

Semantic Web and the New Industrial Revolution

SWAT4HCLS

4 Dec 2018

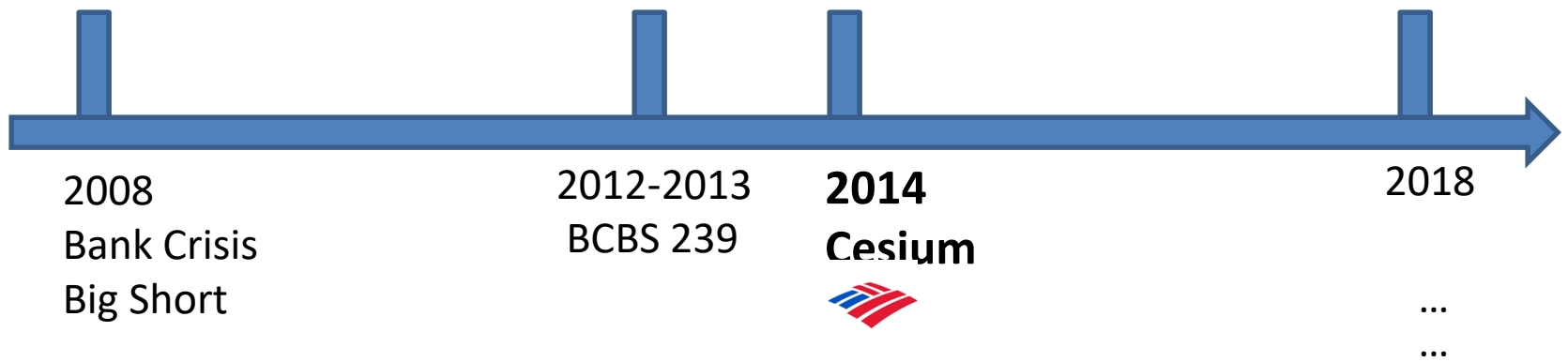
Dean Allemang

Working Ontologist, LLC

dallemang@workingontologist.com







Cesium Reference Data Ontology

August 2015

Cesium is

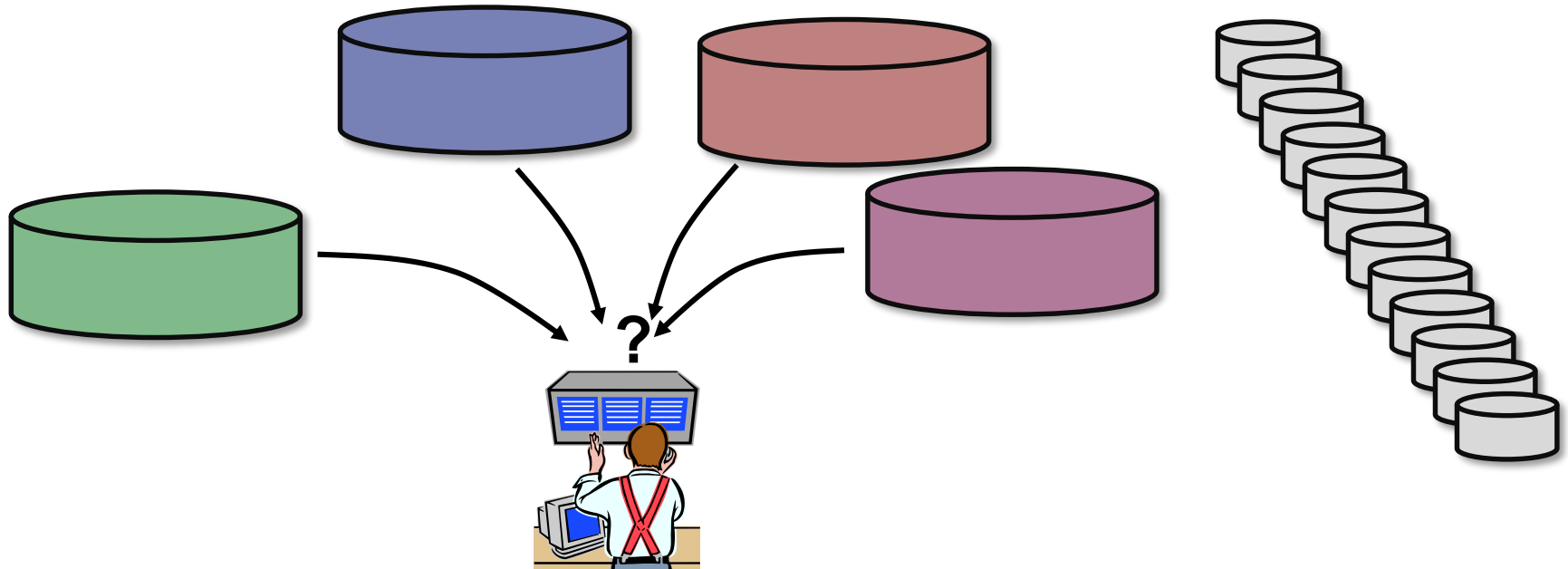
- **A Platform for Reference Data at Bank of America / Merrill Lynch**
- **A Single source for all client data in markets**
 - Integrates and normalizes various systems of records
 - Regulatory attributes
 - What do we need to know about our clients and affiliates to comply to regulations?
 - Consistent linkages between clients, accounts and other aspects
- **Provides a Global consistent footprint**

Cesium went live in Q1 2014

Sustainable Extensibility

The problem of Sustainable Extensibility

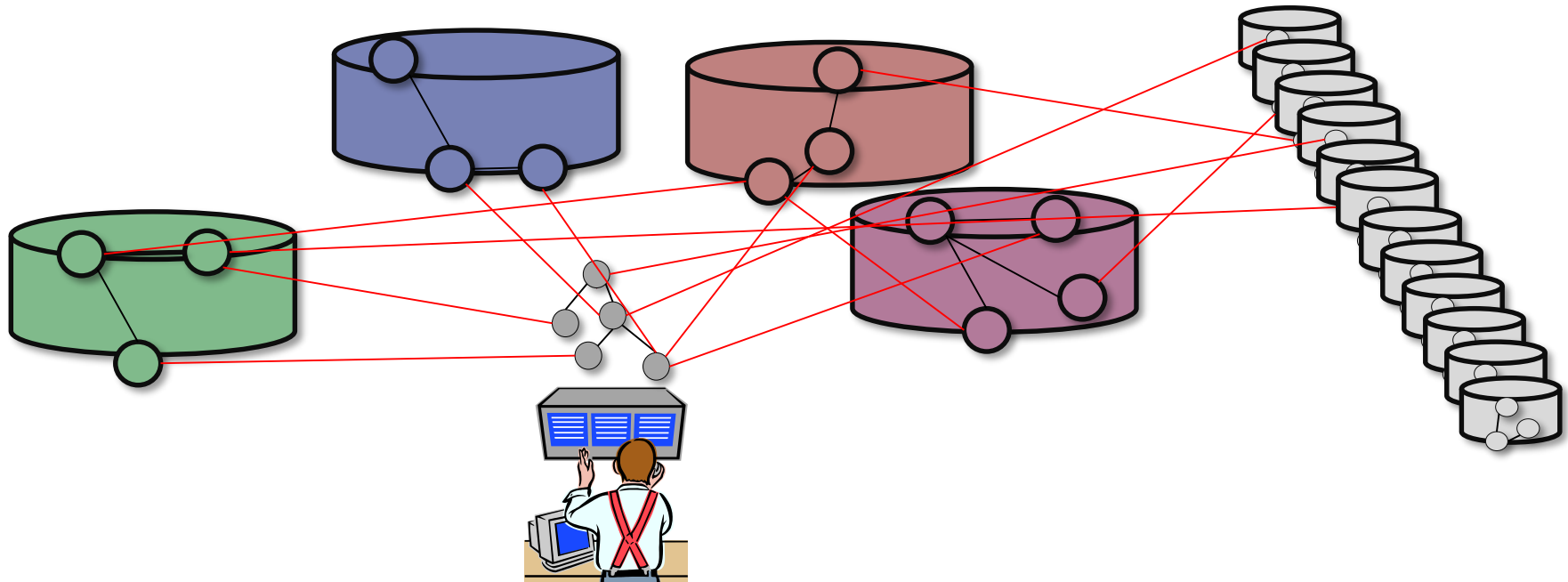
- Bank of America / Merrill Lynch has several systems of record for clients, accounts, affiliates, etc.
- How do you get a single view of all that data ...
- ... especially when there are more databases around the corner?



Sustainable Extensibility

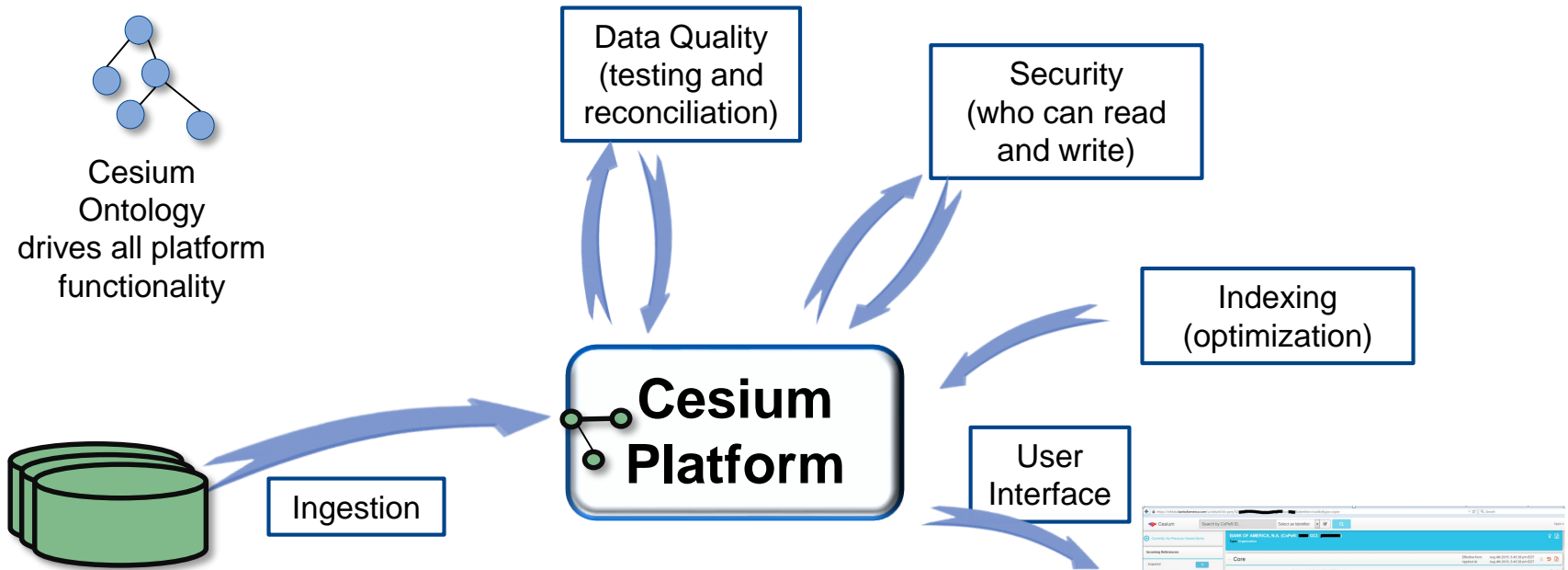
The Cesium solution to Sustainable Extensibility

- Build a model of the data
- Virtualize legacy data as graphs
- Map the graphs and datasets
- Include more data sets as time goes on.



- **Cesium provides a Single model for**
 - Client data
 - Firm data
 - Instrument Data
 - “Primitives”
 - aka controlled vocabularies, code lists, data points, value sets, etc.
 - Uses W3C SKOS for controlled vocabularies
- **Tracks provenance (where the data came from)**
 - Uses W3C Prov-O
 - Displays information about the data source to end users

Model-driven Platform



“One of the key things that has driven the success of our platform is the ability to use the ontology to drive the platform end to end. Starting with ingestion which governs how legacy formats are converted to RDF, data quality checks which attest to the correctness and consistency of the data, security which governs who can publish and see the data, how the data is indexed for efficient retrieval to how the data is actually rendered in the end user UI – these are all driven from a single model. A large part of this is engineering but the engineering would not have been possible without adopting RDF as a strategic choice.”

