IBM InfoSphere Information Server Update & Roadmap

Data Server Day
Stockholm, May 2017

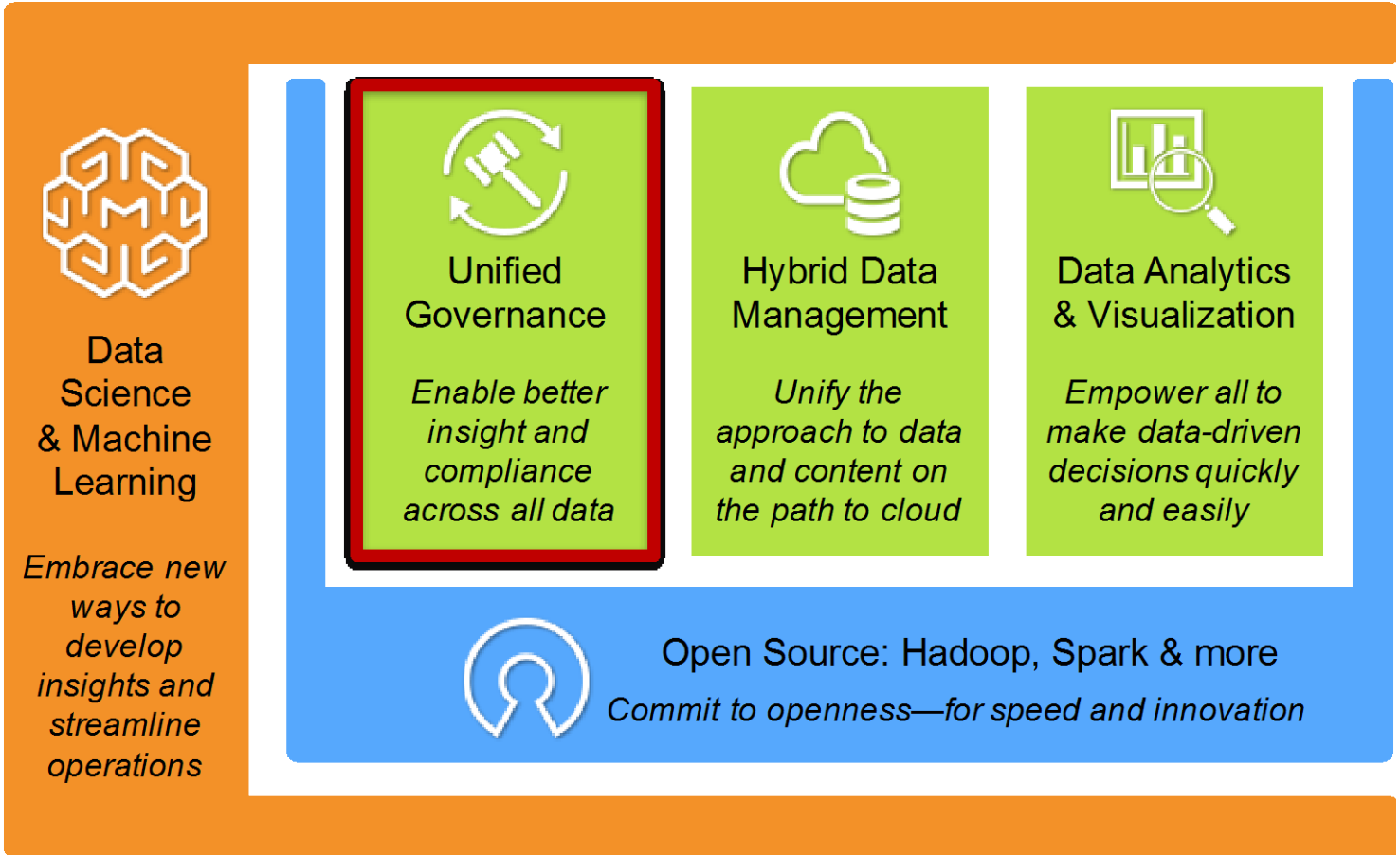
Beate Porst – porst@us.ibm.com
Program Manager – Product Management
IBM InfoSphere Information Server

Disclaimer :

- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.
- Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.
- The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.
- The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

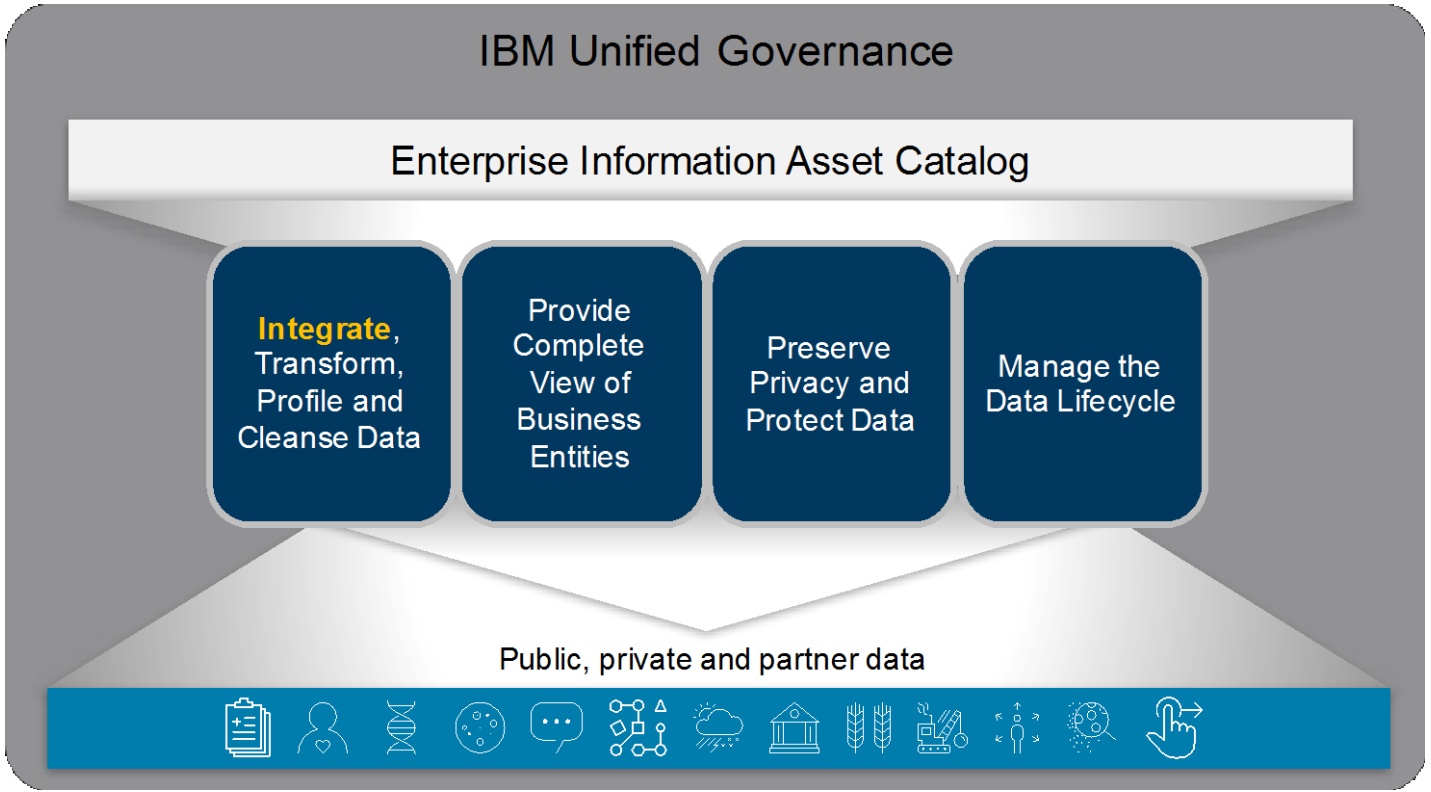
Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Our Portfolio: Power behind and across the portfolio



Unified Governance

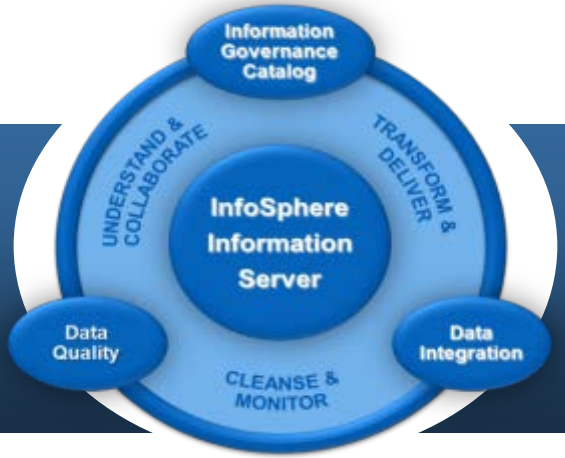
Enable better insight and compliance across all data



Unified Governance…

… powered by Information Server

Integrating and transforming data and content to deliver accurate, consistent, timely and complete information on a single platform unified by a common metadata layer



Information Governance

Understand & Collaborate

- Catalog technical metadata & align w/ business language
- Mange (big) data lineage
- New compliance reporting

Data Quality

Cleanse & Monitor

- Analyze & validate w/ enhanced classification
- Cleanse & standardize
- Define, manage & monitor data rules + exceptions

Data Integration

Transform & Deliver

- Massive scalability
- Power for any complexity
- Deliver in batch and/or real-time with change capture

common connectivity • shared metadata • security • data privacy
common execution engine with flexible deployments → Hadoop • Grid • Cluster • Cloud

Information Integration is Core for many use cases – Traditional and New

Deliver Trusted Information for Data Warehousing and Analytics

Integration & Governance for Data Lakes

Self-service data to sandbox for ad-hoc analytics

Consolidate Applications

Secure enterprise data and ensure compliance

Build and manage a single view

Make Enterprise Applications more efficient

IBM InfoSphere Information Server
Continuous Feature Release & Stack Expansion

GA Information Server **v11.5**

GA Information Server **v11.5.0.1**

New User Experience for the **Business Analyst**

NEW

GA Information Server **v11.5.0.2**

New IA Thin Client Launch User Experience for the **Data Analyst**

New **IBM Information Server on Cloud** offerings

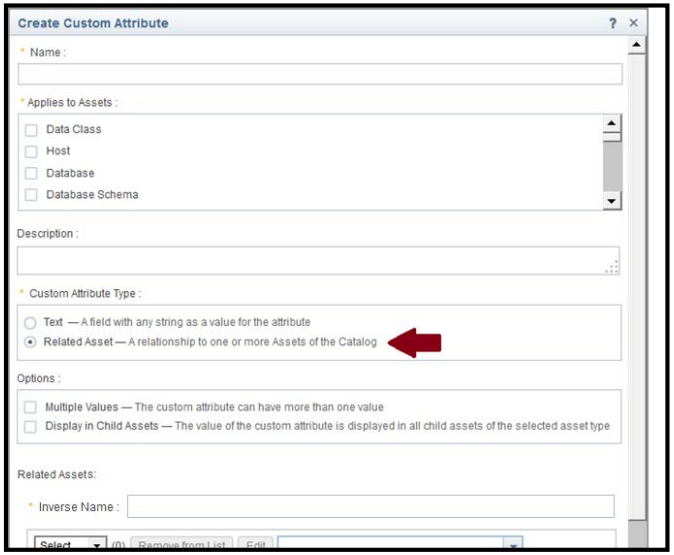
New SAP ERP & BW Packs

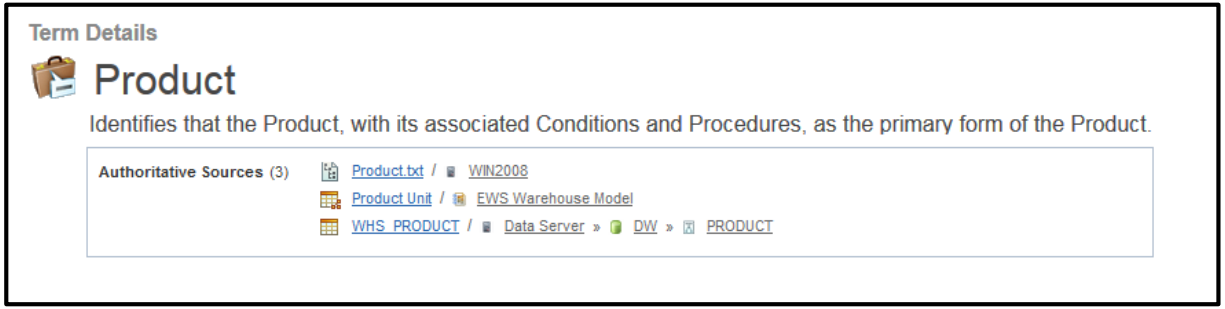
Other Gov/DQ/DI feature patche

IGC – Custom Attribute of Type Relationship

- Define and manage relationships between any two Assets within Information Governance Catalog, such as Business Owners or Asset Dependency.

- Support the ability to extend the Catalog and document impact or referential relationships.





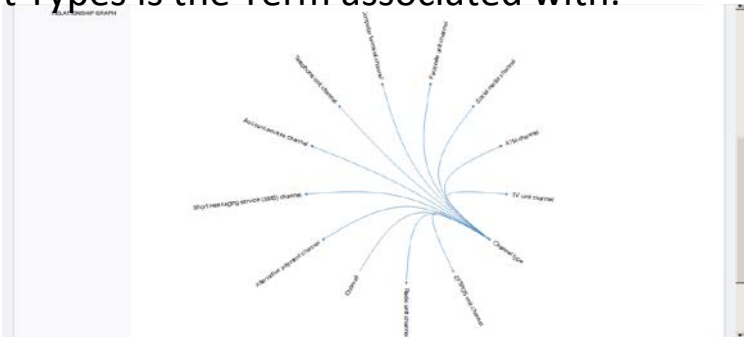
IGC – Graphical Visualization

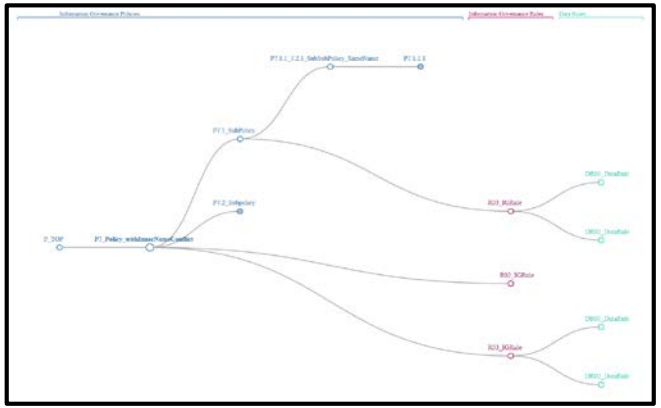
Create and manage graphical relationship reports, such as the visualization of the discovered or implemented Table or Entity relationship, or traverse the relationships between Governance Policies, Governance Rules, Data Rules and Data Sets.

Explore and discover the Neighborhood Graph of any Asset, discover its dependencies and usages.

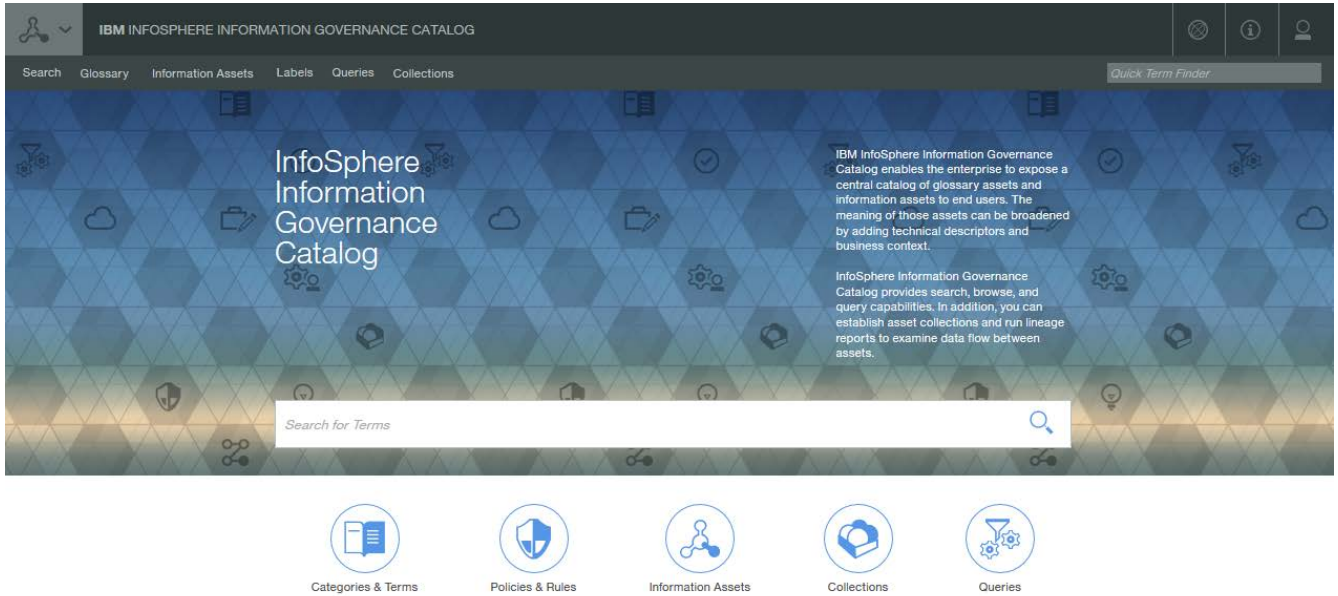
Explore and discover relationships of a Glossary Term

- What are the components of a Term.
- What Types is the Term associated with.





