Lunch n Learn Series: Data Governance

Please bring your lunch and join us on Tuesday, November 23rd, 2021 at 12:30 PM at our third monthly Risk Management Lunch & Learn Series event for Apple Bank colleagues on Data Governance. Dmitriy Khegay and Suresh Jayaraman will be presenting an overview on the Data Governance at the Bank and will provide a Data Governance Platform overview – Collibra/Precisely (Trillium) Tools. The event will be held in the 5th floor conference room behind reception. Attendance is limited to the first 12 people in person, and there will also be a video conference line open.

Topic Overview:

Data Governance helps ensure that Apple Bank's data assets are identified, usable, accessible, and formally managed. Data Quality is an integral part of Data Governance that ensures that the data is trustworthy and suitable to serve its intended business purpose. Good Data Governance is critical for improved data quality, regulatory compliance, decreased data management costs, and increased access to data for all stakeholders. The result is better decision making and better business outcomes.

· Data Governance overview at Apple Bank

· Data Governance Platform at the Bank – Collibra / Precisely(Trillium) Tools overview

· Data Governance Guest Speaker – Philip Black from Data QG (http://dataqg.com/). A trusted platform with leading resources, talent, and opportunities to empower DG Community through connection.

Dmitriy Khegay Bio

· Dmitriy Khegay joined the Data Governance team in December 2019 and leads the Data Governance Operations initiatives. Dmitriy is a Data Governance Officer aligned to CECL and Credit departments and co-lead on Finance, Accounting, and Treasury domain. His prior experience includes Data Governance, Data Quality and Business Analytics roles in Finance, Risk Management from Credit Suisse, American Express, BNY Mellon, and HSBC.

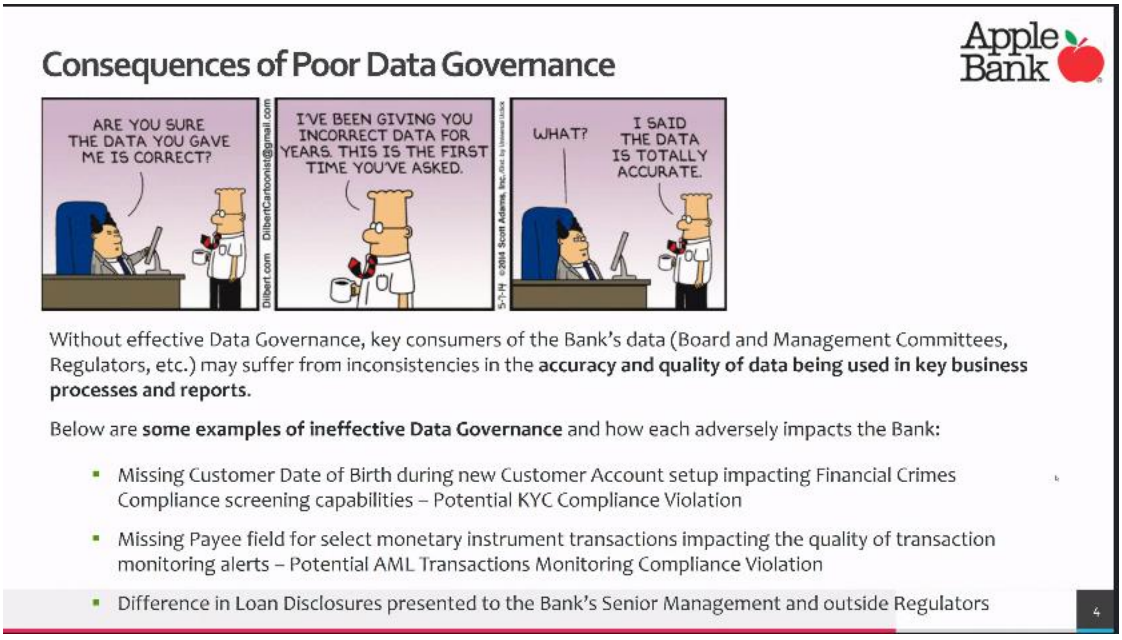
Suresh Jayaraman Bio

· Suresh Jayaraman joined the Data Governance team in May 2021 and leads the Data Quality initiatives. His prior experience includes Data Governance, Data Quality, Analytics, and Technology roles in Risk Management from Goldman Sachs, Morgan Stanley, Bank of America Merrill Lynch, and AIG.

