

# Annual Report 2021-2022 University Data Governance Council

**August 2022** 

# University Data Governance Council Report on Progress 2021-2022 Academic Year

#### Introduction

Building on the work completed in 2021, the UDGC started the 2021-2022 academic year with three working groups: 1) Data Access, Security and Compliance; 2) Metadata and Data Dictionary; and 3) Data Literacy, Training and Tools. The topic of Data Quality was shifted to the University Data Commons group (UDC), which already had an existing working group for that; Abhay Joshi from UDC was added to the UDGC Roster to inform and collaborate.

This year, several members of the UDGC participated in data governance learning and collaboration opportunities with other higher education institutions, gaining insights into the maturity level of peer institution efforts, and sharing information on lessons learned. Neil Sedlak and Jon Teglas Clark enrolled in the <u>EDUCAUSE Learning Lab</u> for data governance, comprised of 4 sessions held online in March 2022, then shared what they had learned from the experience in the April 2022 UDGC meeting. In May of 2022, Brenda van Gelder participated on a panel discussion at the Common Solutions Group spring meeting (CSG), which included a data governance workshop. This event and subsequent discussion with other CSG universities yielded significant insight into how data governance is being implemented in higher education. Virginia Tech is on par with our peer institutions, with a data governance maturity level that is ahead of peers in some areas and behind in others (see Appendix B for results of a survey that was conducted by CSG in May 2022 about data governance activities). Collaboration with these universities will help Virginia Tech's succeed with data governance over the coming years.

#### 1. Milestones Completed

a. The UDGC Progress Report from last year included the milestones listed in the graphic below:



#### Timeline of UDGC activities and milestones Fall 2020-Spring 2023

As the list of deliverables from this academic year show, the UDGC has achieved the milestones set for spring 2022 and is well positioned to continue along this timeline in the coming year. The added organizational structure established through the IT Governance project, realizing recommendation 1.2 of the IT Transformation governance process provides opportunities to further refine and delineate the responsibilities of the UDGC.

#### 2. Working Group Deliverables (see Appendices for complete deliverables which are summarized below)

- a. Data Access, Risk and Compliance Develop data access request flowchart and list of questions for granting access.
- b. Data Access, Risk and Compliance Build a list of Data Trustee and Data Steward role descriptions standardized for consistency across university departments.
- c. Data Literacy, Tools, and Training Outline an internally developed, VT-specific training module.
- d. Data Literacy, Tools, and Training Build a database of free training resources on topics including data literacy, data-informed decision-making, and effective presentation techniques.
- e. Metadata and Data Dictionary Benchmark data collected from Common Solutions Group and EDUCAUSE institutions and assess data dictionary platforms commonly used in higher ed.

These deliverables are consistent with the priorities and activities of higher education peers as found in the results of the Common Solutions Group Data Governance survey of May 2022 (see Appendix C).

#### 3. Moving Forward

The future of data governance at Virginia Tech will include the formation of a new IT governance structure that is arising from the IT Transformation Program. As these structures are developed, the relationship and ongoing composition of the UDGC may also need to shift. It is expected that the UDGC will continue to have a role in advancing data governance maturity of the university. With this expectation in mind, the three UDGC working groups have each suggested next steps for their specific topic areas:

#### Data Access, Risk and Compliance Working Group – Suggested Next Steps

- 1. Potentially in coordination with the IT Job Architecture Project, develop a set of clearly defined university job roles which include named skills, expectations, and/or requirements in an ordered and standardized manner. In the scope of our working group, we recommend defining the Data Trustee Role and the Data Steward Role.
- 2. Identify a consistent and standardized set of resources that are provided by the university to support Data Stewards, and similar roles.
- 3. During the 2022-2023 academic year, create a data access request form for a pilot data domain unit through ServiceNow, to allow exploration and testing of real-world data request scenarios using a ticketing system for data access requests similar to one that is in use at the University of Washington.

#### Data Literacy, Tools, and Training Working Group – Suggested Next Steps

- 1. Develop a university-wide foundational training of data literacy topics with additional, more advanced or specialized training opportunities available based on individual roles and levels of data access.
- 2. In collaboration with the VT HR Talent Development Team, develop and maintain a university portal to house training opportunities. The existing PageUp learning management system could potentially be utilized to enable awareness, increase access, and enable tracking of data literacy training developed and provided by Virginia Tech.

#### Metadata and Data Dictionary Working Group – Suggested Next Step

1. The group recommends that the data dictionary and metadata effort be transitioned to the VT IT Transformation Data Governance Committee for further action. The UDGC working group has taken this topic as far as is possible; at this point a decision needs to be made and resources identified to support the chosen path. The group's assessment is that the iData Cookbook may be the best option for Virginia Tech. The working group has documented benchmarking, surveys, and collaborative discussions with other universities regarding their experience with implementing data dictionaries which can be provided to the IT Transformation governance group for their reference.

In the table below, the suggested next steps are mapped to a matrix illustrating Virginia Tech's readiness and resources for proceeding and implementing each.