IGC – Business User focused User Experience

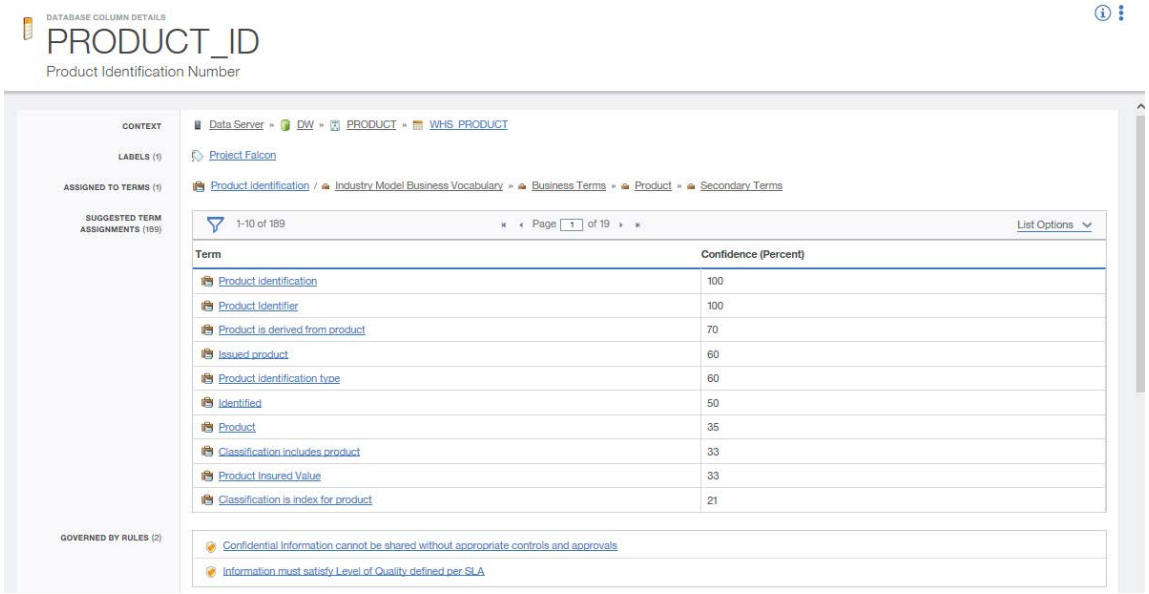


Simplifies and re-define the user experience, while continuing to deliver the ability to define, capture and report upon Data Governance initiative and requirements.

IGC – Suggested Term Assignment

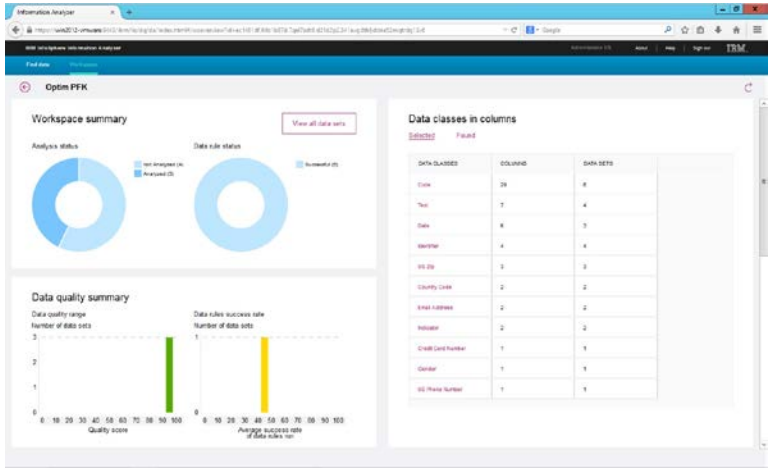
Greater benefit when searching and exploring information within IGC, assets should be aligned with a Glossary Term that can clearly express their definition and requirement.

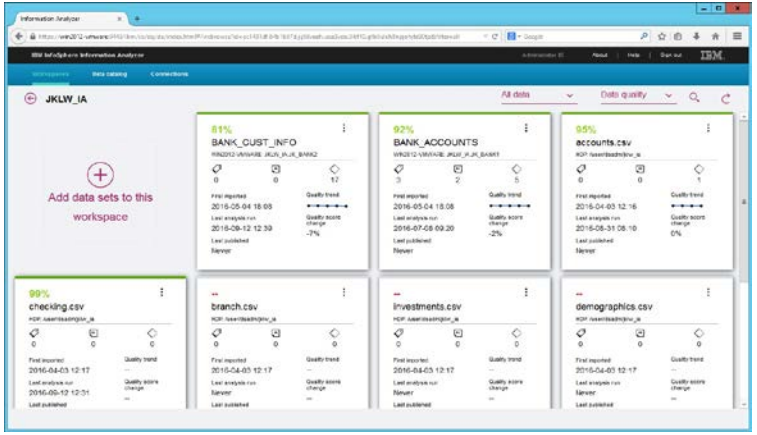
To aid, we introduce a capability to scan a Database or set of Data Files and suggest the Term assignment.



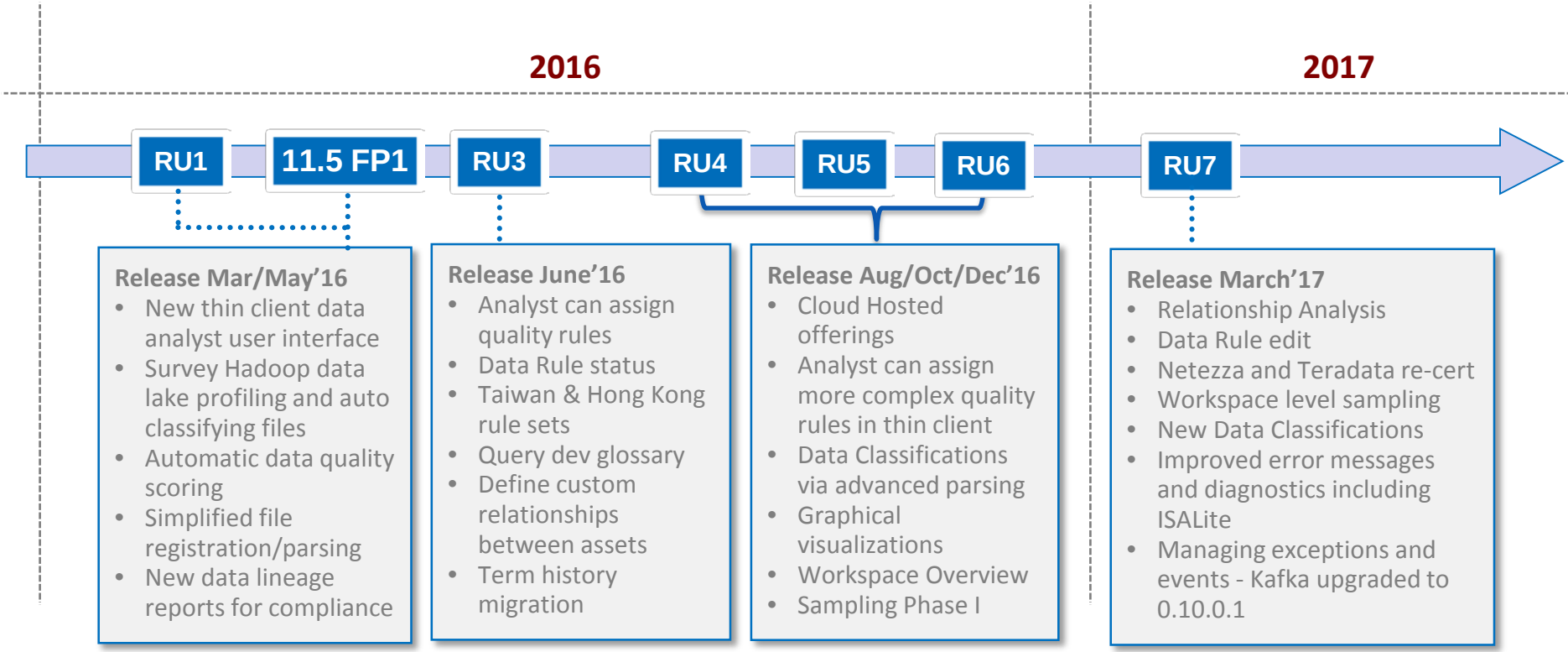
Information Analyzer – Thin Client

- Lightweight, browser-based UI requires no client-side patches
- Streamlined interface is well suited for Business users and Analysts
- Companion to the Information Analyzer Workbench
  - Completing migration in 2017
- Data Quality scores calculated for all columns
- Can browse and analyze delimited files on both Hadoop HDFS and local (engine) file systems
- Data analysts can run, view and edit analysis results for data sets and view data quality scores for files, tables and columns

