The First Step in Information Management

Trends in Emerging Data Technologies

DAMA NY Fall Event

175 Water St, New York – October 18, 2018

Malcolm Chisholm Ph.D.
Chief Innovation Officer
First San Francisco Partners

Quick Introduction

About First San Francisco Partners

First San Francisco Partners is entirely focused on helping our Clients leverage data as a **value-producing asset** through improved information management. We are a group of experts from the industry who can help you create a strategy, align your organization and deliver business value in both the short and the longer terms. We do this via:

- Enterprise Information Management Roadmaps
- Data Architecture Assessments & Strategies
- Data Quality Assessments & Strategies
- Technology Vendor Analysis & Evaluation
- MDM and DQ Assessments & Implementation
- Program Management
- Data Governance Build-out
- Data-centric Project Planning and Management
- Data-centric Architecture and Modeling
- Data-centric Project Implementation (e.g. MDM, Data Warehousing, Big Data)
- Data Analytics Project Implementation
- Data Privacy, Legal and Compliance Planning and Implementation

INFORMATION IS YOUR BUSINESS. *MAKING IT ACTIONABLE IS OURS*.

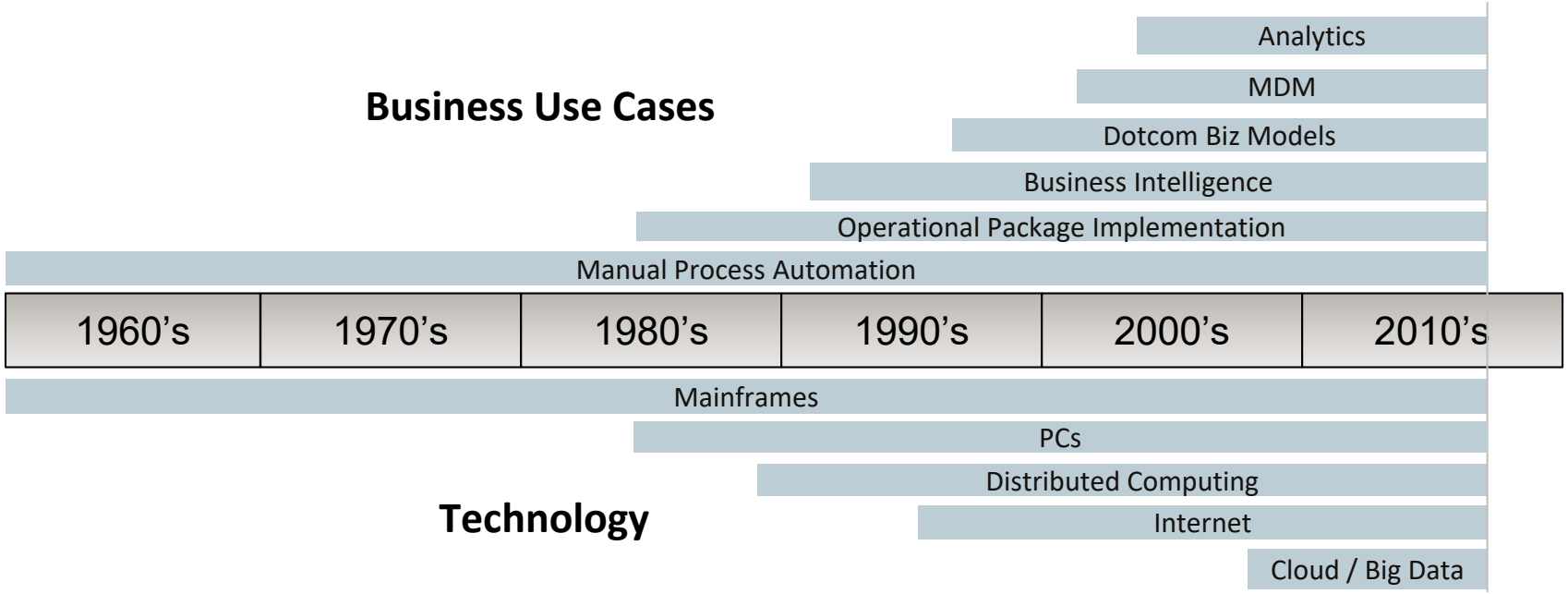
Hello!

Malcolm Chisholm is the Chief Innovation Officer of FSFP and is a recognized expert in data governance and data management with more than 25 years of industry experience. He was the recipient of the prestigious 2011 DAMA International Professional Achievement Award and is also a leading author and speaker at conferences in Europe and North America. Malcolm's published works include *Definitions in Information Management* (how to create and manage high-quality definitions for data management), *How to Build a Business Rules Engine* (how to use metadata engineering to build any kind of business rules engine) and *Managing Reference Data in Enterprise Databases* (the only book on Reference Data Management).

Malcolm's robust EIM facility includes specializations in data governance, data stewardship, master data management, reference data management, Data-centric Development Lifecycle, Semantics (including terminology, definitions, taxonomy and ontology), business rules management, data architecture, data modeling, data integration, big data environments, data quality (detection and data issue management), data change management, data lineage, metadata tools, data legal/privacy/compliance, data monetization, data vendor management and end user computing governance.

Background: The Long Term Trends

From Process Centricity to Data-Centricity and The Golden Age of Data



Over time, there has been a shift from *process-centricity* to *data-centricity*. However, much thinking remains mired in the process-centric era

The Proliferation of Data Technologies

Data Technologies

Data Management

- Data Virtualization
- Data Preparation
- Extract Transform Load (ETL)
- Ingestion
- Reference Data Mgmt (RDM)
- Master Data Mgmt (MDM)
- Data Quality
- Business Rules Engine

…more…

Data Governance

- Business Glossary
- Data Dictionary
- Data Catalog
- Data Discovery
- Data Profiling
- DG Automation
- Data Modeling
- Data Lineage

…more…

Migration To The Cloud

Data-Relevant Infrastructure is Changing

ON-PREMISE

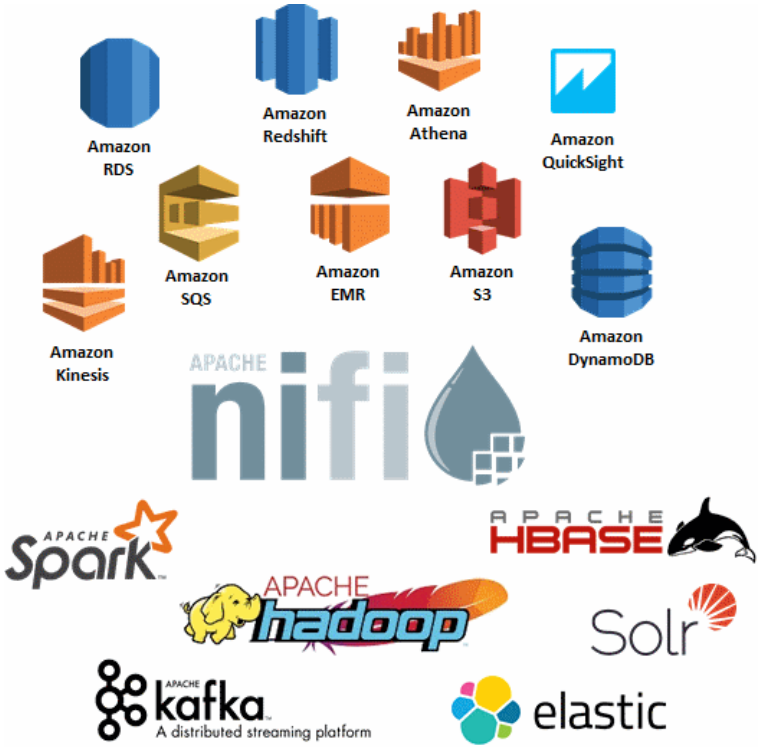
Data Center Infrastructure

Real Estate

CLOUD

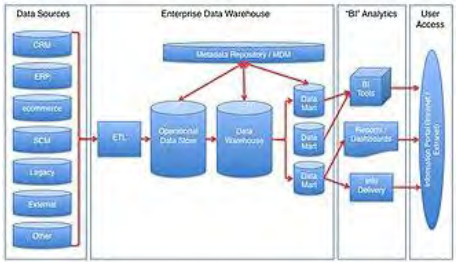
- The shift from on-premise to the Cloud is also significant for data-centric environments
- This started slowly around 2010, but has now accelerated, and Cloud is the norm

The Cloud is Associated with New Technologies

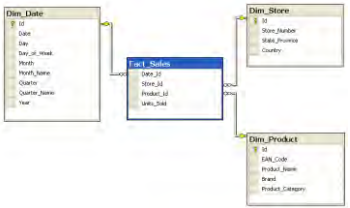


- New technologies have been developed for the Cloud
- These too are having a significant impact on data-centric environments

Legacy Data Warehouses



Architecture



Data Model

ETL

BI / Reporting

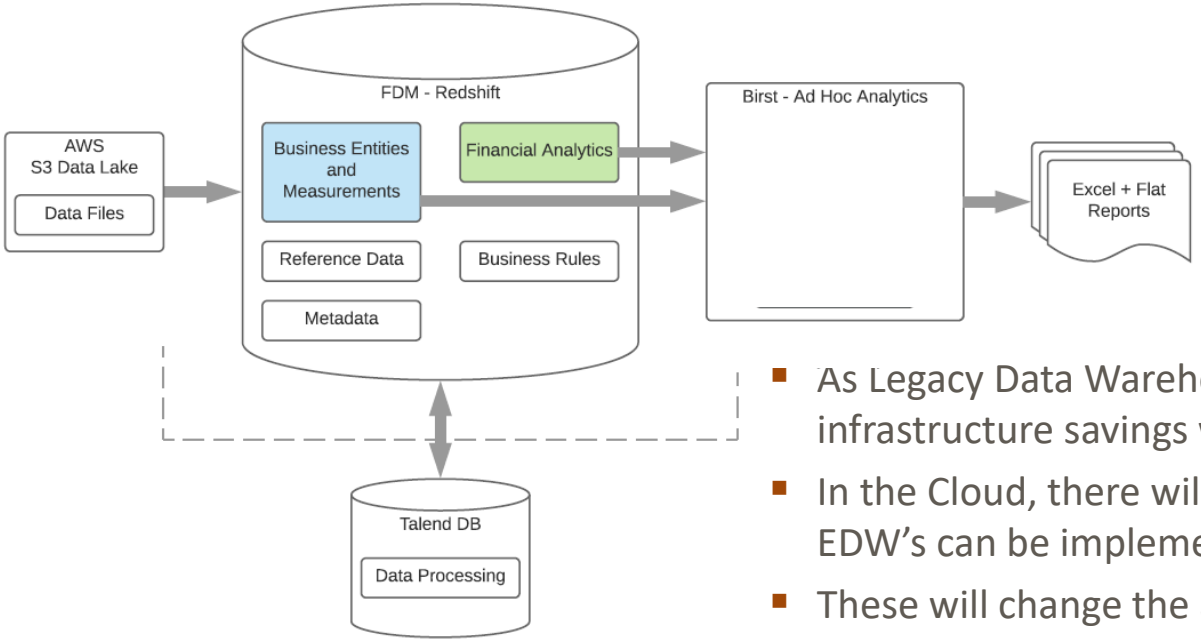
Data Quality

Data Profiling

Tools

- New technologies have been developed for the Cloud
- These too are having a significant impact on data-centric environments

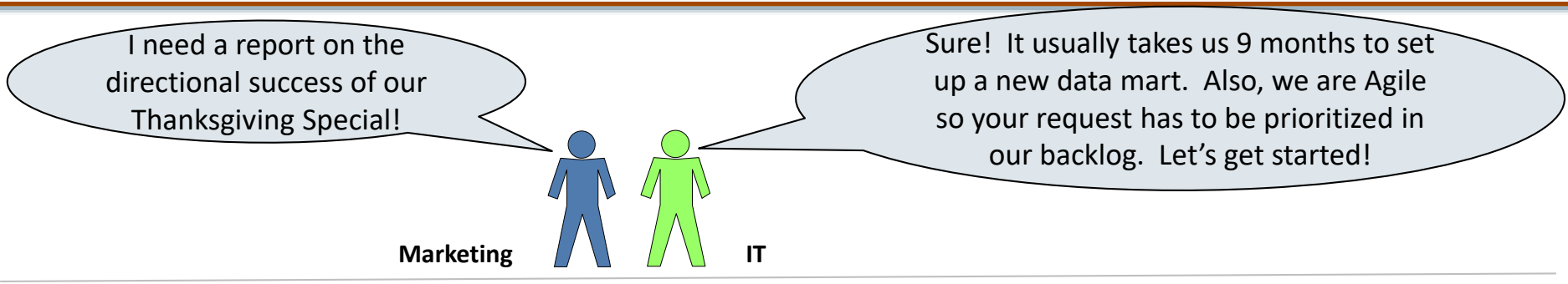
Cloud Data Warehouses

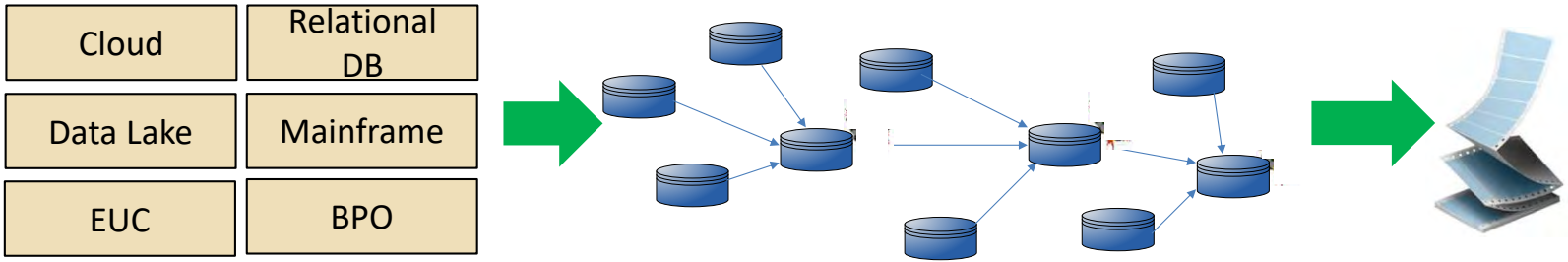


- As Legacy Data Warehouses need to be upgraded, the infrastructure savings will drive enterprise to the Cloud
- In the Cloud, there will be new technologies that the EDW's can be implemented in
- These will change the architecture.
- E.g. Amazon Redshift can provide speed that is faster than Star Schemas in relational databases