Cesium – Ontology Browser Detail

The screenshot displays the Cesium Ontology Browser interface. At the top, the search bar contains the text "Search by CoPeR ID...". The main header shows "BANK OF AMERICA, N.A. (CoPeR: [redacted] GCI: [redacted])" with a "Unified id" label. Below this, the "Core" section lists various attributes such as Full Name, Mixed Case Legal Name, Operational Name, Creation Date, Incorporation Date, Country of Domicile, Country of Incorporation, Country of Taxation, Client Registration Type, Business Classification, Legal Form, NAICS, SIC, Is Internal, RegW Entity Type, RegW Entity Source, Is Regulated, and Business System. The "Identifiers" section at the bottom lists CoPeR ID, GCI, CPR Acronym, and Mnemonic. On the left, the "Incoming References" section lists various roles and their counts, such as "Acquired" (11), "Beneficial Owner For" (1851), "Booking Entity For" (13637), "Credit Parent For" (1088), "Duplicate Invalid Parties" (10), "Investment Manager For" (8612), "Legal Parent For" (606), "Linked Sales Node" (2925), "Merged With" (41), "Primary Owner For" (301), "Primary Sales Entity" (1), "Profile Party" (429), and "Secondary Sales Entity" (27). A "Linked Data" label points to this section. A "Bi-temporal Data" label points to the "Effective from:" and "Applied at:" fields. A "Show History" label points to a history icon. A "Linked to Metadata" label points to the "Country of Incorporation" field. A "Search" button is visible in the top right. A list of features is shown in a pink box on the right.

Annotations:

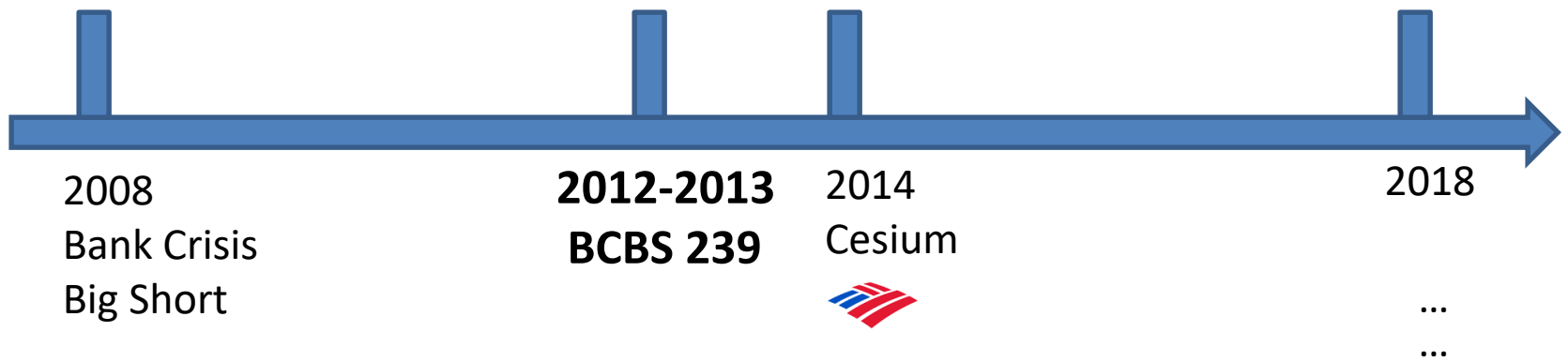
- Unified id
- Bi-temporal Data
- Show History (only appears if there is history)
- Linked to Metadata
- Linked Data
- Aspects

Search across 100+ ids

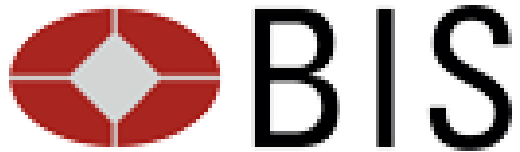
- Search across names
- Filtering
- Faceting
- History
- Navigation
- Dev mode

Platform Features

- **RDF-based open model**
 - **Based on W3C standards including RDF, SKOS and Prov-O**
- **Real-time and Bi-temporal**
 - **Real-time end users**
 - **Current view or bi-temporal snapshot**
- **Extensions, Overrides and Defaults**
 - **The model can be extended to cover new data sets**
 - **Extensions include certain non-monotonic logic like defaults and overrides**
- **Workflow and Data Quality control are integrated into the platform**