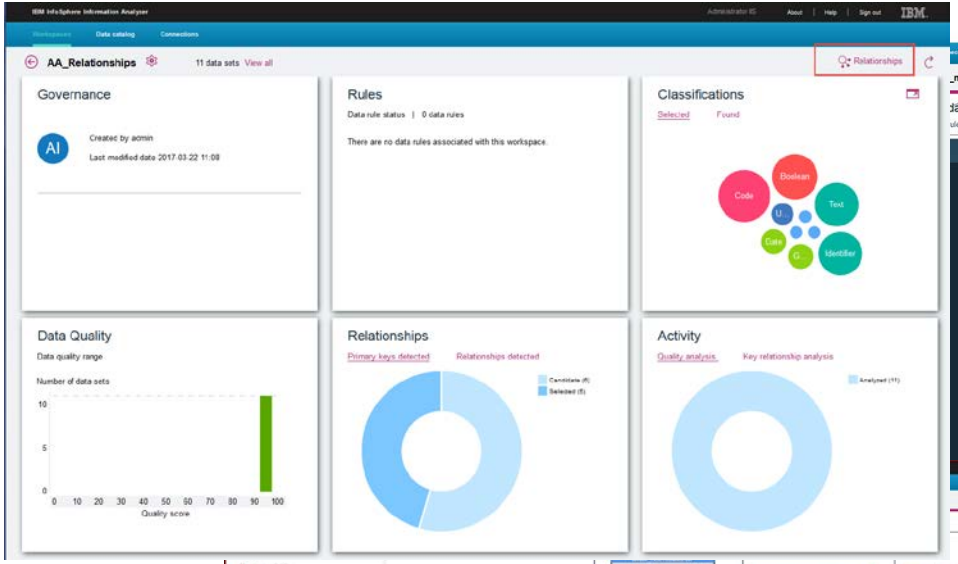


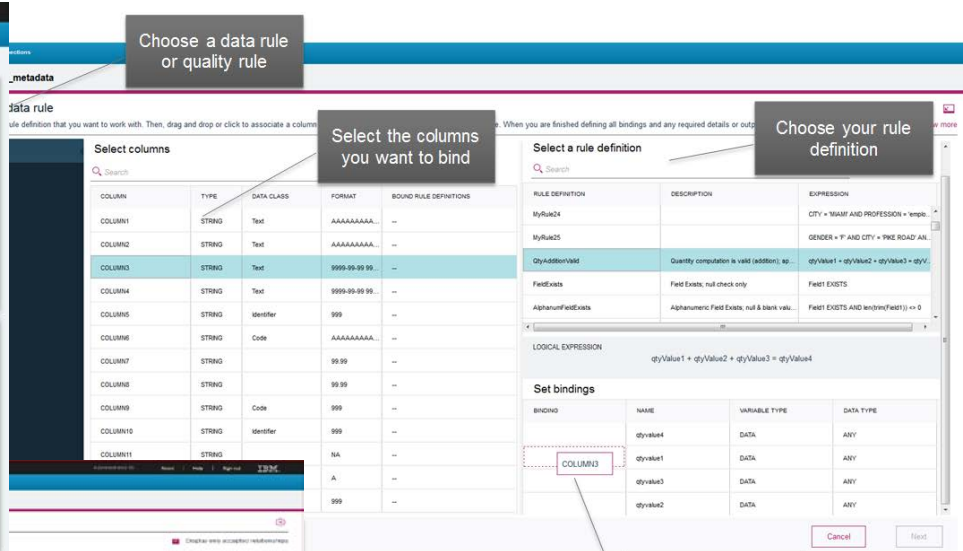


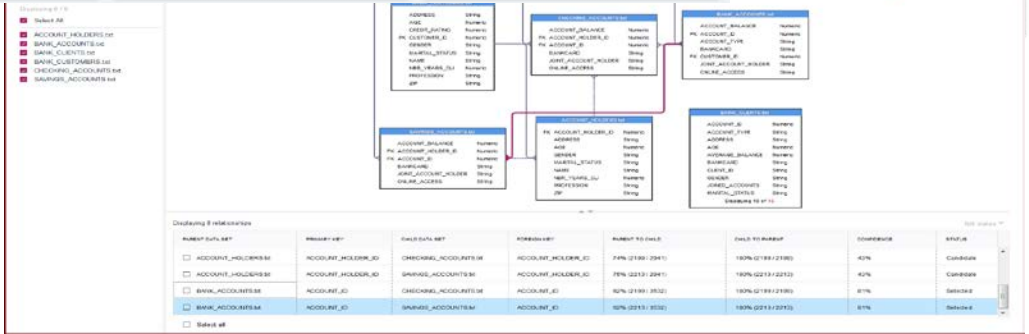
IA – Recent Product Updates



IA – Relationship Analysis & Data Rule Assignment





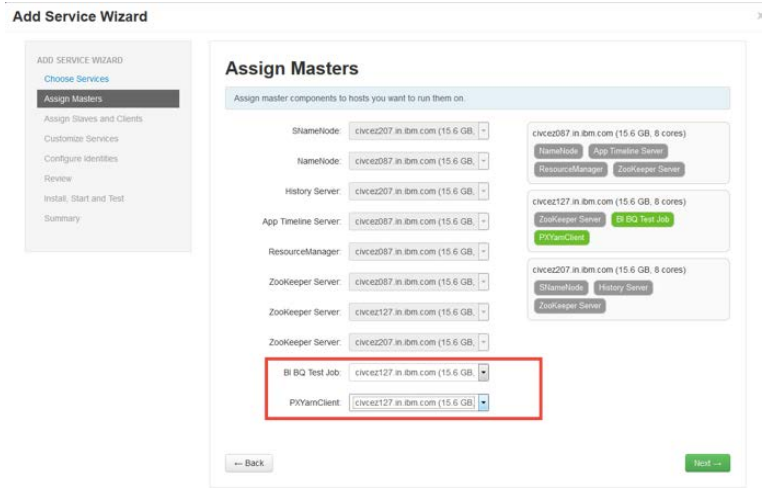


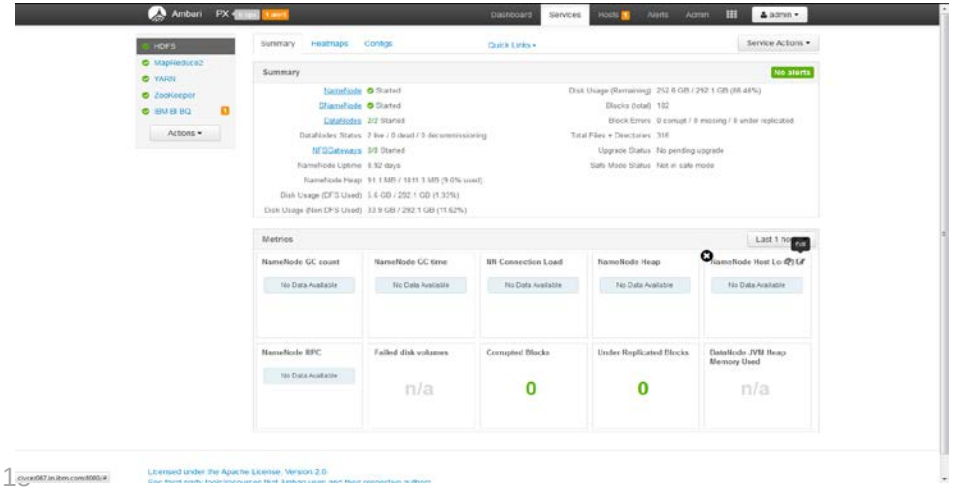
Connectivity – New, Expanded or Enhanced

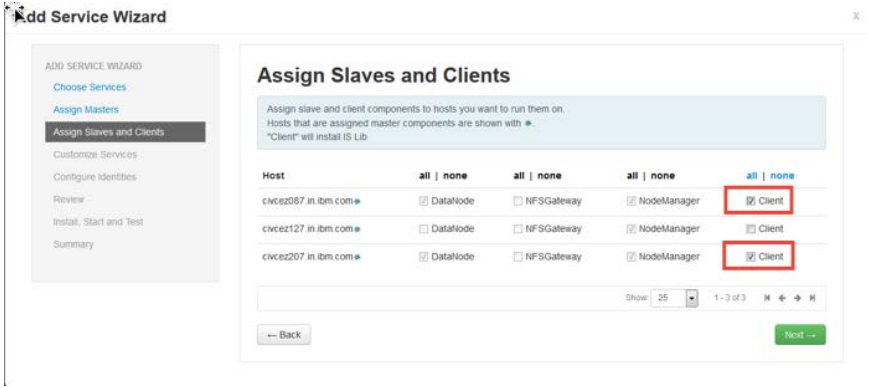
14

Ambari Integration

- Configure, deploy and manage Information Server nodes from Ambari Management Console
- Start/Stop PX Yarn Client from Ambari
- Run IS Verification Service







2016/2017: Information Server *Data Integration* Enhancements

	H1/2016	H2/2016	Q1 / 2017

Connectivity	• MQ v8 certification
	• DB2 LUW v11 certification
	• New **Hive** Connector
	• File Connector enhancements:
	  • **AVRO, ORC, Parquet** File type
	  • HDFS File browsing
	  • Data Locality support
		• SQL Server 2016 certification
		• SFDC Connector API 36 support
		• New **Kafka** Connector
		• **S4/Hana** support
		• **ORC** Parallel read
		• **Netezza** transient table support
		• **Netezza** v7.2 database certification
		• **SQL Server** PDW certification
			• **dashDB** certification
			• **Teradata**
			  • v16 certification
			  • performance enhancements
			• **Cassandra** certification
			• **Salesforce** bulk extract
			• **DB2 Connector** enhancement
			  • Re-establish connection for z/OS load

Platform	• **Ambari** support for Information Server for Hadoop • **DataStage on Cloud** • AIX performance enhancement for large work loads • File Connector lineage support between HDFS File & Hive table • Win 10 support • IE 11 • HDP 2.3, CDH 5.5/5.6, MapR 5	• Aggregate performance improvement • HDP 2.4 certification • **IS Enterprise Edition on Cloud** • **SAML** 2.0 support • BigInsights 4.2 certification • Chrome Browser certification • **New SAP Packs** (ERP & BW) • ***New offerings:*** *Data Integration for Hadoop and Enterprise*	• **CDH 5.8 / 5.9** Certification • **SAML / WebSeal** support • **Rapid Job Update Tool** for v11.5 • **YARN / Ambari** support phase 2 • **BigIntegrate 4.3** Certification

2016: Information Server *Data Quality & Governance* Enhancements

	Q1	Q2	Q3	Q4
Data Quality	• **New thin client Data Analyst user interface** • Survey Hadoop data lake profiling and auto classifying files • Automatic data quality scoring • Simplified file registration / parsing	• Analyst can assign quality rules • Data Rule status • Taiwan & Hong Kong rule sets	• **Data Quality Cloud Hosted offerings** • Analyst can assign more complex quality rules in thin client • Data Classifications via advanced parsing • Historical lineage & DQ reporting	• New Data Quality workspace and data rule experience → (aks IA think client) • New data sampling algorithm
Governance	• **New Compliance Reporting** • Annotate Asset Definitions • Data Quality Scoring & Metric support	• Query Development Glossary Assets • Custom Attributes of type Relationship • Term history migration	• **Governance Cloud Hosted offerings** • **Graphical relationship Visualization** • Additional Visualization Reports for Terms and Entities	• Governance made easy: ***New Business User* focused interface** • Business Lineage improvements • Historical Lineage Reporting support

IIG Strategy / Roadmap 2017 & Beyond

IBM Information Integration & Governance

… the road ahead – *Primary Objectives for 2017 & beyond*

Modernization

- **Re-invent user experiences**
- Strengthen support for **Data Lakes**
- **Modernize** the **foundation**
- Implicit **Governance & Quality** monitoring/enforcement

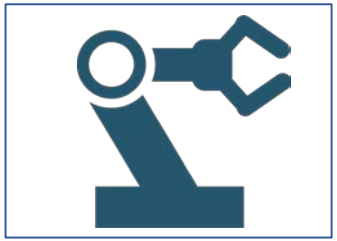
Hybrid

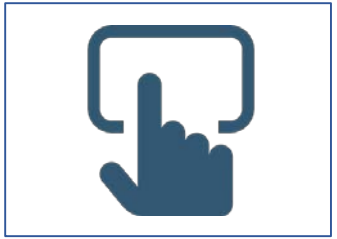
- Build a best in class ***true Hybrid Integration Platform***
- Support **Data Gravity and deployment flexibility**
  - System or user controlled runtime optimization
  - Hybrid elasticity

Unified Governance

- Support **Governance 2.0**
- Governance for **Insights** and **Compliance** for anyone and anything

Simplify, Accelerate & Automate Integration





Modernizing Information Server **Foundation** with build in automation

Self-service cognitive Integration & Quality Design

Redesigning DataStage **User Experiences**

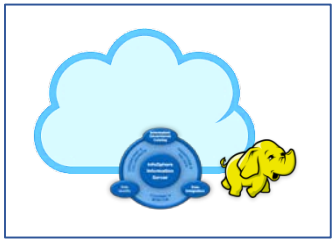
Best in class support for **Data Lakes**

- Connectivity
- Data Security
- Ease of Management & Deployment

Accelerate and simplify the time to value with **Download & Go**

Self-service integration & preparation with implicit quality & governance control & **Shop for Data**

Improving your Time to Value, Management and Productivity of your Data Lakes



Continues Connectivity expansions and enhancement	Tight integrate into Hadoop Cluster Management	Peace of Mind: Enhanced Data Masking functions	Simplifying Maintenance and HA for Hadoop and stand-alone Grid deployments	Continues certification of supported Hadoop distribution

21

Business Agility demands Expansion of Integration User Community

Data Preparation & Curation

WHO:

- Business users / Data owners non-technical users

WHAT:

- Visual data shaping/curation → WYSIWG
- Closely guided and controlled (shop for data paradigm)
- Manipulation of 1-2 data sets at a time

Self-Service Integration

WHO:

- Shadow IT, LOB users, Data Scientist, semi-technical

WHAT:

- Combined visual & flow based design
- Template / pattern approach
- Zero configuration
- Implicit validation
- Collaboration

Enterprise-class Integration

WHO:

- Integration Specialist, Integration Developer, highly technical

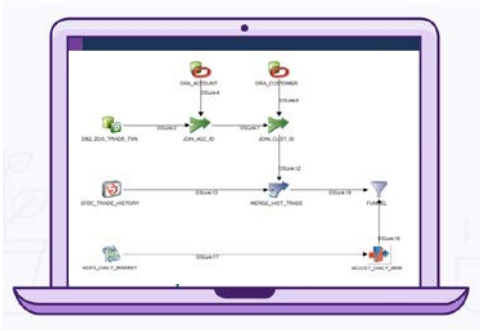
WHAT:

- Comprehensive library of integration, transformation & quality operations
- Support for comprehensive integration flows and projects
- Expandability for custom operations
- Full control for configuration & parameterization
- Top-down or bottom-up design approach
- Support for team development process

Cognitive Integration Design

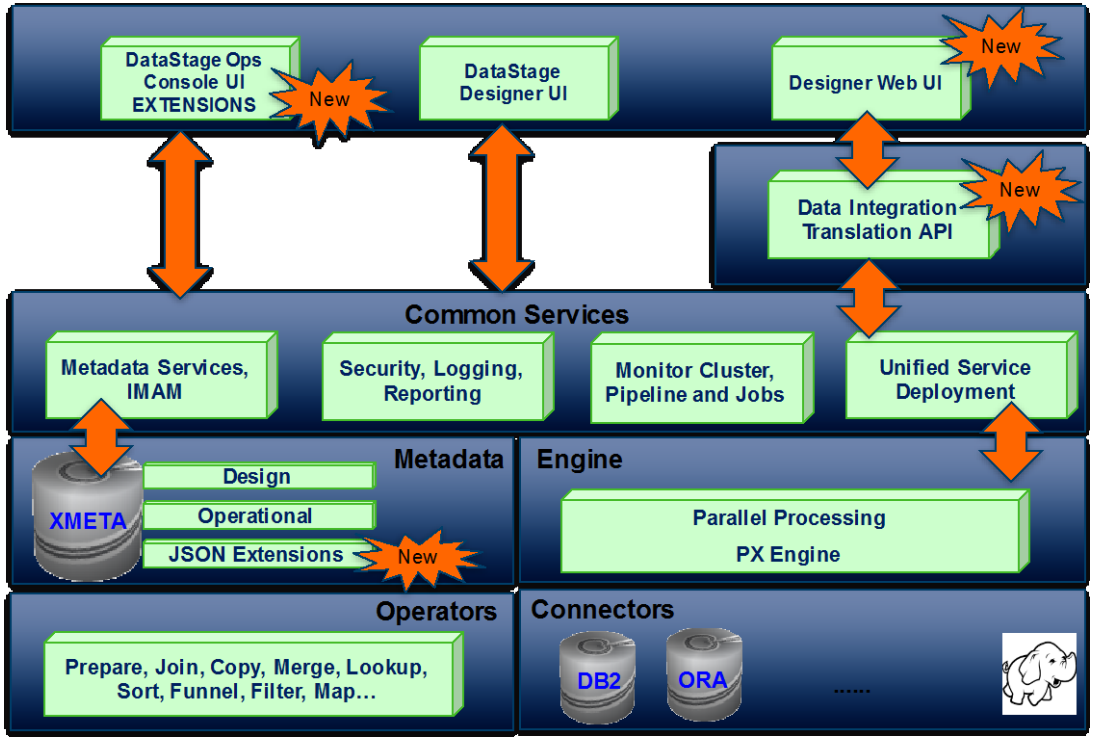
Next Gen DataStage Designer

- *What:*
  - *ZERO* Install → Browser based design
  - *ZERO* migration → view existing jobs/projects in new designer
  - Ability to use new & old Designer side by side
  - New simplified *Design experience* without compromising capabilities



- *Who:*
  - **Phase 1:** New Integration experience for *Integration Specialist / Integration Developers*
  - **Phase 2/3:** *Self service* integration and preparation for Business and LOB users

Enhanced Architecture to support new UI



What do we mean by Hybrid Integration?

Choice of Runtimes based on your data delivery requirements	Distributing workloads across loosely coupled runtimes	Combined SaaS and on-premise self service prep / integration	Optimizing workloads based on Data Locality	On-demand / elastic expansion

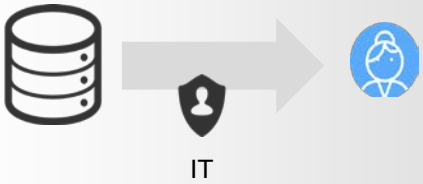
26

Unified Governance

A New Era of Governance

Governance 1.0

- Data within the firewall
- Distinct capabilities for structured & unstructured data
- Compliance use cases: e-Discovery, Records, Archiving, GDPR, BCBS 239, Basel II etc.
- IT led



Governance 2.0

- Data, API's, & Analytics in or outside the firewall (Hybrid platform)
- Common capabilities: Policy Administration, Metadata, Consent Management, & Stewardship
- Compliance & analytics use cases: Information Repositories (e.g. Data Lakes), Self-service analytics, Regulations, & Data Science, GDPR, BCBS 239, Basel II etc.
- IT & Business led

Use Cases Driving a Unified Governance Strategy

GOVERNANCE FOR COMPLIANCE

Discover, classify and manage information in ways that meet the obligations enforced by both regulatory and corporate mandates

Regulations (e.g. GDPR)

Privacy & Protection

eDiscovery

Records & Retention

Archiving

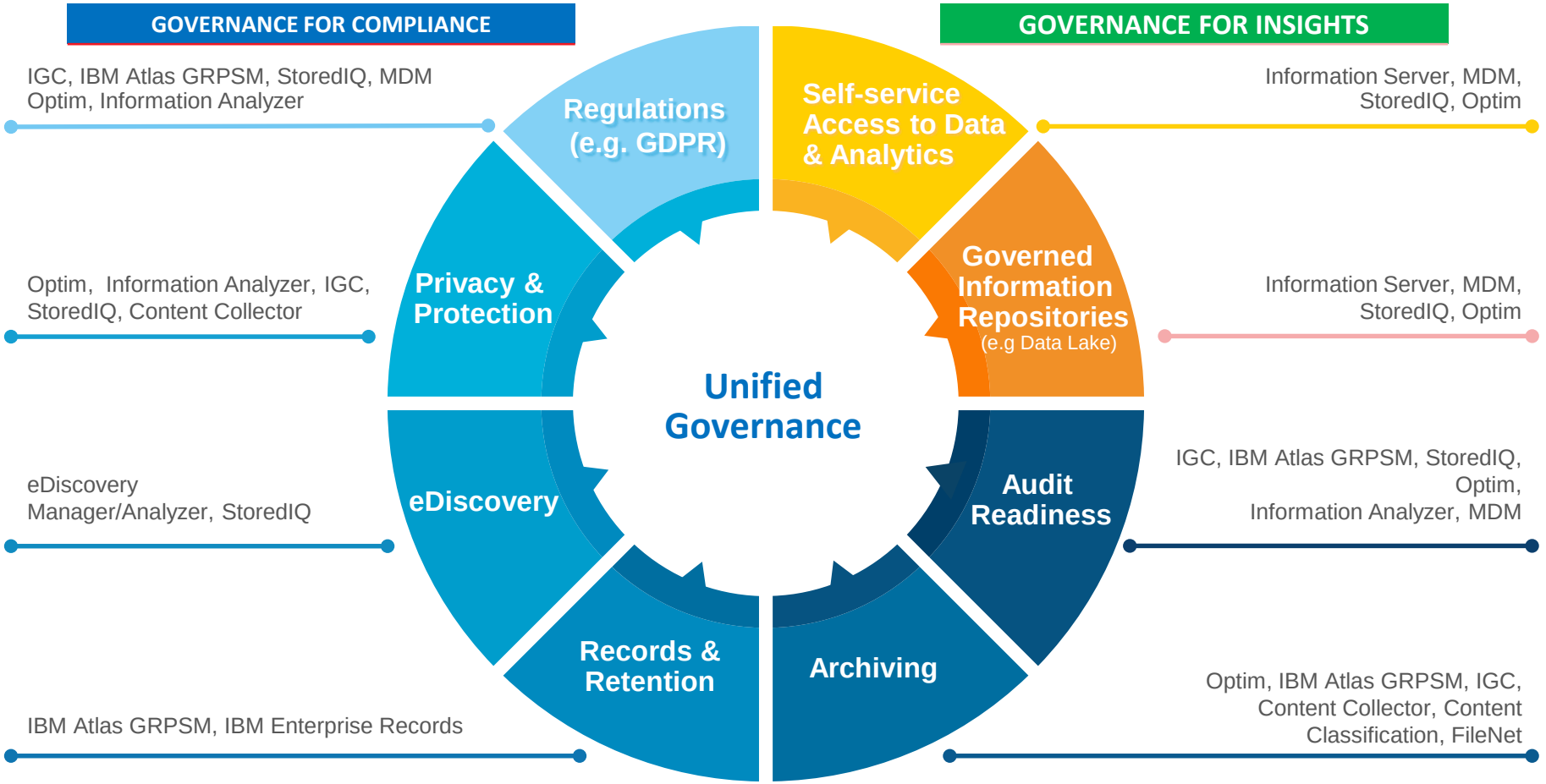
Audit Readiness

GOVERNANCE FOR INSIGHTS

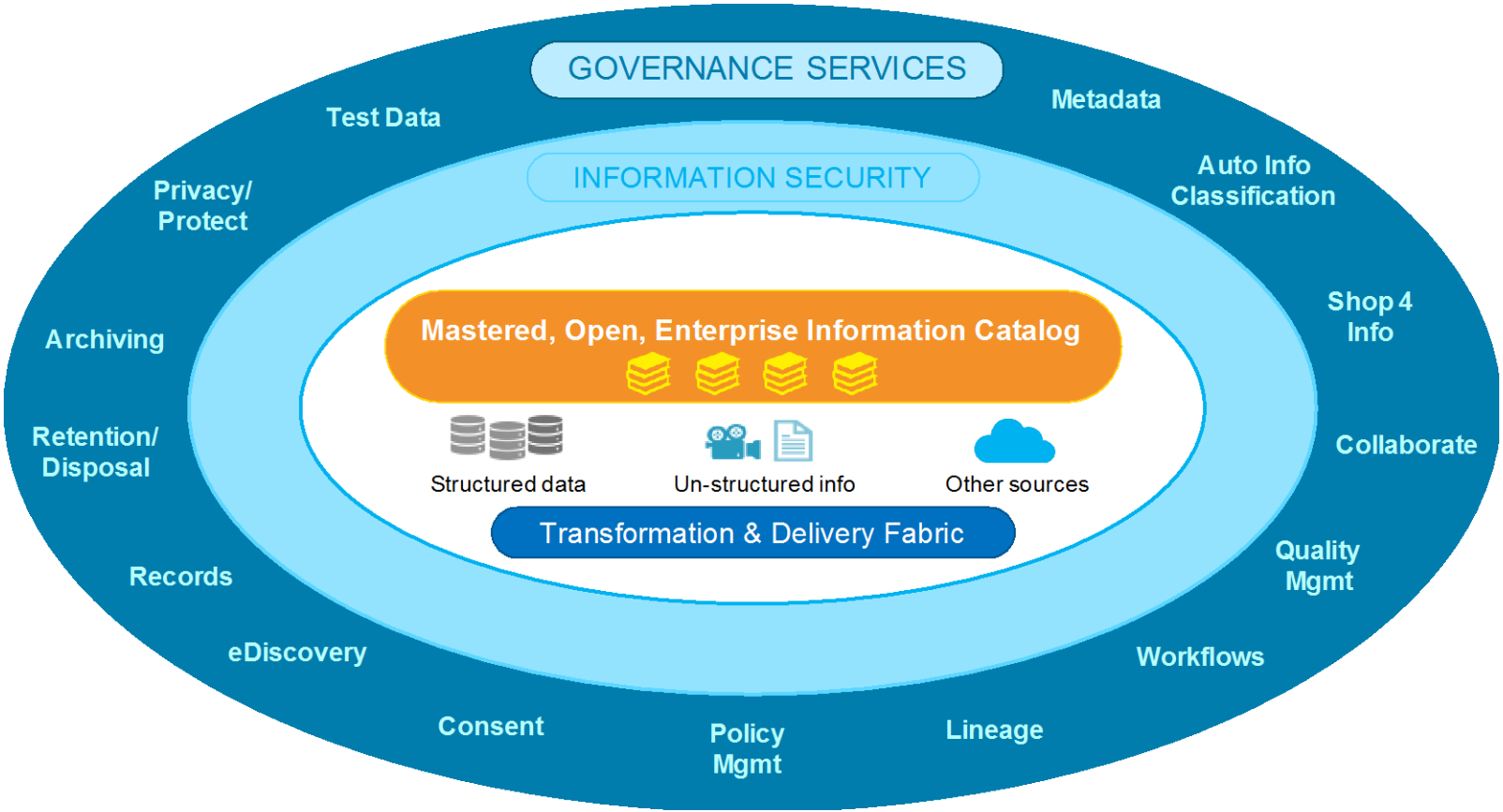
Provide safe access to trusted, high quality, fit-for-purpose data while facilitating effective collaboration among team members

Self-Service Access to Data & Analytics

Governed Enterprise Information Repositories (such as Data Lakes)