When

Tuesday, November 23rd, 2021

12:30 PM to 1:30 PM



Data Governance Implementation

Achieving high performance in Data Governance is a journey that spans multiple years of commitment.

Chaotic	Emerging	Structured	Governed
- Inconsistent data - Business has low confidence in data - Difficult to adapt to business changes - No overall strategy to master or structure data	- **Same data is maintained in multiple systems** - **Limited cross functional use of data due to silos** - Data quality is seen as an IT problem	- Data is viewed as an asset to be actively managed - Fully leveraging our partnership with our Business Data Steward community - A **centralized data dictionary** is fully documented and regularly updated	- Business ownership and control of data - Business has high level of trust in its own data (Partnership with Data Governance to produce DQ Scorecard Metrics)

Current State of the Data Governance Program at AB

NETFLIX Capital One FedEx

Outcomes of Good Data Governance

Good Data Governance → Good Data → Good Business Outcomes

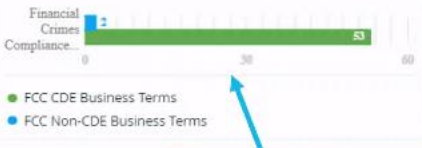
Outcomes (measurable)	
	• **The assurance that high quality data is being consumed in key business reports and processes**
	• The availability of a **Centralized Data Dictionary (key business terms and definitions) in Collibra DG Platform**. This ensures that data consumers clearly understand what data they are consuming

Collibra Platform Demonstration – Financial Crimes Compliance

FCC Dashboard

Welcome to the FCC Dashboard!

FCC Business Terms (Unique Count)
Chart representing asset or task metrics.



Current snapshot of Critical Data Element "Crown Jewels" under the DG Program

FCC Business Term Glossaries

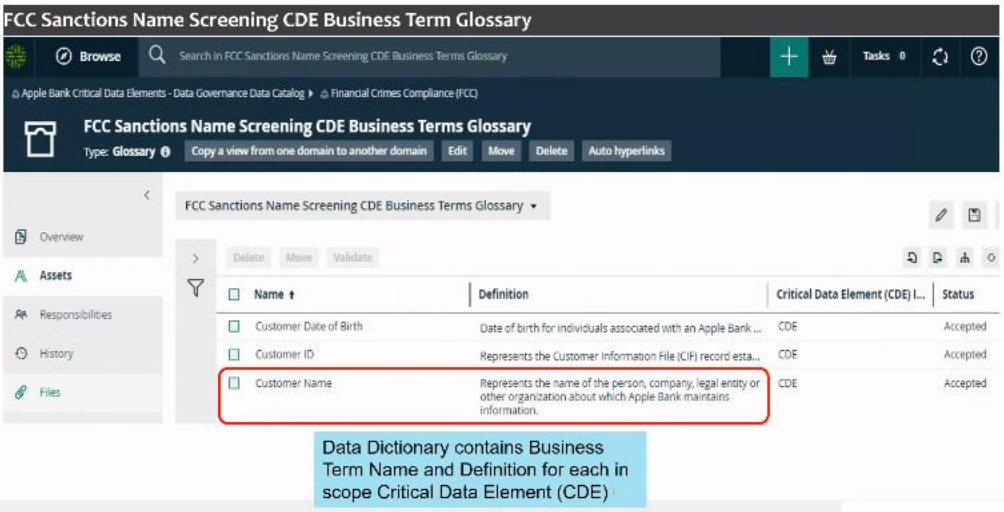
FCC Full Listing of Business Terms
FCC Sanctions Name Screening Business Terms Glossary
FCC Sanctions Payment Screening Business Terms Glossary
FCC Transactions Monitoring Business Terms Glossary

