• Introduction Adrian
• Data Governance Updates Oksana / Adrian
  • A Random Walk Through Data Governance
• AIDA Updates Adrian
• ODG Project Updates Tammana / Sara
• Collibra DGC Updates Chrissy
• Announcements & Events
• Q&A
Data Governance Updates
1. Conduct DG Maturity Assessment

- Assessing data governance maturity level to identify focus areas
- Identify gaps to resolve through Data Governance project

*Outcome: Gartner Document; Data Governance Framework

2. Establish DG Program

- Build robust Data Governance structure at the University level
- Build practical and repeatable standard processes across different business and functional areas
- Establish local Data Governance structure with linkage to the central program - e.g. Abu Dhabi

3. Execute DG Projects/Processes

- Propose the Data Governance projects through Enrollment Management and NYUIT yearly planning process to seek priority decisions and funding support
- Establish data risk control processes and measurements (KPIs, KRIs) based on various compliance and risk management requirements

*Some subject areas exist for this stage: Identity, Student Life, Public Safety, Census, Data Quality Census Validation, Crosswalk

4. Monitor DG Program Effectiveness

- Identify, develop, and deliver Key Performance Indicators and Key Risk Indicators for the data governance program
- Establish ongoing Data Governance processes

*Some subject areas exist for this stage: Identity, Student Life, Public Safety, Census, Data Quality Census Validation (KRIs, KPIs), Crosswalk

Progress across work areas is not even. Some work streams are more developed than others.
Policy
- Policies drive data governance activities

Coordination
- Data and business capabilities are aligned
- Roles and Responsibilities for data asset management are clearly defined and adopted

Data
- Critical data elements (CDEs) inform priorities and development
  CDEs are driven by the business areas (by Stewards)

Metadata
- Clear business definitions
- Technical and business assets mapped
- Uniform accountability

Access
- CDEs are presented and managed within context

Quality
- Improve data quality for CDEs using standards, rules and links

Data Management Risk Control
- Goal 1: Improve data quality to enhance trust in NYU data
- Goal 2: Support easy and secure access to University data assets
- Goal 3: Reduce risk through regulatory, policy and procedural compliance
- Goal 4: Establish a data governance program to improve NYU's ability to create, preserve, and disseminate knowledge
What’s the Big Picture?
“The right data to the right people at the right time”

Note: Cliff’s Journey adapted courtesy of Collibra.
Data Owner / Producer / Author Perspective Value

(Data Trustees, Data Stewards, Data Custodians)
Critical data elements are those that are determined to be vital to the successful operation of the organization.

The business glossary establishes unambiguous definitions for business terms.

**What's the difference?**

FIRST NAME (OFFICIAL) OR FIRST NAME (PREFERRED)?
An **asset** is the building block for which you want to capture information. A **business term** is an asset.
A Random Trip through Data Governance: Domain

Policy

First Name

Identity Glossary

Metadata

The given name by which an individual prefers to be identified.

A domain is an organizational concept in Collibra; it is a logical grouping of assets.

Implementation

Quality

FIRST NAME (PREFERRED) ALIGNS TO THE IDENTITY GLOSSARY.
The NYU Functional Capability Model (Business Capability Model) serves as a reference for university-wide business activities and alignments.
A Random Trip through Data Governance: Community

A community is an organizational concept in Collibra; it usually defines a part of stewardship.

IDENTITY MANAGEMENT ACTIVITIES ALIGN TO A PERSON COMMUNITY, MAINTAINED BY IDENTITY STEWARDS.

Identity Glossary

Identity Management

Person

Identity Glossary

First Name (Preferred)

Metadata

The given name by which an individual prefers to be identified.

Data

First Name

Policy

Coordination

Implementation

Quality
COMMUNITIES ARE USEFUL IN RECOGNIZING AND ASSIGNING ROLES & RESPONSIBILITIES OVER ASSETS.

Assets are aligned to roles in the Data Governance Framework including Data Trustees, Data Stewards, and Data Custodians. Roles are referenced in workflows.
The given name by which an individual prefers to be identified.

The catalog provides technical metadata.
A Random Trip through Data Governance: Catalog

Policy

Coordination
- Identity Management
- Person
- (Identity Data Stewards)
- Identity Glossary
- First Name (Preferred)

Data
- First Name

Metadata
- The given name by which an individual prefers to be identified.
- (AIDA Information Mart; Alphanumeric; “Joe”; etc.)

Implementation
- SLIP
- Sail Point
- Snowflake

Quality

Catalog may show data location, attributes & examples but not actual data.

We see how the data is implemented.

The catalog provides insight on technical metadata.
A Random Trip through Data Governance: Risk Category

<table>
<thead>
<tr>
<th>Policy</th>
<th>Implementation</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance (GDPR, HIPPA, etc.)</td>
<td>SLIP</td>
<td>Snowflake</td>
</tr>
<tr>
<td>Electronic Data &amp; System Risk Classification Policy</td>
<td>Sail Point</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coordination**
- Identity Management
- Person
- (Identity Data Stewards)
- Identity Glossary
- First Name (Preferred)

**Data**
- First Name

**Metadata**
- The given name by which an individual prefers to be identified.
- (AIDA Information Mart; Alphanumeric; “Joe”; etc.)

**Implementation**
- SLIP
- Sail Point
- Snowflake

**Quality**

**Notes**
- Risk and access restrictions can be described.
- Policies may inform such attributes.
- Business knowledge is critical.

Policy references are available.
A Random Trip through Data Governance:
Quality

WHEN RELATED, THIS METADATA WILL HELP ASSURE DATA QUALITY.
A Random Trip through Data Governance: Data Governance Model and Relationships

Data Management Risk Control

1. Improve data quality to enhance trust in NYU data
   
   Key Supports: Coordination; Data; Metadata; Quality

2. Support easy and secure access to University data assets
   
   Key Supports: Policy; Coordination; Data; Metadata; Implementation

3. Reduce risk through regulatory, policy and procedural compliance
   
   Key Supports: Policy; Coordination; Data; Metadata

4. Establish a data governance program to improve NYU’s ability to create, preserve, and disseminate knowledge
   
   Key Supports: Policy; Coordination; Data; Metadata
Data Consumer Perspective Value

(Analysts, Data Consumers, BI Developers, API Developers, etc.)
Note: Cliff’s Journey adapted courtesy of Collibra.
AIDA Updates
**AIDA Project Snapshot FY20**

**Institutional Data Assets**
- **1000+** Data Attributes
- **22** Business Glossaries

**Org Enablement**
- **5+** Strategic Projects
- **5** Efficiency Opportunities

**Data Governance Support**
- **900+** Business Terms
- **2300+** Data Points’ Lineage

**Details**
- 20+ subject areas
- 173 Information Mart tables, 1044 attributes
- 400 Data Vault tables, 1392 attributes
- 400+ source tables, 2389 attributes

**Highlights**
- **NYU Returns**
  - Emma (Email Direct replatform)
  - Badging (Registry replacement)
  - Efficiencies via standardization, modularization, and reusability

- **Consolidating legacy NYU Business Glossary**
- **Content across functions (e.g., Student Life, Registrar, Public Safety, Finance, Identity, Space and more)**

- **Identity**
  - Emma (Biographic, Contact)
- **Student**
  - Courses, Roster
  - Engagement
  - Athletics
  - Residential Services
  - Career Outcome
- **Public Safety**
  - Tap Access
- **Space / Building**
  - Space Attributes
  - Reference Data
Clients use **Emma** to bulk email select constituencies. Emma relies on Mulesoft (APIs) for integration via AIDA:
- consistent
- reusable

**Critical data elements** aligned with business needs:
- consistent
- reusable

Emma-related data is modeled and integrated with other University data sets and data assets.

Data is staged for modeling and integration.

The Emma team identified critical elements, terms and context, and explained business processes.

Reference: [AIDA Development Process](#)
ODG Project Updates
### Objective:
To standardize academic and administrative names & codes across NYU enterprises affecting reports that determine data access and engage with senior leadership.

**Project Objective:**
- Short Term: Restore historical logic to update Reporting Department Crosswalk (RDC) - Complete
- Long Term: Fully automate the manual processes of the RDC – On hold

### Project Outcomes
- Restored historical logic for Crosswalk update
- Processed 4 cycles of data refresh (Last Update: Nov, 2020)
- Increased engagement between data custodians and NYU IT teams to support technical processes

### Next Steps
- Streamline & clean up the Crosswalk and Hierarchy reports
- Data refresh process to update Crosswalk will be run bi-weekly, next refresh: Nov 21, 2020
Objective: Review, update, and confirm the structure and names within the Data Governance Participants List to ensure we appropriately represent and have buy-in from areas of data around the university.