Data Virtualization

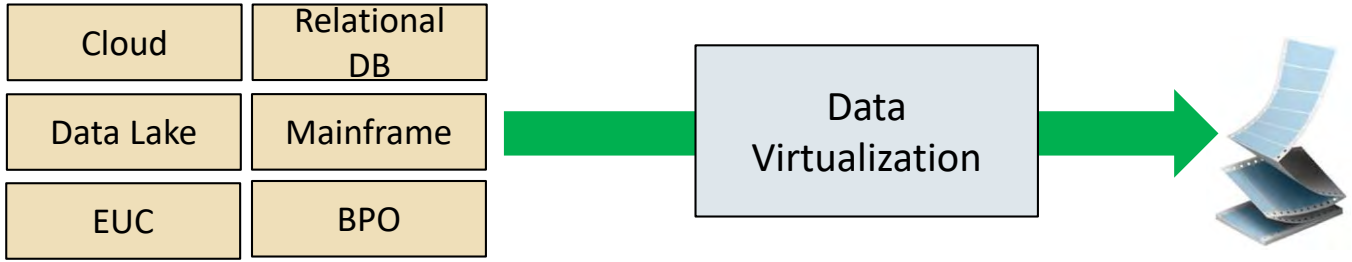
Problems of Time and Complexity





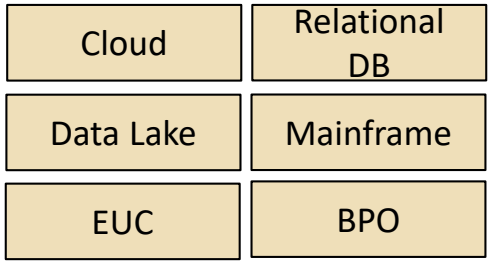
- Two important trends:
  - There are increasing requirements for information that cannot be predicted, are ephemeral, and need to be fulfilled in a short time
  - The accreted architecture of the past 50-60 years makes these requests difficult to fulfil.

Data Virtualization

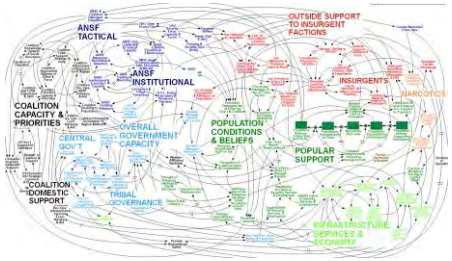


- Data Virtualization provides a way that these two trends can be overcome
- Now, Data Virtualization has been around for a number of years:
  - Initially there were problems in terms of speed of execution
  - These problems are now solved
- What you get is often the same as more costly solutions like Data Integration, ETL, ESB, Database Replication, Data Federation
  - So you get lower cost, more agility, and less need for persistent data stores

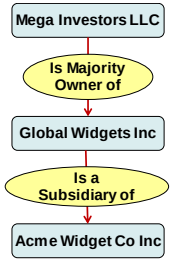
The Challenge for Data Management



Sources



Physical Data Model



What the User Wants

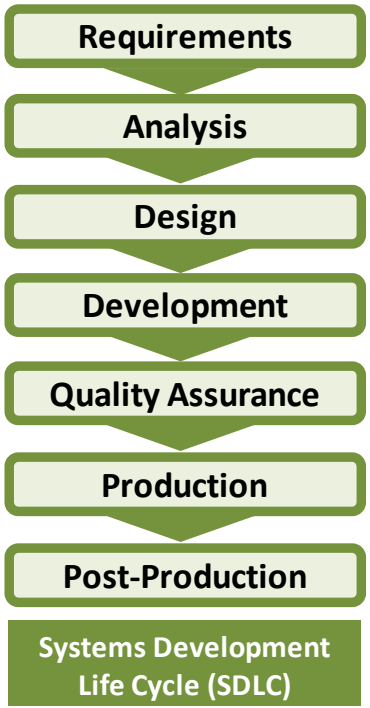
- Now that Data Virtualization is mature, and the technology issues are largely solved, the focus is on the data itself.
- Conceptual models become more important in order to abstract business views of information from the underlying physical data assets.
- Change control is also important as changes in the sources can impact the results sets (structure and content) – but this is truly a more general challenge.

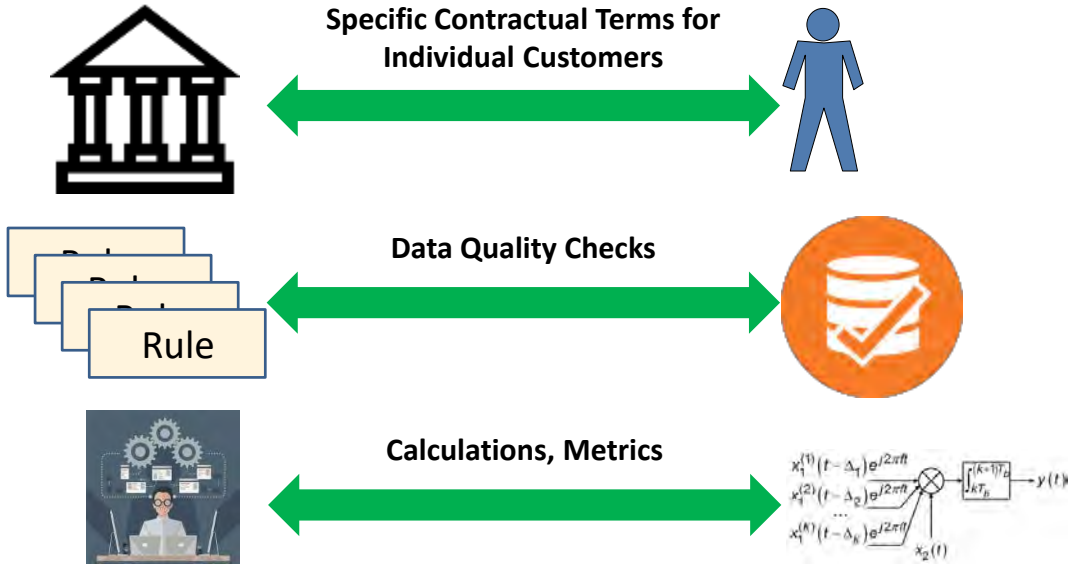
Business Rules Engines

The Significance of Business Rules for Data Goverance

- Business Rules are atomic items of logic that operate on data.
- Common classes of Business Rules include:
  - Calculations: These yield new data elements.
  - Derivations: Similar to calculations, but logic rather than mathematics is used to create new data elements.
  - Data Quality Business Rules: These check for issues of data quality (to the extent such issues can be checked for). These rules do produce a result, but not new data elements
  - Trust and Survivorship Rules: In MDM these determine whether source records (or data elements) make it through into the golden copy. No values are produced by these rules

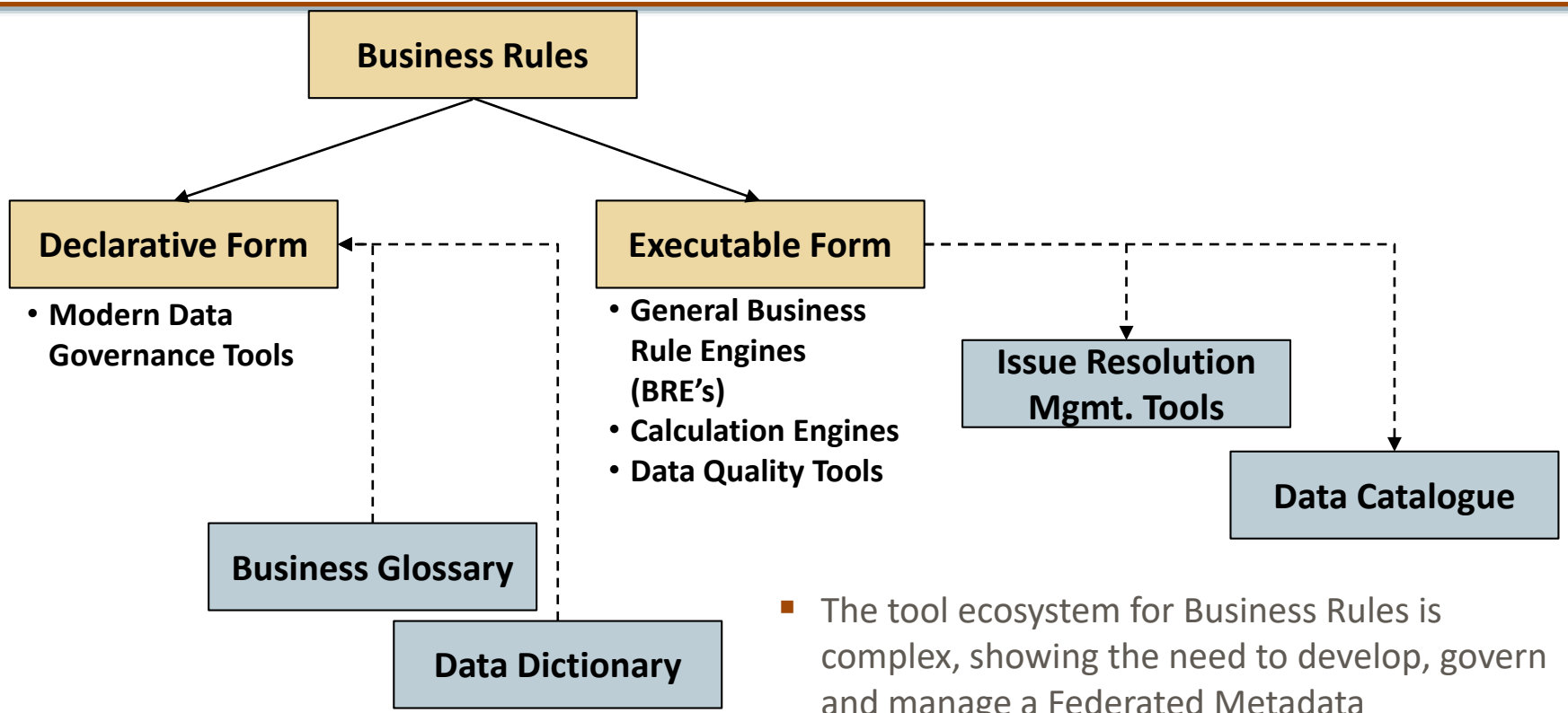
Business Rules: A Way to Get Out of The IT-Driven SDLC





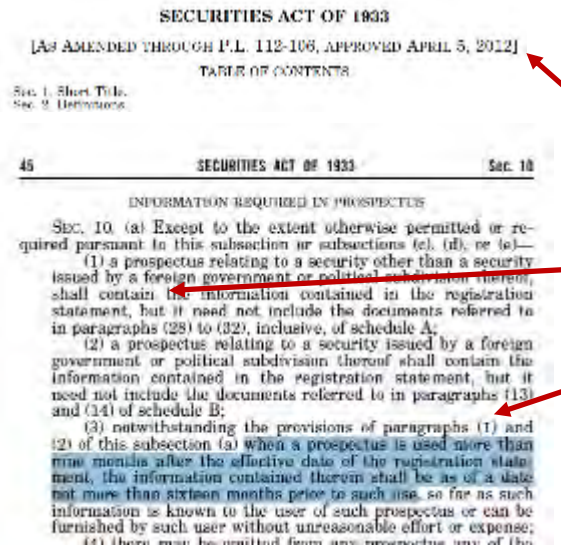
- There are a lot of use cases for the Business Rules approach
- The promise is that the business can manage them and bypass having to get IT involved, where their needs would be deprioritized, require additional funding, be subject to misinterpretation, etc.

The Tool Ecosystem for Business Rules



- The tool ecosystem for Business Rules is complex, showing the need to develop, govern and manage a Federated Metadata Architecture

Challenge for Data Governance: Rule Tracability



SECURITIES ACT OF 1933

[AS AMENDED THROUGH P.L. 112-106, APPROVED APRIL 5, 2012]

TABLE OF CONTENTS

Sec. 1. Short Title.
Sec. 2. Definitions.

45 SECURITIES ACT OF 1933 Sec. 10

INFORMATION REQUIRED IN PROSPECTUS

SEC. 10. (a) Except to the extent otherwise permitted or required pursuant to this subsection or subsections (c), (d), or (e)—

(1) a prospectus relating to a security other than a security issued by a foreign government or political subdivision thereof, shall contain the information contained in the registration statement, but it need not include the documents referred to in paragraphs (28) to (32), inclusive, of schedule A;

(2) a prospectus relating to a security issued by a foreign government or political subdivision thereof shall contain the information contained in the registration statement, but it need not include the documents referred to in paragraphs (13) and (14) of schedule B;

(3) notwithstanding the provisions of paragraphs (1) and (2) of this subsection (a) when a prospectus is used more than nine months after the effective date of the registration statement, the information contained therein shall be as of a date not more than sixteen months prior to such use, so far as such information is known to the user of such prospectus or can be furnished by such user without unreasonable effort or expense;

Authority = SEC

Authority Document

Authority Document Section

Citation

R0001: *Prospectus First Used Date* must be populated with a valid date

Conceptual Rule

R0001IPS: Ensure field "Prospectus First Used Date" is not null in the Security Master table. If it is, raise Error Code ER1734

Logical Rule

R0001IPS01:

```
INSERT INTO ERR_TAB (ERR_NUM, ERR_CT)
VALUES ("ER1734", (SELECT COUNT(*)
FROM SEC_MSTR WHERE PRSPT_FUSED_DT IS NULL))
```

Physical Rule

- Being able to track where Business Rules come from has never been adequately addressed from an industry-wide viewpoint.
- This represents a significant Data Governance challenge

Robotic Process Automation (RPA) and Business Rules

FOR ENTERPRISE ARCHITECTURE PROFESSIONALS February 13, 2017

The Forrester Wave™: Robotic Process Automation, Q1 2017

The 12 Providers That Matter Most And How They Stack Up

FIGURE 2 Forrester Wave™: Robotic Process Automation, Q1 '17

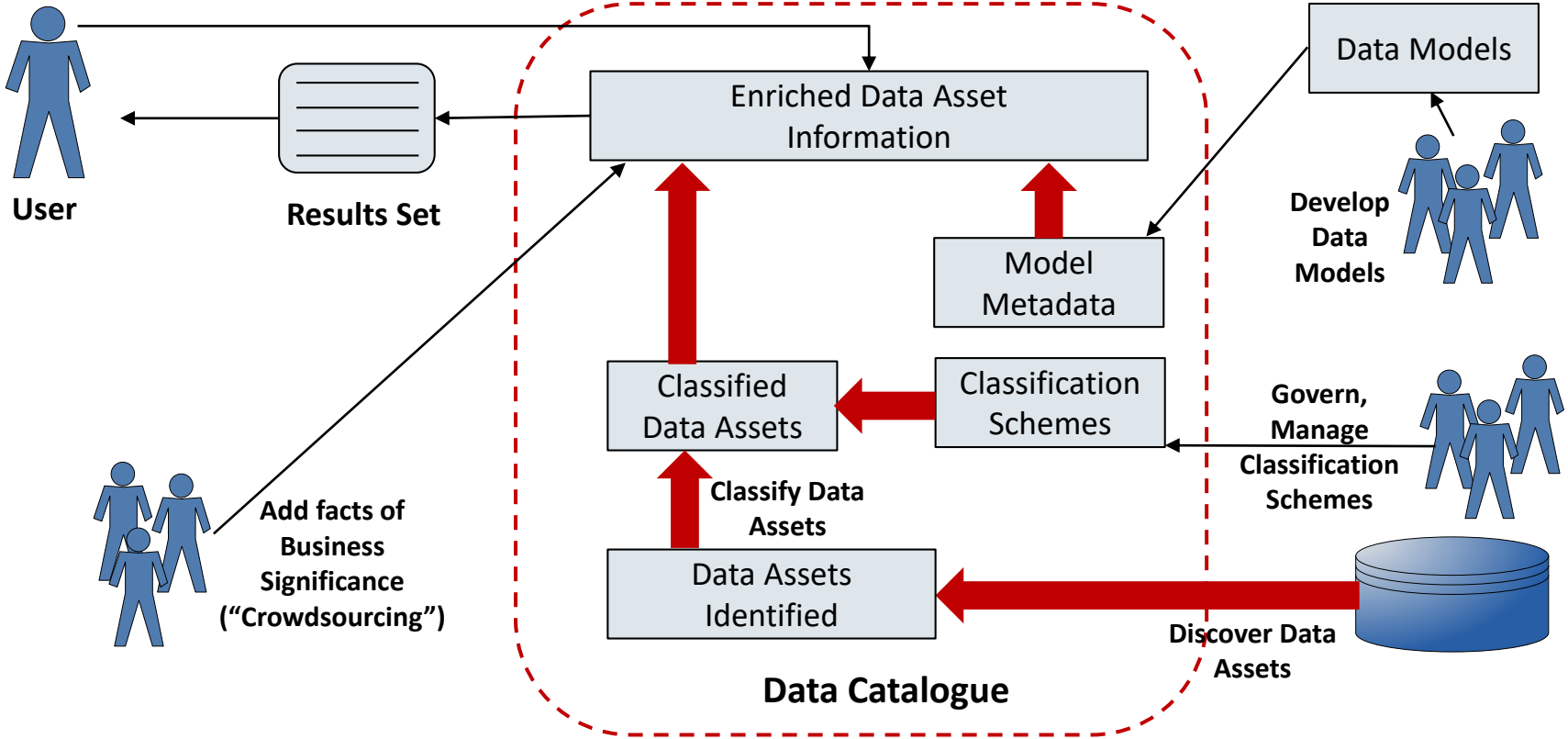
- RPA is maybe the most significant trend of all this year.
- It is the automation of many data management (and other) tasks; anywhere a UI is exposed, it can work.
- The bots have to be programmed – they work off Business Rules. This is extending the pressure on Data Governance.

Self-Service Analytics

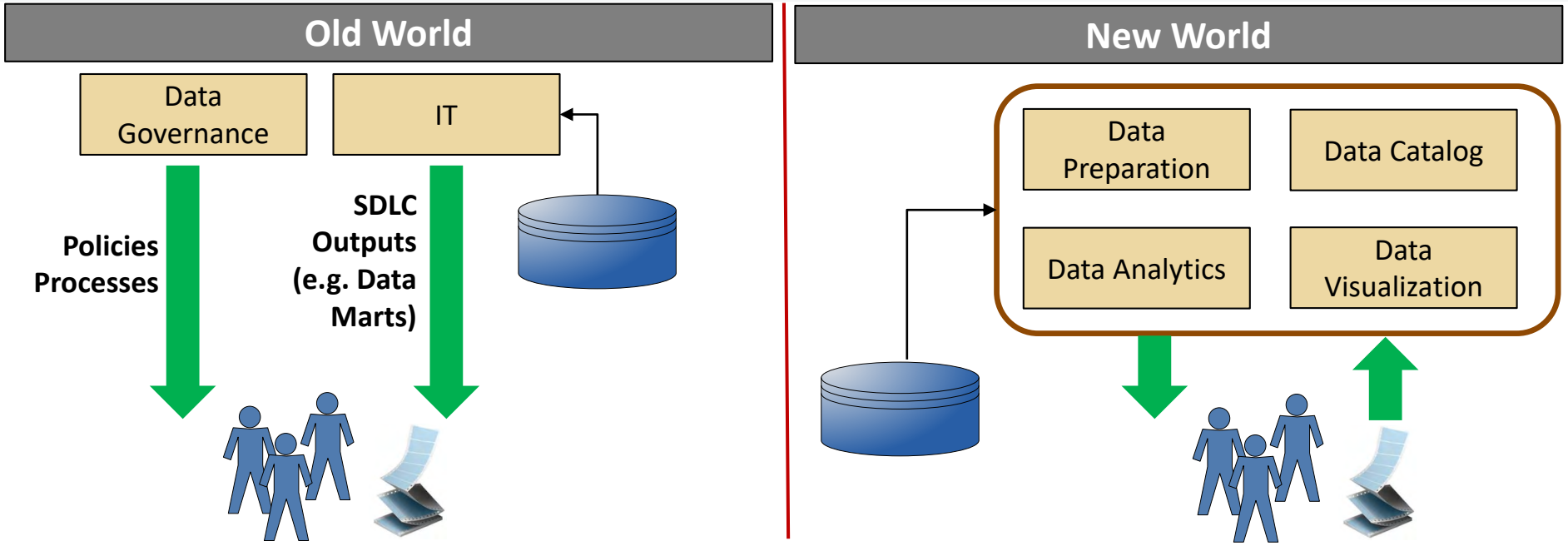
The Core Requirements of Self-Service Analytics

- If Self-Service Analytics is to be a reality, then users must know:
  - Where the data is
  - What the data means
  - What coverage the data has
  - What issues there are with the data
  - How to get the data
  - Whether it is permitted to use the data for the proposed purpose
- This information will ideally be in one location that is business-facing and easy to navigate
- Today, this one location is identified as the Data Catalog
- Once all this is known, the data wrangling tools, the analytic tools, the data visualization tools, etc. can be used to satisfy the request

How Should a Data Catalog Be Populated?

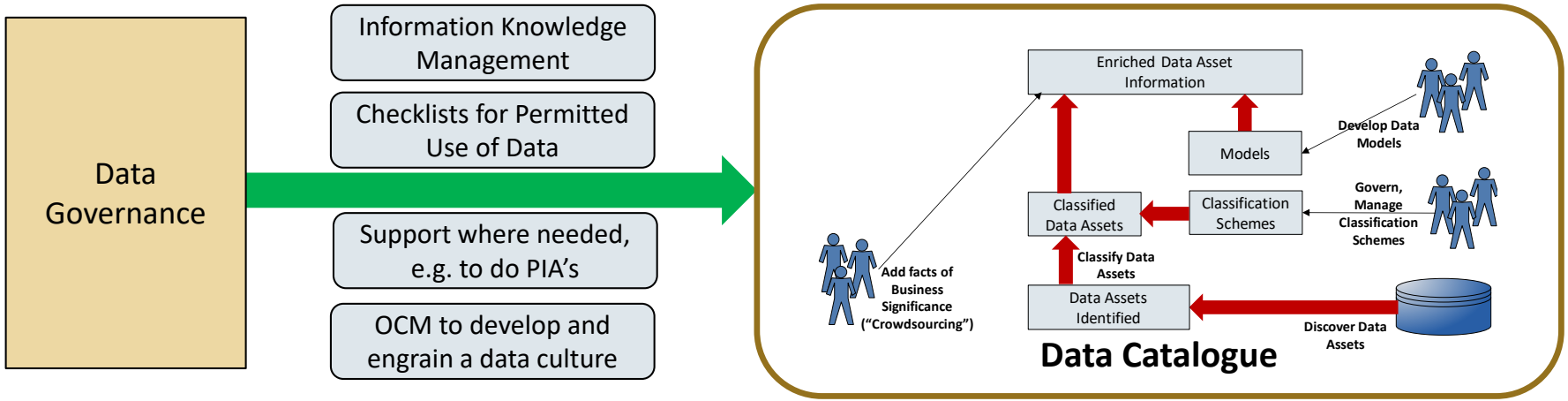


Challenge for Data Governance



- The old way of doing Data Governance does not match the empowerment provided by the new data technologies, and this creates a challenge for Data Governance

Meeting The Challenge of the Data Catalog



- Data Governance is going to have to shift from managing operational risk in data and data management, to supporting the extraction of value from the data resource
- The Data Catalog is the focal point via which Data Governance can carry out its mission and communicate meaningfully with data citizens

Thank you!

Malcolm Chisholm Ph.D.
malcolm@firstsanfranciscopartners.com
732-687-9283