BCBS 239

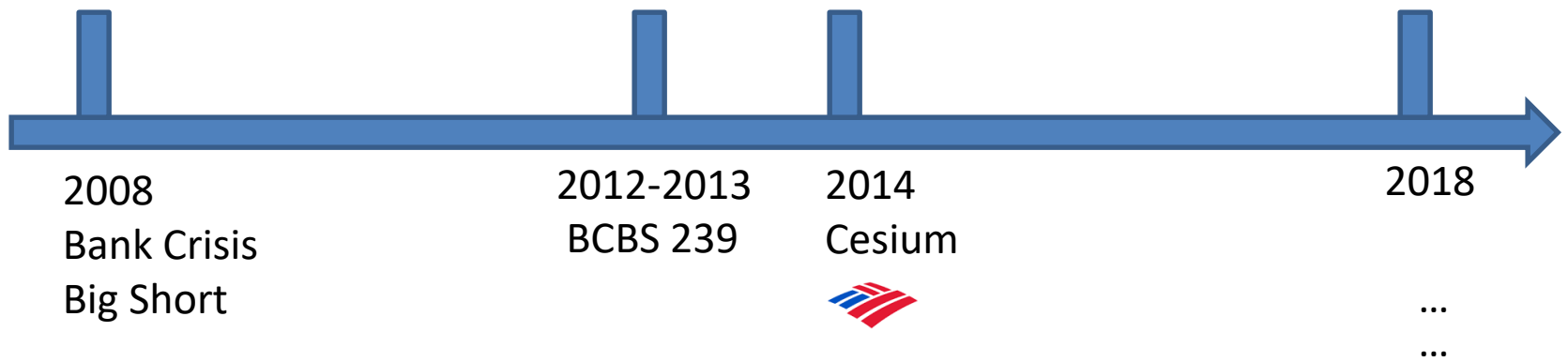


Banks need to manage their risk data better.

Principles for doing that:

Summary of BCBS 239 Principles

- Governance - govern your risk data management and reporting
- Infrastructure - in good times and bad
- Accuracy and Integrity - Aggregate automatically to get integral picture
- Completeness - from all viewpoints
- Timeliness - automated
- Adaptability - respond to lots of stakeholders
- Accuracy - reconciliation and validation
- Comprehensive - all aspects of risk data
- Clarity and Usefulness - Data for use
- Frequency - let me know when you'll report
- Distribution - responsibility to provide (not just need to know)



FIBO Basics



FIBO Basics



SKOS



FIBO-V



FIBO Basics



HTML/JS etc.



FIBO-Glossary

FIBO Use Cases

- 1. Data Harmonization:** factual reference point for MEANING (not words) replaces spreadsheet-driven reconciliation and promotes process automation [STP, trust and confidence, save \$]
- 2. Structural Validation:** alignment to precise meaning tests conformance of content to ensure required properties and allowable values [quality assurance; smart contracts; Blockchain]
- 3. Data Integration:** alignment of content to explicit meaning makes it easier to process and integrate data from federated sources [reduce errors; reusable concepts, save \$]
- 4. Flexible Analysis:** separates meaning from structure and links concepts without having to restructure columns and rows [graph capability; inference; classification; aggregation]
- 5. Machine Learning:** Ontologies are used as inputs into machine learning models and can be coupled with algorithms for data discovery [build inventory and enhance learning models]
- 6. Enterprise Data Rationalization.** Describe what a data asset (e.g., table in an RDB) means by reference to external meaning.

Metadata Management in Moviemaking



<http://www.etcentric.org/etcusc-tests-production-in-the-cloud-with-the-suitcase/>



Insert Suitcase Presentation Here



Conclusions?

- You are doing Science!
 - More formal notion of data, experiment, etc.
 - Publish or Perish
 - Audience is willing to think hard about data, metadata, etc.

Data Categories in various industries

Data Category	Media	Finance	HCLS
Image	Video, Stills	??	Satellite images, Xrays, Crop photos,
Streaming Data	Twitter	Transactions, offers	Clinical data, field measurements
Measurement	Tagging	??	Experimental data
Derivative data	Market data	Ratings	Published results
Vocabulary	Character lists, Authorities	LCC, statuses	Phenotypes, SNOMED, ICD,
Schema	Ontology (EIDR, media ontology)	FIBO	???