Suggested Next Steps	Ready to implement	Some new resources	Requires further study
	using existing	needed	and new funding
	resources		support
Data Access			
1. Develop standardized Data	During routine or		
Trustee and Data Steward role	periodic position		
descriptions	description updates, the		
	verbiage for these roles		
	could be added		
2. Training support for Data	This item spans 2 columns	because there is already	
Trustees and Data Stewards	existing training available; I	however new training is always	
	welcome and helpful to kee	ep skills and knowledge up to	
	date. UDGC data literacy w	orking group next steps	
	address this also.		
3. Develop and test a Service		Select a pilot data domain	
Now data access request		unit and create a "liger	
process		leam" of cross-functional	
		Experts to develop and test	
Data Literary		Service Now request process	
Data Literacy			
1. Aggregate existing, curated	The UDGC working group	New, VT-specific training	
training resources & develop	has identified training	modules may need to be	
new, vi-specific training as	already paid for/ or free	created over time as needs	
needed	to Virginia Tech	arise	
2 Create a controlized online	employees		
2. Create a centralized online			
portal and means of requiring/			collaborate to determine
training			whether Pageup can be
Motadata / Data Dictionary			useu
1 Dilet implementation of iData		There is an estimated east	
Cookbook or other solution		for a data distingant tool in a	
		CNR for IT Transformation	
		recommendation 4.1. If/	
		when these new funds are	
		allocated a data dictionary	
		platform can be	
		implemented	

#### APPENDICES

- A. Full written suggested next steps from each working group
  - i. Data Access, Risk and Compliance
  - ii. Data Literacy, Tools and Training
  - iii. Metadata and Data Dictionary
- B. Full CSG data governance survey results

# **APPENDIX A: Working Group Suggested Next Steps**

- i. Data Access, Risk and Compliance
- ii. Data Literacy, Tools and Training
- iii. Metadata and Data Dictionary

# Suggested Next Steps from the UDGC Data Access, Risk and Compliance Working Group

**Challenge**: While Data Trustee and Data Steward roles are documented in Policy 7100 and its associated <u>Standard</u> <u>for Administrative Data Management</u>, individuals having those responsibilities are often not aware or informed about what those duties entail within the context of their day-to-day jobs. In addition, there is no structured mechanism for holding these roles responsible for the performance of these data-related tasks which are essential to the success of the university's data governance.

**Recommendation 1**: Develop a set of clearly defined university job roles which define a named set of responsibilities in a centralized, standardized manner. Our suggested next steps include defining the Data Trustee Role and the Data Steward Role. Policies, standards, and documents exist covering various duties and requirements for Trustees and Stewards, but they do not clearly define it as a position. The proposed definitions add the following benefits:

- Centrally defined, mandatory language to standardize these roles in position descriptions
- Uniform responsibilities, leading to uniform evaluations across university units
- Definition of onboarding and continuous training requirements to hold the role
- Definition, if necessary, of who many authorize an individual to have the Job Role
- Predefined Job Roles on a position help differentiate for compensation
- Pre-authorization for access to tools, systems, and data sets, or participation in groups, governance, and committees.

The Job Role concept applies across many core university responsibilities. Example Job Roles could include: Fixed Assets Coordinator; Communication Liaison; Records Manager; Student Advisor; On-Call IT Security; System Administrator; Facilities Manager; etc. Example verbiage for job descriptions:

*For Data Trustee Roles*: This position serves as the Data Trustee for \_\_\_\_\_\_\_ data domain as documented in the <u>Standard for Administrative Data Management</u>. Data Trustee responsibilities include assigning and overseeing one or more Data Stewards and identifying training needs and resources for Data Stewards and Data Custodians to ensure high performance with regard to data quality, data security, and data access for administrative data in your purview. Data Trustees are accountable and Data Stewards are responsible for ensuring that your functional area complies with applicable laws, regulations, data security best practices, and policies. Data Trustees may delegate tasks to Data Stewards and others as needed and appropriate; however, Trustees alone serve as the escalation point for issues related to requests for access to data or for issues related to privacy and ethics associated with data in your purview.

Training Requirements as vetted and made available through the UDGC Data Literacy, Tools and Training Working group and others, for example:

• Data Strategy Course (45m): https://www.linkedin.com/learning/data-strategy/why-data-strategy

Authorization: Authorized by and tied to the Vice President level during the definition of a new data domain.

For Data Steward Roles: This position serves as the Data Steward for \_\_\_\_\_

• Work with the relevant Data Trustee and the Information Technology Security Office to assure that data is inventoried and classified by the <u>Virginia Tech Risk Classifications</u> low, moderate, or high risk. Identify procedures for maintaining data confidentiality as they relate to data under the Data Steward's management. As needed, work with the IT Security Office to enforce the procedures.

data

- Assure that there are documented and published processes for granting access to data within assigned data domains and subdomains.
- Review and respond to requests for data based upon legitimate university business objectives that would benefit from the use of requested data.
- Work with the Information Technology Security Office to validate that all systems, including externally hosted systems supporting business processes within the Data Steward's domain or subdomain, conform to Virginia Tech's standards for security and data handling.
- Works with the Information Technology Security Office where needed to evaluate data access or use alternatives and to stay up to date on training and compliance requirements.
- As needed, participate in the management of shared data in ERP systems (such as Banner) supporting the Data Steward's domain or subdomain
- Establish and maintain an appropriate structure and review process for responsible management of the data access lifecycle, including deprovisioning access when appropriate.
- Communicate to stakeholders regarding the enterprise use, policies, and decisions around the stewarded data.

Training Requirements (examples):

- Data Strategy Course (45m): <u>https://www.linkedin.com/learning/data-strategy/why-data-strategy</u>
- Data Stewardship Foundations (1h 47m): <u>https://www.linkedin.com/learning/data-steward-foundations</u>
- Learning Data Governance (1h 24m): <u>https://www.linkedin.com/learning/learning-data-governance-14224082</u>

Authorization: Must be authorized by Data Trustee Role in the same data domain

### **Recommendation 2**

This section contains recommendations of resources provided by the university to support the Data Stewards, Data Experts, and similar roles. Examples would include but are not limited to:

- 1. Develop, centrally publish, and maintain via governance data access approver guidelines and questions to assist stewards with decision making (draft)
- 2. Management should ensure capacity to assess/address (directly or within organization) the IT security aspects associated with a given data set
- 3. Provide support for university-wide Communities of Practice regarding data stewardship, cross-datamart reporting, security, and other topics.
- 4. Replace the <u>Guidelines for Data Stewards</u> with the above job descriptions and question guides.

#### Next Steps for UDGC Data Access Working Group

During AY 2022-2023, the UDGC data access working group should proceed with a pilot project to create a data access request mechanism through Service Now, perhaps modeled on the process being used by the University of Washington, see below. This type of process could incorporate a list of questions for data access requestors and approvers that was developed by this working group in 2022.

#### https://itconnect.uw.edu/work/data/data-security/request-access/

goes to a Service Now request, which takes you to a list of questions designed to determine what University role is required to obtain the types of data being requested:

Support for ana	alytics: Guided access to data		Piet   UW
	Do you need access to HR or Student Data		
	HR and Student	* *	
	*Do you need access to social security numbers (SSN)?		
	No	×	
	*Do you need access to disability info?		
	No	× *	
	*Do you require access to any of the following? Benefits, FMLA, Medical C Race/Ethnicity, or Veteran Status	enditions,	
	*Do you require access to any of the following? Citizenship, Contributions, Deductions, Home Address/Phone, Personnel Actions, or Sex	, Date of Birth,	
	No	× *	
	*Describe your needs with respect to financial aid data		
	None Needed	× *	
	*Do you need access to non-charge related (contains sensitive student acc	count information)?	
	No	× *	
	ASSESS ROLE		W

A matrix showing the different university roles and what types of data they are allowed to access should also be created, for example:

https://s3-us-west-2.amazonaws.com/uw-s3-cdn/wpcontent/uploads/sites/70/2014/12/23214411/DMCAccessRolesMatrix.pdf

# Data Literacy, Tools and Training



Working Group Report

University Data Governance Council 2021-2022

# Table of Contents

Data Literacy Deliverables	2
Data Analysis, Visualization, and Communication Deliverables	5
Overall Deliverable from Working Group	6
Data Literacy, Tools and Training Committee Members	8
Appendix a	9
Appendix b	11

# Data Literacy, Tools and Training

# Data Literacy Deliverables

- Outline for an internally developed, Virginia Tech specific training module.
- Database of free training resources on several topics including data literacy, understanding data, data-informed decision-making, and effective presentation techniques.

## Deliverable 1

Outline for an internally developed, Virginia Tech specific training module. To recommend Virginia Tech specific training for university faculty, Administrative and Professional faculty, and staff members who work with university data.

## • Section 1: Importance of Data Literacy

- Why Virginia Tech cares about data literacy and what does the university expect
- o Why it is important to protect the university data
- What are the responsibilities of an employee to protect data
- How does being data literate benefit one not only as an employee but personally

## • Section 2: Data Literacy Overview

- How does Virginia Tech define data literacy
- What it means to be data literate
- What are some models/examples of data literacy
- Where to go for more information or training [Refer to Deliverable 1b related to Data Literacy for additional resources.]

### • Section 3: Virginia Tech Data Specifics and Security

- o What data is available at the university
- What are data threats, how to keep data secure/data management, and what responsibilities do employees have regarding data security (University resource already available: <u>https://sites.google.com/vt.edu/deptheadtraining/it-security</u>)
- o What ethical considerations are there related to presenting data

## Deliverable 2

Database of free training resources on several topics including data literacy, understanding data, data-informed decision-making, and effective presentation techniques. (See Appendices A and B for additional information on trainings)

## • Data Literacy:

 Data Literacy Group: A Culture of Data Literacy (<u>https://thedataliteracyproject.org/learn</u>)

- LinkedIn Learning:
  - Presenting Data Effectively to Inform and Inspire; Section 1: Developing a Plan (<u>https://www.linkedin.com/learning/presenting-data-effectively-to-inform-and-inspire/develop-your-data-literacy?autoAdvance=true&</u>)
  - Data Fluency: Exploring and Describing Data; Section 1: Think with Data (https://www.linkedin.com/learning/data-fluency-exploring-anddescribing-data/gather-greater-insight-and-make-better-decisions-withyourdata?autoAdvance=true&autoSkip=false&autoplay=true&resume=false&u =57888345)
- Understanding Data:
  - Data Literacy Group:
    - Why Analytics <u>https://thedataliteracyproject.org/learn</u>
    - Understanding Data (<u>https://thedataliteracyproject.org/learn</u>)
    - Understanding Aggregations <u>https://thedataliteracyproject.org/learn</u>
    - Understanding Distributions <u>https://thedataliteracyproject.org/learn</u>
  - LinkedIn Learning: Data Fluency: Exploring and Describing Data (<u>https://www.linkedin.com/learning/data-fluency-exploring-and-describing-data/gather-greater-insight-and-make-better-decisions-with-your-data?autoAdvance=true&autoSkip=false&autoplay=true&resume=false&u=57888 345)
    </u>
    - Section 2: Prepare Data
    - Section 3: Explore Data
    - Section 4: Describe Data
- Data-Informed Decision-Making:
  - Data Literacy Group: Introduction to Data-Informed Decision Making (<u>https://thedataliteracyproject.org/learn</u>)
  - Data Literacy Group: Data-Informed Decision-Making Framework (<u>https://thedataliteracyproject.org/learn</u>)
  - Data Literacy Group: Decision Making Analytic Techniques (<u>https://thedataliteracyproject.org/learn</u>)
- Effective Presentation Techniques:
  - LinkedIn Learning: Presenting Data Effectively to Inform and Inspire (<u>https://www.linkedin.com/learning/presenting-data-effectively-to-inform-and-inspire/develop-your-data-literacy?autoAdvance=true&</u>)
    - Section 2: Story telling
    - Section 3: Putting it in Writing (Presentations)

- Section 4: Pictures
- Section 5: Pre-Presentation
- Section 6: Present
- Overview of Training Resources presented in Appendix A and B:
  - Data Literacy Project (<u>https://thedataliteracyproject.org/about</u>) (see Appendix A)
    - Provides a free eight-module series on data fundamentals with additional modules focusing on data fluency can be purchased
    - Modules are provided in various areas including presenting data, data fluency, data visualizations and storytelling, etc.
    - Committee members reviewed modules to determine if they can be used as part of a Virginia Tech training program.
    - Example from Appendix A: A Culture of Data Literacy
      - Learning areas: What the definition of data literacy is, What a data literacy culture is, How proper culture influences an organization, Steps to ensure your organization has the proper culture in place
      - Time: 29 minutes
      - Audience: Director, Department Head, VP, etc.
      - Committee evaluation: Fairly good quality for specific targeted audience in understanding what a culture of data literacy is and how one can move the organization forward to be more data literate
      - Audience level (beginner, intermediate, advanced): Intermediate
  - LinkedIn Learning (see Appendix B for some resources available on Virginia Tech's Learning Management System):
    - Modules are provided in various areas including presenting data, data fluency, data visualizations and storytelling, etc.
    - Committee members reviewed modules to determine if they can be used as part of a Virginia Tech training program.
    - Example from Appendix B: Presenting Data Effectively to Inform and Inspire
      - Learning areas: Background of data literacy, Keys to communicating data, Types of data analysis, Understanding audience
      - Time: 14 minutes
      - Audience: All; pretty basic for most but provides a solid foundation
      - Committee evaluation: Ranged depending on section viewed. Sections 1 through 4 are very good. Sections 5 and 6 are not bad; just very basic.
  - Other Potential Sources of Training:
    - Online

- Data Literacy @ Johns Hopkins (Coursera)
- <u>Data Literacy: What is it and why does it matter?</u> @ University of Copenhagen (Coursera)
- Professional organizations
  - Data Management Association International (DAMA)
  - International Institute of Business Analysis (IIBA)
  - <u>Association for Computing Machinery (ACM)</u>

# Data Analysis, Visualization, and Communication Deliverables

- Database of data analysis tools and training available through publicly accessible platforms.
- Potential focus areas for training and a collaboration portal for housing information on data analysis and visualization tools.

## Deliverable 1

Database of data analysis tools and training available through publicly accessible platforms.

- Data Analysis Basics: Tools and trainings available by university to analyze data:
  - Tools currently available: Excel, R, SAS, SPSS, Python, Microsoft Power BI, MicroStrategy, and Tableau
  - Trainings currently available through the university for data analysis:
    - Lynda@vt.edu (https://video.vt.edu/media/Lynda.vt.edu+-+A+free+resource+for+Virginia+Tech+faculty%2C+staff%2C+and+students./ 1\_h5v1t03b)
    - https://guides.lib.vt.edu/c.php?g=580714&p=4559973
    - https://lib.vt.edu/research-teaching/data-services.html
    - Software Carpentries Workshops @ University Libraries (Unix, Git, Python, R)
    - <u>Professional Development Network</u> @ TLOS (R, JMP, Data Analysis, etc.)
    - <u>SAIG Short Courses</u> @ Department of Statistics
- Other potential sources of training:
  - Online:
    - <u>https://powerbi.microsoft.com/en-us/learning/</u>
    - LinkedIn Learning example: Data Visualizations: A Lesson and Listen Series [(https://www.linkedin.com/learning/data-visualization-a-lesson-and-listenseries/introduction?autoAdvance=true&autoSkip=false&autoplay=true&res ume=true&u=57888345
    - Become a data analyst path: <u>Become a Data Analyst (linkedin.com)</u>
  - Professional organizations:

- Educause: <u>https://events.educause.edu/educause-institute/data-literacy-institute/</u>
- ACCS: Data integration and analysis SIG: <u>https://accsva.org/sigs</u>

# Deliverable 2

Potential focus areas for training and a collaboration portal for housing information on data analysis and visualization tools.

- Potential focus areas for training:
  - How to effectively visualize and communicate data, data communication techniques, best visualizations for specific types of data, data modeling, and general awareness of how to connect the dots with data
  - The language of the business specific to Virginia Tech, where to access university data and what is available, data governance and policies at Virginia Tech
  - How to use or read interactive reports (PowerBI, Tableau) for executive leaders and relational database basics (for PowerBI, Tableau, MicroStrategy)
  - Reporting university data (reflecting Virginia Tech branding guidelines) and making data dashboard accessible
- Collaboration portal example: <u>https://virginiatech.sharepoint.com/sites/TTL\_Working\_Group/SitePages/Data-</u> <u>Literacy.aspx</u>

# Overall Deliverable from Working Group

Three possibilities as to how the university can define roles or levels of expertise for different data literacy, data analysis, and visualization training.

# Possibility 1

Roles based on University policy 7100 - <u>Administrative Data Management and Access Policy</u>; Section 2.1 Data Management Roles and Responsibilities

- Data Trustees: Senior university officials (typically at the level of Vice President, Vice Provost, or Dean) who have planning and policy-making responsibilities for university data.
- Data Stewards: University officials (typically at the level of Associate Vice President, Associate Vice Provost, University Registrar, University Bursar, or Director) who oversee the capture, maintenance and dissemination of data for a particular operation.
- Data Managers: University staff in a functional area with day-to-day responsibilities for the capture, maintenance, and dissemination of data for a particular operation.
- Data Users: Individuals who access university data in order to perform their assigned duties or to fulfill their role in the university community.
- University Information Technology Security Officer: The university official responsible for maintaining a plan for security policies and practices and for keeping abreast of security-

related issues internally within the university community and externally throughout the information technology marketplace.

# Possibility 2

Roles based on job titles or other criteria

- Have at least some small trainings for all teaching, research, and A/P faculty
- Different training paths for people with different job types
- Identifying people dealing with different databases
- Groups of possible trainings by role; people could be required to do a certain number people decide what they need
- Example: Administrative data users at Virginia Tech
  - Leadership / Executive Level
  - Director / Manager Level
  - Data Analyst (Person asked to pull together data for decision making)

# Possibility 3

Self-evaluation of expertise level

- Recommending a short course with embedded questions guiding individuals to appropriate training opportunities. Training opportunity links will be provided upon completion.
- Developing a matrix where individuals identify their perceived skillset and the matrix guides them to either required or optional training.

# Data Literacy, Tools and Training Committee Members

Arsenault, Rhonda Senior Associate VP for Advancement and COO, Advancement Division
Austin, Janice Assistant Dean and Director of Graduate Admissions Dean, Graduate School
Bliss, Marie Assistant VP for Human Resources Administration
Bodo, Bethany Director, Institutional Effectiveness, Office of Analytics and Institutional
Effectiveness
Cooper, AI Director Business Management Systems, President's Office
Gursoy, Hikmet Director, Student Affairs IT
Ogier, Andi Assistant Dean & Director, Data Services Library
Potter, Mary Director, Privacy and Research Data Protection, Scholarly Integrity and Research
Compliance

Teglas, Jon Clark Chief of Staff to VP Campus Planning, Infrastructure, and Facilities

# Appendix a

# Data Literacy Project

# https://thedataliteracyproject.org/about

# Beginner Level

Name of the session	Short description of the session	Time commitment	Best target audience
Why Analytics	Understand what analytics is Why it is important in the world today Understand how it is applied in real world situations Introduction into how Qlik is here to help in the analytical journey	11 minutes	Employee who has no significant background in using data
Understanding Data	Gain an understanding of data Learn about the different types and attributes of data Understand why learning about data is important in the world of analytics today	25 minutes	Employee who has no to low background in using data
Understanding Aggregations	Learn what data and analytical aggregations are Learn about specific aggregation types like mean and median Understand why aggregations matter in the world of analytics today	22 minutes	Employees who do not have a background in basic statistical analyses and definitions
Understanding Distributions	Understand the theory of distributions Learn the different types and characteristics of distributions Learn the importance to the world of analytics Explore standard deviation	22 minutes	Employee who has no to low background in basic statistical data analysis and visualization

# Appendix a (continued) Data Literacy Project

# https://thedataliteracyproject.org/about

# Intermediate Level

Name of the session	Short description of the session	Time commitment	Best target audience
Into to Data-Informed Decision Making	Explain what Data-Informed Decision Making is Explore how organizations can turn data into value Gain an understanding of skills and competencies required to properly leverage data	32 minutes	Director, Department Head, VP, etc.
Data-Informed Decision-Making Framework	Explain the steps to take when making a Data-Informed Decision Explore various modules and tools that can help with the decision-making process	39 minutes	Director, Department Head, VP, etc.
Decision Making Analytic Techniques	Understand various types of analytics available to decision makers Identify use cases where one or more of these techniques can be applied	35 minutes	Director, Department Head, VP, etc.
A Culture of Data Literacy	What the definition of data literacy is What a data literacy culture is How proper culture influences an organization Steps to ensure your organization has the proper culture in place	29 minutes	Director, Department Head, VP, etc.

# Appendix b LinkedIn Learning Presenting Data Effectively to Inform and Inspire

#### https://www.linkedin.com/learning/presenting-data-effectively-to-inform-and-inspire/develop-your-dataliteracy?autoAdvance=true&autoSkip=false&autoplay=true&resume=false&u=57888345

**Description:** "We all deal with data, and at some point, we all need to be able to present it effectively to others. But how can you really make an impact? How can you make sure you're conveying the information that matters most, and that you're going to inspire your audience to action? In this course, instructor Bill Shander walks through how to plan for a data presentation: finding the story in your data, focusing on the "so what", defining and reaching your audience, using visuals most effectively, editing your presentation, and presenting your data to an audience. Bill covers the whole gamut, from concept to execution, even including some presentation skills training." (citation link)

Name of the session	Time commitment	Best target audience and audience level (i.e., beginner, intermediate, advanced)	Training category
Section 1: Plan	Total time for all sections: approximately 97 minutes	Section 1: People with very little experience with data; Beginner	Section 1: Defining data literacy; data literacy basics
Section 2: Storytelling	Section 1: Approx. 11 minutes	Section 2: People that use data and need to present it; Beginner	Section 2: Storytelling; creating a story with the data
Section 3: Putting it in writing (presentations)	Section 2: Approx. 19 minutes	Section 3: People that need to present data; All	Section 3: Presenting data; making slides
Section 4: Pictures	Section 3: Approx. 12 minutes	Section 4: People who	Section 4: Presenting data effectively; using the right visuals
Section 5: Pre-	minutes	Beginner	Section 5: Presentation basics; how to present
	Section 5: Approx. 9 minutes	Section 5: People new to presenting; Beginner	Section 6: Presentation basics; audience and
Section 6: Present	Section 6: Approx. 16 minutes	Section 6: People new to presenting; Beginner	methodology

### Appendix b (continued) LinkedIn Learning Data Fluency: Exploring and Describing Data

#### https://www.linkedin.com/learning/data-fluency-exploring-and-describing-data-2019/gathergreater-insight-and-make-better-decisions-with-yourdata?resume=false&u=57888345&autoplay=true&autoAdvance=true&autoSkip=false

**Description:** "Data analysis isn't just for specialists who need to make sense of massive datasets. Decision-makers in every industry can benefit from a basic understanding of the goals and concepts of applied data analysis. In this course, join Barton Poulson as he focuses on the fundamentals of data fluency, or the ability to work with data to extract insights and determine your next steps. Barton shows how exploring data with graphs and describing data with statistics can help you reach your goals and make better decisions. Instead of focusing on particular tools, he concentrates on general procedures that can help you solve specific problems. Find out how to prepare data, explore it visually, and use statistical methods to describe it." (citation link)

Name of the session	Time commitment	Best target audience and audience level (i.e., beginner, intermediate, advanced)	Training category
This LinkedIn Learning contains quizzes at the end of each section Introduction	Total time for all sections: approximately 4 hours, 19 minutes	Section 1: Solid foundation/overview of data literacy; Could be foundation for all but definitely good for beginners	Section 1: Overview of data literacy, different levels of fluency, when to use data to answer questions, how to frame questions
Section 1: Think with data Section 2: Prepare data	Introduction: 1 min. 30 sec. Section 1: Approx. 60 minutes	Section 2: Good general foundation on ethics and data information; Could be foundation for all but definitely good for beginners	Section 2: Covers data ethics, in- house data, talk about PII, open data sources, gathering data, third party data, data quality, generalizability, data meaning
Section 3: Explore data Section 4: Describe data	Section 2: Approx. 78 minutes Section 3: Approx. 68 minutes	Section 3: Pretty much anyone who is trying to figure out what kind of visualization works best with your data Section 4: This section is related to	Section 3: Data Visualizations; Good overview on different types of visualizations like bar charts, pie charts, etc.
Next steps section at the end on other areas that someone would want to examine	Section 4: Approx. 56 minutes	working with data	variability, effect sizes, correlations and regressions

### Appendix b (continued) LinkedIn Learning Data Visualizations: A Lesson and Listen Series

## https://www.linkedin.com/learning/data-visualization-a-lesson-and-listen-

#### series/introduction?autoAdvance=true&autoSkip=false&autoplay=true&resume=true&u=578883

45

Time commitment [Total time is about 11 hours]	
Introduction [4 minutes]	Daily practice [3 minutes]
Creative and unique visualizations (9 minutes]	Data visualization at scale in industry [5 minutes]
• Visualization in the real world [8 minutes]	Concept visualization [6 minutes]
Data storytelling for regular folks [12 minutes]	Data visualization research [5 minutes]
Historical visualizations [4 minutes]	Visualizing complex and controversial topics [4 minutes]
Visualizing large data sets [6 minutes]	Color in data visualizations [4 minutes]
Unicorns [5 minutes]	Visualizations in excel [4 minutes]
Text analysis and visualizations [8 minutes]	Science visualizations [4 minutes]
Maps [5 minutes]	Data art [4 minutes]
• Visualization for a cause [6 minutes]	Making decisions with data viz [ 6 minutes]
It's all in the details [4 minutes]	Visualizing health and healthcare data [4 minutes]
Becoming an informed designer [5 minutes]	Diversity and inclusion in data viz [5 minutes]
Data literacy [3 minutes]	• Data visualization tools of the future [3 minutes]

#### Overview of Two Potential Metadata/Data Dictionary Platforms: iData and Collibra

For academic year 2021-22, the objectives of the metadata/data dictionary working group of the University Data Governance Council were:

- 1. Review potential metadata/data dictionary platforms;
- 2. Query other universities for insights relating to these platforms and critical elements for successful implementation, and;
- 3. Summarize our findings in an informational report for the UDGC.

Most universities we queried were either using iData or had no tool in place. Therefore, we focused our efforts on iData but also include some content on Collibra, which offered an onsite demo to VT approximately four years ago.

#### iData - Data Cookbook - http://www.idatainc.com/

Several members of the working group attended a webinar by iData in Fall 2021. A copy of the webinar presentation slide deck is available on the Google Drive: *Achieve Data Governance with the Data Cookbook* – *Webinar and Demonstration\_V11.pdf*.

#### Summative Points of Webinar:

- 1. Offers a comprehensive framework for data governance by consolidating governance content, establishing workflows, allowing impact analyses and addressing facets of stewardship, users, information management, and processes.
- 2. Three editions available depending on needs:
  - a. Knowledge Edition Business glossary only.
  - Enterprise Plus Edition Business glossary with data models, metadata framework, data lineage analysis, data quality modules, integration with various reporting tools, data governance workflows, etc.
  - c. +Knowledge Edition Business glossary with partial integration for data processing, governance workflows, etc.
- Banner specific implementation is possible where the Banner database is already documented. However, VT has customized our version of Banner. Therefore, many of tables are custom built and owned by VT so iData has no supporting framework.
- 4. Connections to additional databases such as warehouse and third-party databases would incur higher cost and more complex challenges. Moreover, the technical support for these connections would be less well developed or not available.
- 5. Data quality monitoring and reference data system (once set up) appears robust.

- 6. Contains option to establish a ticket system for data requests and quality issues.
- 7. To some degree, successful implementation assumes data dictionary/metadata development is already in some stage of completion/maturity.
- 8. Allows embedding into various systems, tools and websites via API Integration.
- 9. Data governance services offered from iData include assessment, roadmap, training, consulting, etc.
- 10. Implementation estimated at 9-12 months by iData.

#### Summative Points of Peer Discussion:

- 1. Successful (partial) rollouts typically ranged 1-2 years. Dependent on product.
- 2. In most instances, a single individual was appointed to develop a plan and oversee the rollout as a core responsibility. Further, special positions (e.g. Data Governance Manager) were created to provide more involved functional/technical support and handle the monitoring and support of data governance. Cross functional teams also seemed to be critical to a robust implementation.
- 3. The Data Cookbook product was often purchased to provide a catalyst and consistent framework for data stewards and functional areas to create data glossaries and to respond to data classification requirements.
- 4. A few schools started their data governance initiatives using Collibra but switched to iData due to financial constraints.
- 5. User training, especially those users who are part of the main workflows, is an absolute requisite for success and often involved several sessions over a 6 9 month period.
- 6. The most significant challenge cited was the sheer size of the lift for users/functional areas to get data identified and entered into the system. As an additional responsibility (without at least some dedicated support personnel), the initiatives proved especially burdensome and fraught with setbacks.
- 7. A second notable challenge was the variance in the data steward's comfort with the technical aspects of this exercise. Many functional areas require significant ongoing assistance.
- 8. A third common challenge is determining which data steward has approval power for certain data/definitions that are common across multiple areas of the institution and defining data lineages and responsibilities for quality assurance/control.
- 9. A fourth common challenge mentioned is iData's largely passive and manual process of metadata management and the workload on data engineering teams.

### Collibra - https://marketplace.collibra.com/

Approximately four years ago, Collibra provided an onsite demo for Virginia Tech. The link to that demo is: <u>https://video.vt.edu/edit/0\_hc7cvne0</u>.

#### Summative Points of Demo, Internal (VT) Discussion, and Peer Discussion:

- 1. A full-time, expert position is required for the managing the application.
- 2. Implementation can be expensive and time consuming. Often requiring much consulting before operational.
- 3. The product has a very high level of functionality. Institutions often mentioned the product offered more than they could practically implement and use.
- 4. Where implemented, an entire team (numbers not disclosed) was specially assigned to ensure implementation was successful.
- 5. The rate of product enhancement is high, which is desirable for a rapidly changing data environment.
- 6. An economical metadata ingestion licensing model (licensing is based on technology platforms, such that Oracle counts as one license regardless of number of databases—versus licensing for some based on each application or system ingested).
- 7. User-friendly interface to aid buy-in and enhance usage on campus.
- 8. Greater mobile device support options.
- 9. Good product support reputation.
- 10. An excellent data quality module with the addition of OwIDQ (a predictive data quality vendor).
- 11. Robust data lineage capabilities.
- 12. Solid crowdsourcing framework (for collective and cross-functional implementation and management).

Other products/comments from peers (EDUCAUSE Data Governance Community Poll):

- Many institutions have no form of structured data governance in place. Most that do are using options available in the information systems being used or maintaining Excel/Access databases within functional areas.
- Other platforms mentioned:
  - Azure Purview
  - o Informatica
  - Oracle (free tools that are sufficient for technical analysts).
- The return on investment on creating data dictionaries was often brought into question. Particularly as it related to non-technical users.

#### Summary

**Challenge**: An abbreviated survey completed by this working group in FY 2021 of select functional areas revealed that Virginia Tech's approach to generating data data definitions and metadata standards are highly disjunctive and processes (where they exist) are non-standard across data domains. In order to properly advance data governance and data intelligence, the university must identify the appropriate solution/platform and develop a roadmap for implementation with clear, actionable goals to ultimately enable advanced analytics and data access capabilities. To assist with decision-making this working group completed in FY 2022 an informal survey and assessment of data dictionary/metadata tools being used by our higher education peers, as well as produced a high level summary of the two most common products used.

**Recommendation**: We recommend that the data dictionary and metadata efforts be transitioned to the VT IT Transformation Data Governance Committee for further action. An administration decision must be made on the proper data governance solution to advance to the next phase of implementation. Based on the information gathered, this group's assessment is that the iData: Data Cookbook may be the best option for Virginia Tech.

**Next Steps:** The UDGC Data Dictionary and Metadata group will not be continued for AY 2022-2023. In addition to the group's summary report, members of the IT Transformation IT Governance – Data Governance Committee may wish to more closely review iData: Data Cookbook by watching the free webinars the company provides, here: <u>https://www.datacookbook.com/recorded-webinars</u>

Furthermore, the committee may wish to consult with points of contact at other universities who have implemented this product, available from the UDGC working group documentation.

# APPENDIX B: Common Solutions Group (CSG) Workshop on Data Governance Survey May 2022

See the <u>full text of the survey</u>.

Out of 21 total survey responses, the following institutions self-identified:

- > Brown University
- > Carnegie Mellon University
- > Columbia University
- > Cornell University
- > Duke University
- > Emory University
- > Georgetown University
- > Harvard University
- > MIT
- > New York University
- > Northwestern University
- > Princeton University

- > University of Iowa
- > University of Michigan
- > University of Nebraska
- > University of Notre Dame
- > University of Washington
- > University of Wisconsin Madison
- Virginia Polytechnic Institute and State University
- > Yale University

# Survey: Drivers for Data Governance

Are the following topics relevant to data governance at your institution?

	Not a concern	lt's a driver	Actively being addressed	Already taken care of
Meeting compliance requirements	0	7	7	7
Governance of data warehousing	2	4	8	7
Responding to emergencies (e.g., COVID)	2	5	8	6
Security or privacy issues	0	7	11	3
Enabling data analytics	0	8	9	4
Enabling institutional views/uses of data	0	9	9	3
Enabling discovery & understanding of data	1	12	6	2
Data quality problems	0	14	6	1
Ensuring easier access to data	1	13	6	1
Increasing data literacy	3	9	7	1

# Survey: Scope of Data Governance

	Not in scope	Plan to govern	Some influence	Actively governed
Data Security	2	2	7	8
Data Warehousing & BI	2	3	7	7
Data Integration & Interop.	1	7	8	3
Metadata Management	3	6	5	5
Data Quality	4	5	6	4
Data Architecture	5	3	8	3
Data Storage & Operations	6	2	9	2
Reference & Master Data	3	7	6	2
Data Modeling & Design	9	4	5	1
Document & Content Mgt.	12	2	3	2

Does data governance at your institution govern:

# Survey: Organization of Data Governance

# Does data governance at your institution include (check all that apply): 21 responses



# Survey: Tools

	Not considering	Investigating	Implementing	In use
Dictionary/catalog/glossary	2	5	3	9
Data access management	1	6	2	9
Metadata management	4	7	3	5
Master/core data mgt.	7	3	4	4
Automated data retention	7	4	2	4
Data marketplace	8	5	1	4
Data lineage/Impact analysis	9	7	1	2

Indicate if your institution uses software tools/platforms for functionality related to data governance:

# **Survey: Products**

Does your institution use any of the following in relation to data governance:



Footnote to the survey question above: Like many universities, Virginia Tech is currently using one or more home grown solutions for data dictionary and metadata. The second most commonly used is the Data Cookbook -- our working group has made contact with several universities that are currently using that platform and are very satisfied with it for their higher education data definitions.

# Survey: Resourcing Data Governance

#### How is data governance resourced? (check all that apply)

21 responses



# Survey: Decision-Making About Data

In general, does decision-making about data at your institution include: (Central policies, standards, strategies, or plans)



# Survey: Impact of Data Governance

	No real change	Some improvement	Good progress
Security or privacy issues	3	8	7
Enabling institutional views/uses of data	3	8	7
Enabling data analytics	4	7	7
Ensuring easier access to data	4	8	6
Responding to emergencies (e.g., COVID)	6	4	8
Governance of data warehousing	6	5	7
Meeting compliance requirements	6	7	5
Enabling discovery & understanding of data	5	11	2
Data quality problems	6	10	2
Increasing data literacy	9	8	1

As a result of data governance, has there been improvement with regard to these drivers?

# Survey: Data Literacy Training

# