diagrams: 
A Paradigm Shift for Information Architects

https://www.colibra.com/blog/data-lineage-diagrams-paradigm-shift-information-architects/
Would Collibra post more "visionary" blogs, like this blog on data lineage?

- No, these blogs are confusing and a waste of time
- Yes, but please make it more obvious that it is still a vision
- Yes, I really like that Collibra shares their vision for the future

100%
SUBSCRIBE TO THE COLLIBRA BLOG

The place to go to learn about true data governance

The Collibra Blog is the place for CDOs, data stewards, and data citizens to go to learn about true data governance. We cover popular topics such as data governance readiness, data lineage, big data governance, and more.

As a subscriber, you'll enjoy:

- Regular posts written by the foremost thinkers and leaders in data governance
- Education, insight, and best practices to help you start and grow your data governance initiative
- A monthly recap so you can catch up on what you missed

Subscribe to the Collibra Blog

First Name

Last Name

Company

Email Address

SIGN ME UP

Subscribe

http://learn.collibra.com/Blog-Subscription
Some Data Facts and Figures

1 Zettabyte
Global internet traffic surpasses 1 Zettabyte (1 billion terabytes) in 2016
Cisco Research paper

90%
90% of the world’s data was generated in the past two years
Science Daily

50X
Data volume in the enterprise is expected to grow 50x year-on-year
Hortonworks CEO Rob Bearden

85%
The most important thing to recognize is that 85% of that data is coming from net-new data sources
Hortonworks CEO Rob Bearden
Collobra: the System of Record for Data Change Agents

Know where the data comes from
Find Catalog

Know what the data means
Understand Data Dictionary
Business Glossary
Reference Data

Know that the data is right
Trust Policy Manager
Data Helpdesk
Stewardship

All activities and information surrounding the data, its meaning and its use.
All activities and information surrounding the data, its meaning and its use.
Data Lineage is Huge Challenge Today

**Scattered application landscape**
- Front to back office
- Several data warehouses
- Multiple reporting tools

**How do you manage gigantic spider web?**
- Rethink the paradigm of data lineage
- Consider data lineage as active, not passive
Rethink the Paradigm

Interactive versus static
• Allow different kind of interactions from within the diagrams

Starting place instead of ending place
• Allow to start from a conceptual model first
• Allow to work on a future state (to be) instead of current state

Rely on the community instead of a happy few
• Include crowdsourcing and collaboration features

See a diagram as a governed asset
• Save a snapshot of a data lineage diagram
• Govern diagrams in terms of ownership, definition and documentation, etc.
In the Meantime?

Leverage today’s capabilities
• Automate what can be automated (Catalog / Connect)
• Focus on critical data elements and data usage (since the only lineage that really matters)
• Add a conceptual, logical and business layer on top

Leverage tomorrow’s capabilities
• Our new release will focus on drastically better look and feel, better navigation and easy configuration

Help us shape your future capabilities
• Engage with us at the Product Booth
• Join the User Participation Program: upp.collibra.com
DATA LINEAGE DIAGRAMS
ROADMAP AND DEMO

Erwin Dral
Disclaimer

This presentation is intended to outline our general product direction. It is for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Collibra products remain at the sole discretion of Collibra.
Diagrams 5.1

Makeover: *Style & substance*

- User-defined colors and symbols
- Optimized layout to fit larger diagrams
- Toolbar and node menu with more ways to interact
- Ability to share diagram views with specific users, groups or roles
Diagrams 5.1

**Navigate:** Find your way through volume and complexity

- Start with a condensed diagram, then expand nodes to see more details inside
- Investigate only the incoming side (‘lineage’), the outgoing side (‘where-used’), or both
- Explore individual relation types
- Trace path between two nodes
Diagrams 5.1

Overlays: Pack more information in less space

- Data quality score, status, attributes, tags
- Responsibilities
- On complex relations, too
Diagrams 5.1

View Editor: Graphical UI to define views

- Construct the diagram view by selecting asset types and relation types from the metamodel
- Choose layout and settings in the UI
- No more JSON!
KEEP CALM IT IS DEMO TIME
QUESTIONS?