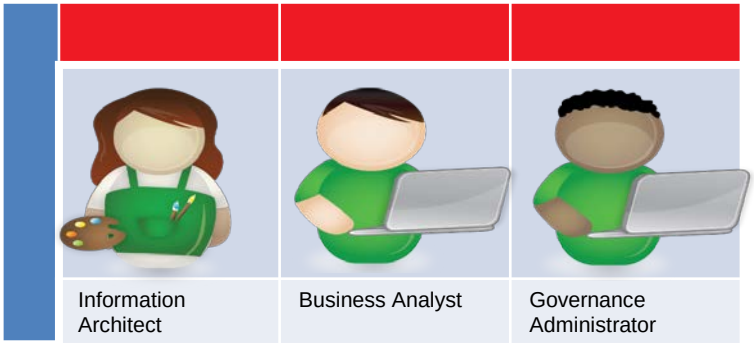

GOVERNANCE FOR COMPLIANCE
GOVERNANCE FOR INSIGHTS
IGC, IBM Atlas GRPSM, StoredIQ, MDM Optim, Information Analyzer
Information Server, MDM, StoredIQ, Optim
Optim, Information Analyzer, IGC, StoredIQ, Content Collector
Information Server, MDM, StoredIQ, Optim
eDiscovery Manager/Analyzer, StoredIQ
IGC, IBM Atlas GRPSM, StoredIQ, Optim, Information Analyzer, MDM
IBM Atlas GRPSM, IBM Enterprise Records
Optim, IBM Atlas GRPSM, IGC, Content Collector, Content Classification, FileNet
Regulations (e.g. GDPR)
Self-service Access to Data & Analytics
Privacy & Protection
Governed Information Repositories (e.g Data Lake)
Unified Governance
eDiscovery
Audit Readiness
Records & Retention
Archiving

Unified Catalog – The Core of the Unified Governance



User Experience

Support varied users (Business Analyst, Developer, Author, Architect) within IGC, where each user has different requirements - what information do they wish to view, the scope of that information, the actions available - and each user would like to include personalized shortcuts and bookmarks, and included dashboards or reports.



HOUSE

Search

Activity

2/3/17, 5:35 PM Pre-qualify added to Terms

1/31/17, 12:55 PM 010048582 added to Collections

1/16/16, 9:15 AM 23984_new modified by Andrew Smith

1 2 ··· 50 >

IGC Latest New Assets

Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget.Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo.

Queries

- Lorem ipsum dolor sit amet
- Consectetuer adipiscing elit, sed diam nonummy
- Nibh euismod tincidunt ut
- Laoreet dolore magna aliquam erat
- Volutpa tation ullamccommodo conse

Collections

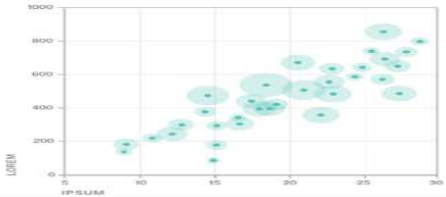
Lorem ipsum | 010048582 | quis_erat-new

Ut wisi enim | Ut wisi enim02 | AT VERO 001

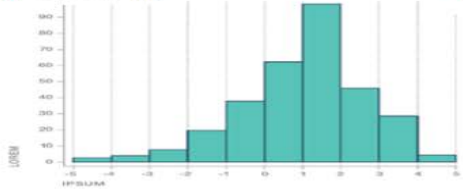
Subscriptions

Ut wisi enim
Ad minim
Veniam
Exerci tation
Commodo conse

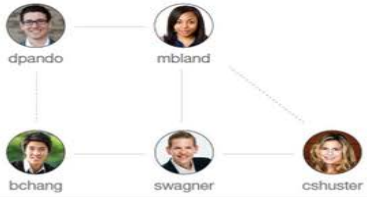
Workflow



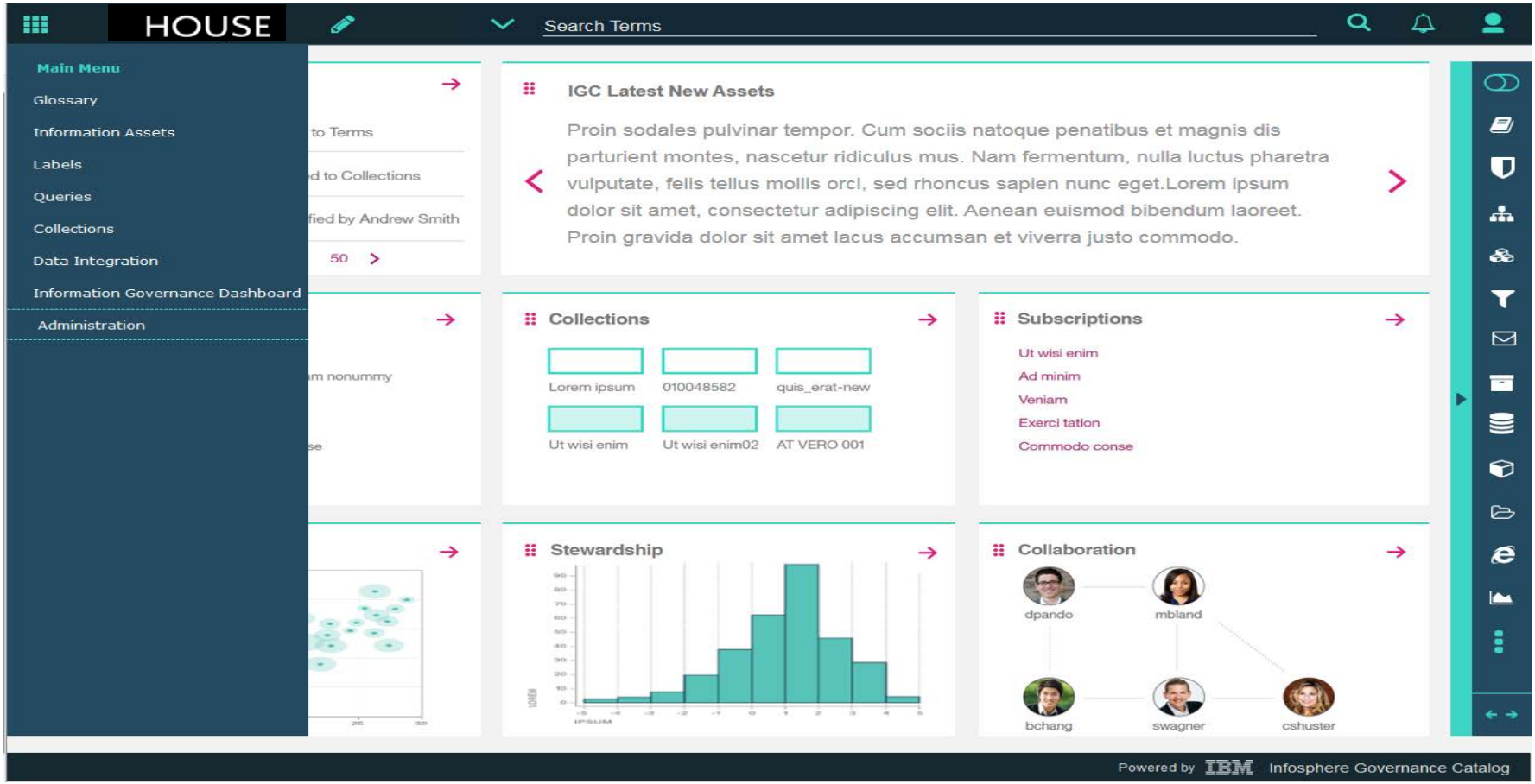
Stewardship



Collaboration



Powered by IBM Infosphere Governance Catalog


HOUSE
Search Terms
Main Menu
Glossary
Information Assets
Labels
Queries
Collections
Data Integration
Information Governance Dashboard
Administration
to Terms
to Collections
fied by Andrew Smith
50
IGC Latest New Assets
Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget.Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo.
Collections
Lorem ipsum
010048582
quis_erat-new
Ut wisi enim
Ut wisi enim02
AT VERO 001
Subscriptions
Ut wisi enim
Ad minim
Veniam
Exerci tation
Commodo conse
Stewardship
Collaboration
dpando
mbland
bchang
swagner
cshuster
Powered by IBM Infosphere Governance Catalog

Contextual Search – The heart of LabBook is a graph that captures **context**

Business Users

Business Analysts

Data Scientists

IT Staff

Social Widgets

Contextual Search

Recommenders

Contextual Graph Browser

Activity Streams

Embedded Widgets

slack

Social Networking Tools

IBM Watson Analytics

tableau

Business Analyst Tools

jupyter

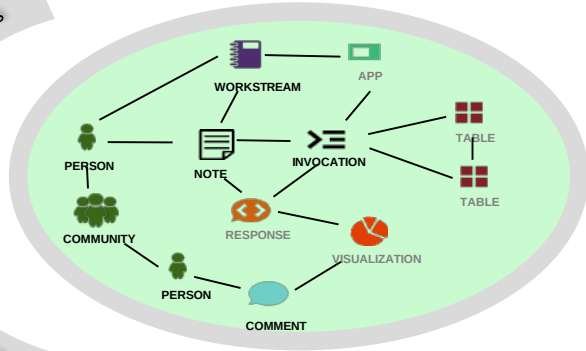
Zeppelin

Data Science Tools

IBM InfoSphere Spark

Data Integration Tools

I/O APIs + LOG EXTRACTION + HI-END APIs



Contextual Usage Graph

Contextual Search

The Heart of LabBook is a contextual Graph

Business Users

Social Widgets

Contextual Search

slack

Social Networking Tools

Business Analysts

Recommenders

IBM Watson Analytics

tableau

Business Analyst Tools

Data Scientists

Contextual Graph Browser

jupyter

Zeppelin

Data Science Tools

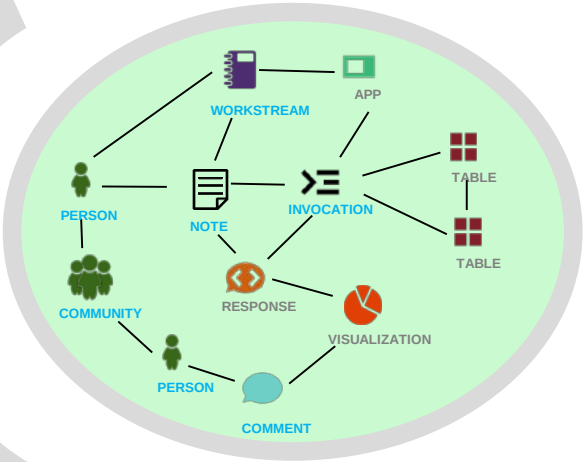
IT Staff

Activity Streams

IBM InfoSphere Spark

Data Integration Tools

I/O APIs + LOG EXTRACTION + HI-END APIs



Contextual Usage Graph

Embedded Widgets

… populated and consumed by 3rd party tools

Source systems

Contextual Usage Graph

Embeddable Widgets

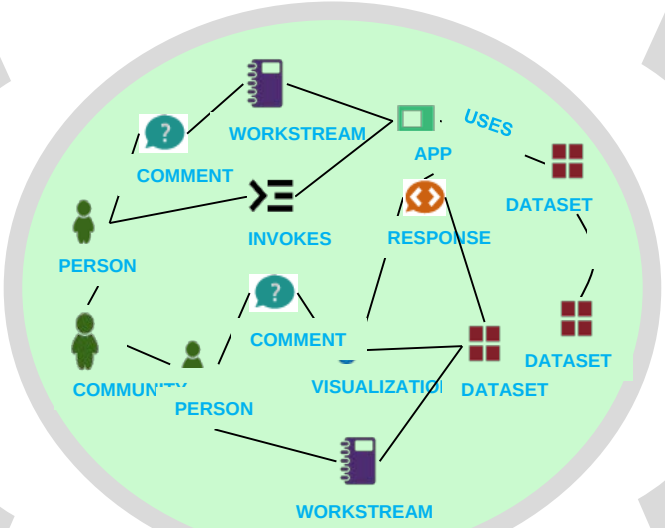
User Interfaces

Social Networking Tools

Business Analyst Tools

Data Integration Tools

REST API + LOG EXTRACTION



REST API

Social Widgets

Contextual Search

Activity Streams

Contextual Graph Browser

Recommendations

Business users

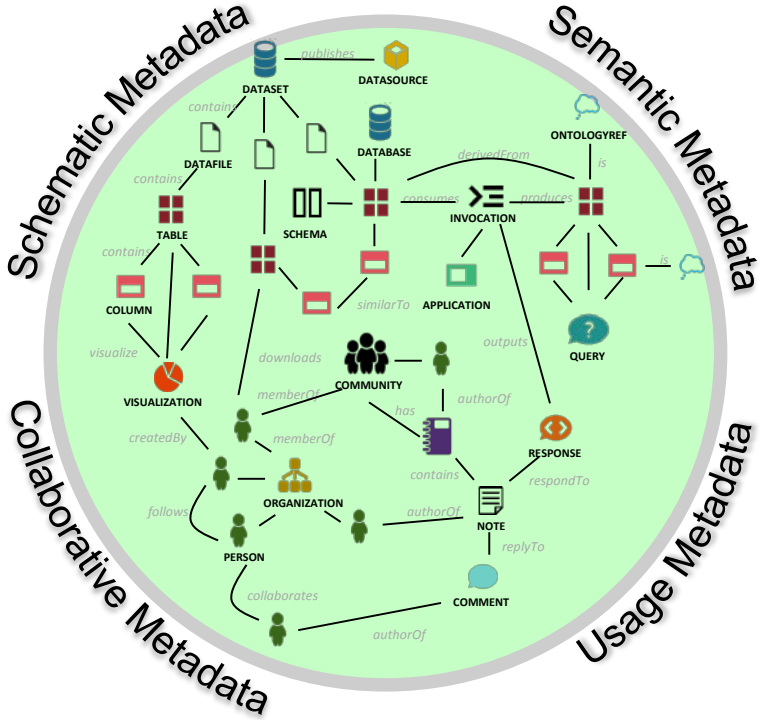
Business analysts

Data Scientists

IT staff

What context do we capture in the Graph?

- Schematic
  - How data is structured
- Semantic
  - What data means
- Collaborative
  - How people work together
- Usage
  - How data is used



Business & Historical Lineage

- *Support the requirement of the business to save and compare historical Lineage requirements to support compliance and regulatory requirement.*
  - Support ability to save and retrieve Lineage Reports, with additional descriptors
  - Support ability to compare versions of saved Lineage Reports, or compare to the current Lineage View
- *Support the requirement of the business to view Lineage according to the language and level of understanding of the business – according to domain or project.*
  - Support ability to create and manage Lineage Projects or Domains
  - Support ability to render, filter and customize Lineage according to Lineage Projects or Domain
  - Support ability to expand Lineage and view detailed view of data flow

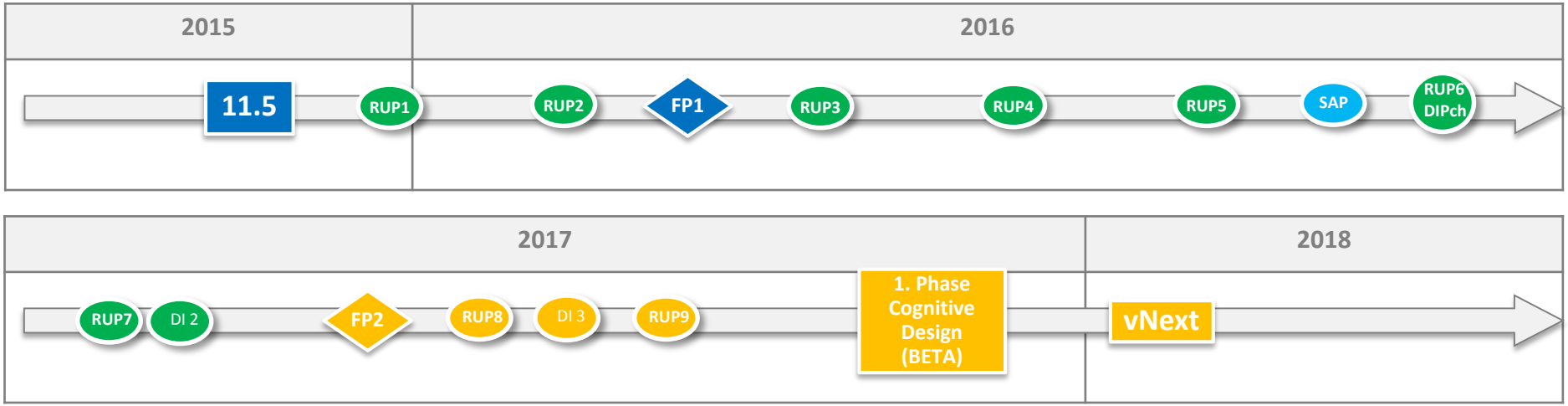
Near Term Mile Stones

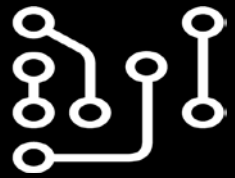
Time line

Short Term -- Q2/2017	Mid Term --- 2H 2017	Long Term -- 2018
Fixpak 2 for Information Server (v11.5.0.2)	1. Phase of Modernized platform foundation: • New Platform API	Information Server vNext • UG – LabBook • IA Thin client completion
Continues Connectivity enhancements	BETA: 1. MVP Cognitive Designer • Design, compile and run DS Jobs through a browser-based UI	Cognitive Designer: • Self-service Data Pre/Integration support • Shop for Data Experience
Continues Hadoop certifications	BETA: 1. Phase Dual Engine Support	Continuation on Dual Engine support: • run jobs on PX Engine or Spark
Download & Go for Governance		Continuation on Platform Modernization
		Download & Go for Integration
	Continues Connectivity enhancements	Continues Connectivity enhancements
	Continues Hadoop certifications	Continues Hadoop certifications

2016 / 17 Information Server Milestones

- ***Agile it is !*** – Continuous delivery of new features through rollup patches. No more waiting for new features until the next release.



RUP = Governance/Quality Rollup Patches
DI = Connectivity / Integration Rollup Patch
FP = Fixpak
vNext = Next release of Information Server

THANK YOU