Hyperlink for easy access to Business Glossary/Data Dictionary information

Data Quality Issue - Setup "Short" Form (GRC Tool)

GRC DQ Issue Short Form Link

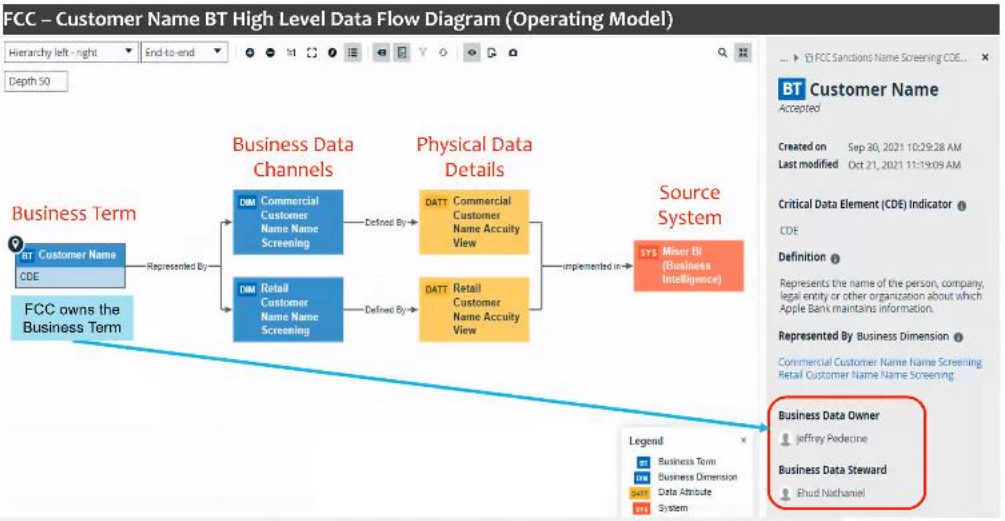
Collibra Platform Demonstration – Financial Crime Compliance



Data Dictionary contains Business Term Name and Definition for each in scope Critical Data Element (CDE)

8

Collibra Platform Demonstration – Financial Crime Compliance



9

Collibra / Precisely – Data Quality Profiling/Scorecard Process

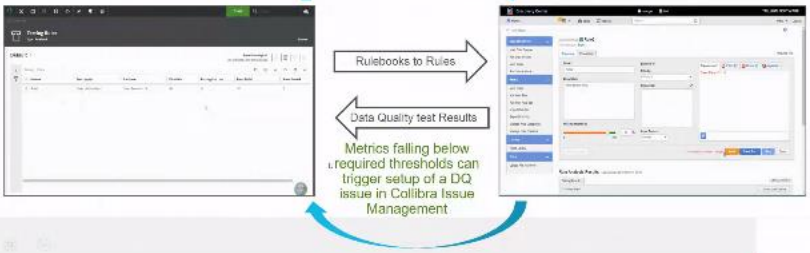
Apple Bank

Collibra Data Governance Center

- Enables non-technical users to define business policies and data quality rules in plain language
- Makes data quality performance available to all users

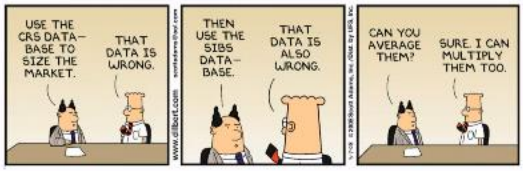
Trillium Discovery (Precisely)

- Imports business rules so a technical user can convert to executable data quality rules
- Constantly runs data quality metrics on **a near real-time basis**, and passes results back to the Collibra data quality dashboards



What is Data Quality?

- An integral part of Data Governance
- Data quality is the ***measure of suitability of data to serve its intended business purpose*** (i.e., Trust Worthiness and Fitness for Purpose)
- **Poor data quality leads to poor decisions.** Analysis paralysis leads to failure
- Ensuring data used for analysis and interpretation is of good quality is critical
- This is easier said than done



"If you can't measure it, you can't manage it."

This maxim ranks high on the list of quotations attributed to Management Guru **Peter Drucker.**

Spot the Data Quality Issue - 1

Common data quality issues that are easy to spot are missing or duplicate values, and inconsistent data

MLB 2021 Top Batters

Player	Team	Pos	Age	Height	Weight	G	AB	R	H	HR	RBI	AVG	Top-100
Anthony Rendon	LAA	3B	31	73	200	58	217	24	52	6	34	0.240	8
Christian Yelich	MIL	OF	29	75	195	117	399	70	99	9	51	0.248	9
Cody Bellinger	LAD	1B	26	75		95	315	39	52	10	36	0.165	10
DJ LeMahieu	NYY	2B	33	76	220	150	597	84	160	10	57	0.268	14
Fernando Tatis	SD	SS	22	75	217	130	478	99	135	42	97	0.282	6
Fernando Tatis	SD	SS	22	75	217	130	478	99	135	42	97	0.282	6
Juan Soto	WAS	OF	23	74	224	151	502	111	157	29	95	0.313	5
Mike Trout	LAA		30	74	235	36	117	23	39	8	18	0.333	1
Mookie Betts	LAD	OF	29	69	180	122	466	93	123	23	58	0.264	2
Nolan Arenado	STL	3B	30		215	157	593	81	151	34	105	0.255	13
Ronald Acuna	ATL	OF	23	72	205	82	297	72	84	24	52	0.283	7
Teoscar Hernandez	TOR	OF	29	74	93	143	550	92	163	32	116	0.296	

Spot the Data Quality Issue - 2

- Other issues are difficult to spot
- Detecting and handling such problems require domain knowledge

Can you spot the three data issues in the data set below ?

MLB 2021 Top Batters

Player	Team	Pos	Age	G	AB	R	H	HR	RBI	AVG	Top-100
Anthony Rendon	LAA	3B	31	58	217	24	52	6	34	1.240	8
Christian Yelich	MIL	OF	29	117	399	70	99	9	51	0.248	9
Cody Bellinger	LAD	1B	62	95	315	39	52	10	36	0.165	10
DJ LeMahieu	NYY	2B	33	150	597	84	160	10	57	0.268	14
Fernando Tatis	SD	SS	22	130	478	99	135	42	97	0.282	6
Juan Soto	WAS	OF	23	151	502	111	157	29	95	0.313	5
Mike Trout	LAK	OF	30	36	117	23	39	8	18	0.333	1
Mookie Betts	LAD	OF	29	122	466	93	123	23	58	0.264	2
Nolan Arenado	STL	3B	30	157	593	81	151	34	105	0.255	13
Ronald Acuna	ATL	OF	23	82	297	72	84	24	52	0.283	7
Teoscar Hernandez	TOR	OF	29	143	550	92	163	32	116	0.296	

Spot the Data Quality Issue - 2

MLB 2021 Top Batters

Player	Team	Pos	Age	Height	Weight	G	AB	R	H	HR	RBI	AVG	Top-100
Anthony Rendon	LAA	3B	31	73	200	58	217	24	52	6	34	1.240	8
Christian Yelich	MIL	OF	29	75	195	117	399	70	99	9	51	0.248	9
Cody Bellinger	LAD	1B	62	75	203	95	315	39	52	10	36	0.165	10
DJ LeMahieu	NYY	2B	33	76	220	150	597	84	160	10	57	0.268	14
Fernando Tatis	SD	SS	22	75	217	130	478	99	135	42	97	0.282	6
Juan Soto	WAS	OF	23	74	224	151	502	111	157	29	95	0.313	5
Mike Trout	LAK	OF	30	74	235	36	117	23	39	8	18	0.333	1
Mookie Betts	LAD	OF	29	69	180	122	466	93	123	23	58	0.264	2
Nolan Arenado	STL	3B	30	74	215	157	593	81	151	34	105	0.255	13
Ronald Acuna	ATL	OF	23	72	205	82	297	72	84	24	52	0.283	7
Teoscar Hernandez	TOR	OF	29	74	205	143	550	92	163	32	116	0.296	