Overview:
- Review DG Ptp List for unconfirmed or missing information
- Work with Data Trustees to identify and update the structure of and names within their area
- Work with NYU IT to identify technologies and Data Custodians within areas
- Develop procedures for making future changes to this list and onboarding new participants

Progress Updates on Review

- All (17) Business Areas analyzed
- 10 Business Areas complete / confirmed
- 2 Business Area pending response
- 5 Business Area require further analysis, will be moved to a Phase II

<table>
<thead>
<tr>
<th>Project Phase Updates</th>
<th>Completed Items</th>
<th>Outstanding Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery – Completed</td>
<td>• Added new communities to increase representation of Data Stewards</td>
<td>• Complete Phase I - Confirmations</td>
</tr>
<tr>
<td>Design – Completed</td>
<td>• Restructured certain areas to better align with the organization and ownership within NYU</td>
<td>• Start Phase II - DG Ptp List Expansion</td>
</tr>
<tr>
<td>Implementation – In Progress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-Implementation – Not Started</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Objective: Strengthen the validity and reliability of NYU's census data by streamlining and making more efficient the process by which we validate this data.

Overview:
- Map the current state process and existing business rules for census data validation
- Develop Census Data Validation Framework and identify areas within current state for improvement and increased efficiency
- Build new solution including automated query, validation rules, KRI scorecard, and KPI dashboard

Progress Updates on New Assessment Process

New Data Validation Process run for Fall 2020 Term - Tableau Dashboards & Scorecards, KPI Scorecards

| Outcomes - KRI:                      |  ➢  Overall DQ score increased over validation period  
|                                   |  ➢  100% DQ Score for one high-priority CDE (unprecedented)  
|                                   |  ➢  ~50% of DQ rules’ logic refined to be more accurate  

| Outcomes - KPI:                      |  ➢  ~10 hrs / week of IR effort saved due to automation  
|                                   |  ➢  ~5 hrs / week of Registrar effort saved due to self-service dashboards  

Term Enrollment Census Validation Scorecard in Tableau
Census Data Validation Process Improvement Project: Proof of Concept for Data Quality Processes

<table>
<thead>
<tr>
<th>Project Phase Updates</th>
<th>Completed Deliverables</th>
<th>Next Steps</th>
</tr>
</thead>
</table>
| • Initiation – Completed  
• Discovery I & II – Completed  
• Implementation – Completed  
• Post-Implementation – In Progress | • Census Data Validation Framework  
• New business and technical process  
• Automated query  
• Automated dashboards  
• Semi-automated DQ and KPI scorecards  
• Detailed procedures and documentation | • Review and align ETL Processes  
• Leverage deliverables of this project to scale DQ within NYU |

Gartner Data Governance Goals

● Goal 1: Improve Data Quality and enhance trust in NYU data

Strengthen data quality & reliability of NYU’s data by:

● Formalizing Data Quality dimensions  
● Creating Data Quality rules  
● Saving results from each validation run to track DQ scores over the validation process  
● Increasing communication with clients via working sessions  
● Improving process efficiency
  ○ Data validation was automated via Tableau and processed daily  
  ○ Data validation was built on ongoing data, allowing the process to start earlier
Collibra DGC Update
What will be asking you to do in Collibra?

<table>
<thead>
<tr>
<th>Task</th>
<th>Stewards</th>
<th>Custodians</th>
<th>Trustees</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add / Update Business Terms and Definitions</td>
<td>✔️</td>
<td></td>
<td></td>
<td>Develop a common vocabulary and improve data literacy</td>
</tr>
<tr>
<td>Review / Approve Business Terms</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>Ensure that the information we have is vetted by experts</td>
</tr>
<tr>
<td>Review Technical Metadata</td>
<td></td>
<td>✔️</td>
<td></td>
<td>Capture up to date physical data structures</td>
</tr>
<tr>
<td>Link Business and Technical Metadata</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>Help data analysts connect what data means with <em>where</em> they can find it</td>
</tr>
<tr>
<td>Review / Approve Reference Data</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>Improve data literacy and reporting capabilities</td>
</tr>
</tbody>
</table>
Announcements & Events
Questions?
Thank you!
Appendix
Detailed Data Governance Operating Model & Relationships

[Diagram showing various data governance processes and their relationships, including compliance policies, data classification, critical data elements, and data quality metrics.]

NYU Capability Model is a DRIVER to Calibra's Operational Model:
Level 1. University Area align to Community (Governed by Data Trustee)
Level 2. Functional Area align to Sub-Community (Governed by Data Steward)
Level 3. Business Function align to Business (Governed via Sub-Community/Data Steward)
### Value Proposition & Alignment between DG Model and DGSS

<table>
<thead>
<tr>
<th>ID</th>
<th>DG Model Description</th>
<th>Data Governance Model’s Value Proposition</th>
<th># RMap</th>
<th>DGSS Services and Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drive from Policies</td>
<td>Supplements Business Terms with related Policy alignment details, mandatory training</td>
<td>1.6</td>
<td>Data Privacy and Protection</td>
</tr>
<tr>
<td>2A</td>
<td>NYU Capability Model v18</td>
<td>Maintain consistent alignment between data governance participants and enterprise architecture business functions.</td>
<td>1.1.1</td>
<td>Assign Participants</td>
</tr>
<tr>
<td>2B</td>
<td>Support Collibra Data Operating Model</td>
<td>Facilitate the responsibility and workflows around data assets with consistency and alignment to the University’s capabilities and business functions.</td>
<td>1.1.1</td>
<td>Assign Participants</td>
</tr>
<tr>
<td>2C</td>
<td>Align with Data Governance roles (Trustees &amp; Stewards)</td>
<td>Facilitate the responsibility, maintenance, workflows, and accountability over data assets through specified roles (e.g., Trustees, Stewards, etc.)</td>
<td>1.1.2</td>
<td>Approval Process</td>
</tr>
<tr>
<td>3</td>
<td>Critical Data Element (CDE)</td>
<td>Identify and develop priority data elements.</td>
<td>1.1</td>
<td>Data Governance Workflow Management (What type of data do we have?)</td>
</tr>
<tr>
<td>4A</td>
<td>Use Case 1: Database</td>
<td>Develop the Critical Data Element, Business Glossary, and Data Catalog assets represented through a Database and in support of a particular business function.</td>
<td>1.4 &amp; 1.2.2</td>
<td>Master Data Management &amp; Technical Metadata Management</td>
</tr>
<tr>
<td>4B</td>
<td>Use Case 2: Reports</td>
<td>Develop the Critical Data Element, Business Glossary, and Data Catalog assets represented through a Report and in support of a particular business function.</td>
<td>1.4 &amp; 1.2.2</td>
<td>Master Data Management &amp; Technical Metadata Management</td>
</tr>
<tr>
<td>4C</td>
<td>Use Case 3: Application</td>
<td>Develop the Critical Data Element, Business Glossary, and Data Catalog assets represented through an Application and in support of a particular business function.</td>
<td>1.4 &amp; 1.2.2</td>
<td>Master Data Management &amp; Technical Metadata Management</td>
</tr>
<tr>
<td>5A</td>
<td>Business Glossary</td>
<td>Deliver clear, unambiguous understanding of the meaning of business terminologies, taxonomies, and other hierarchies, and how they differ across Business / Functional areas.</td>
<td>1.2.1</td>
<td>Maintain Business Metadata (Business Glossary)</td>
</tr>
<tr>
<td>5B</td>
<td>Data Catalog</td>
<td>Deliver documented reference material for logical and physical data structures and elements across the IT space; deliver traceability by relating business glossary assets and technology assets.</td>
<td>1.2.2</td>
<td>Maintain Technical Metadata (Map)</td>
</tr>
<tr>
<td>5C</td>
<td>Access Control Certification Process</td>
<td>Assure accountability over entitlements / access to data.</td>
<td>1.6.1</td>
<td>Access Request (Access Control)</td>
</tr>
<tr>
<td>6</td>
<td>Data Quality</td>
<td>Manages improvement of data quality over time for critical data elements using standards, rules, and links to 5A</td>
<td>1.3</td>
<td>Data Quality Management</td>
</tr>
<tr>
<td>7</td>
<td>Data Management Risk Control</td>
<td>Implements data management risk control aligned to 1 and 5</td>
<td>1.6</td>
<td>Data Privacy and Protection</td>
</tr>
</tbody>
</table>

**Note:** ID references align to Model and Relationships