1. Mike Trout plays for the LA Angeles not the LA Lakers basketball team
2. It is rare for a Major League Baseball Player to be over the age of 40, much less be 62 years old
3. Batting average of 1.24 is invalid (perfect batting average is 1.000); in fact very few batters have come close to the great Ty Cobb's 0.3662 lifetime average

- Domain specific business rules are required to check for Accuracy, Consistency, and Validity
- Apple Bank uses appropriate Data Quality Business Rules to assess regulatory compliance and to implement business practices to optimize quality assurance

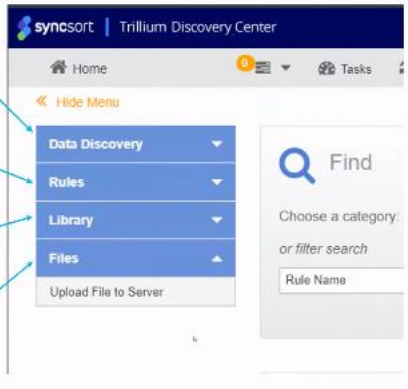
Trillium Discovery Center - Capabilities

Understand your data, discover data distributions, relationships, dependencies

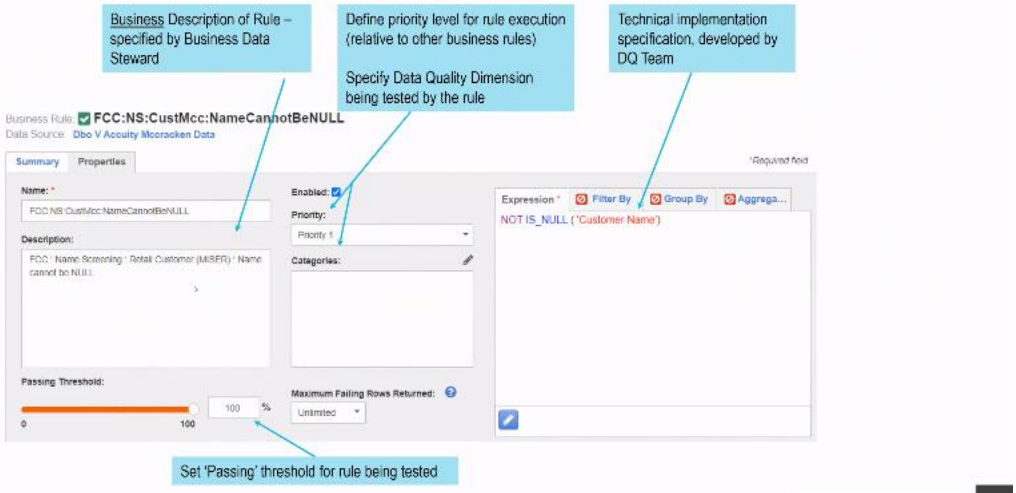
Create and execute Business Rules to assess regulatory compliance and to implement business practices to optimize quality assurance

Organize, store and share business rules in Libraries

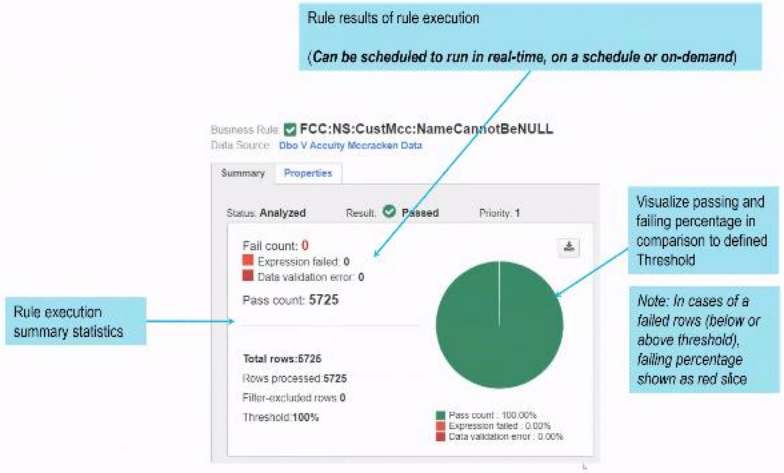
Upload files from local system sources to for ad-hoc (informal) data discovery



Business Rules Creation - Example



Business Rules Execution, Results - Example



Data Quality Scorecard – FCC Example

Banded (Impact-Level): Pass/Fail results by Data Element and Business Rule

Drill-down to view details of a Failing rule

Discover underlying cause(s) for Root Cause Analysis

Impact Level (Defect Rate)

- N/A
- Very Low, and Low (< 2%)
- Medium (2% - 5%)
- High (5% - 10%)
- Very High (>= 10%)

Max of Impact Level Number		DQ Test Name								
Data Source	Number of Key Data Elements Tested	Blanks/All Whitespaces	Date Open < Date Close	Default Dates	Default Numeric Values (All 0s/9s)	Duplicate Values	Max Length	Negative Transaction Amounts	Null values
Accounts	5	PASS			PASS	PASS	PASS		PASS
Customers	3	PASS		PASS	PASS	PASS	PASS		PASS
master_portfolio	9	PASS		PASS	PASS		PASS		PASS
TransactionAchDetails	2	PASS					PASS		PASS
TransactionMonetaryInstrumentDetails	2	PASS					PASS		PASS
Transactions	10	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
TransactionWireDetails	2	PASS					PASS		PASS
VDM_FK_InclearingTransactions	6	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
VDM_FK_PODTransactions	6	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
VDM_WireTransactions	8	PASS					PASS		FAIL
v_Accuity_MISERBI_Data	8	PASS		PASS	PASS		PASS		PASS

19

Data Governance Guest Speaker

Phil Black
Founder @DataQG

Definition of Data Governance...

Preparing Data for Insights and Compliance

Data
Biggest Asset
Biggest Liability

"The ~~three~~ four Ps"
People, Process, Platform,
~Politics~

DataQG's
Philosophy

- No Standard
- Opinion Based
- Choose the best fit

"What do you think is the most important factor when starting a data governance program?"

36.8% Business Case Identification
31.6% Company Culture
26.3% Executive Leadership
5.3% Robust Software Solution

Treat Risk, Save Money, Make Money

Every Bank takes a different approach.

Culture, resources, business drivers and sponsorship will dictate your path.

"It is neither cost-effective nor prudent to devote the same level of resources to every data element"

Data Quality for Risk in Banking

Completeness: Will missing values impede the Bank's ability to meet a regulatory requirement?

Accuracy: Lack of standardization, be identified and limit the usability of the data.

Consistency: Relationships with other data elements on that record. Integrity of data.

Timeliness: Monitor if the data is refreshed as frequently as the business need dictates.

Defense vs. Offense

Foundation

Regulations - Privacy - Insights

Food for Thought.

More data sources & more regulations
Fastest growing area of data outside data science
Competitive Insights
AI/ Machine Learning
Why do you think I started this company?

The global enterprise data management market size was estimated at USD 72.79 billion in 2020 and is expected to reach USD 122 billion in 2025

Join Us

DataQG.com/community

Pblack@dataqg.com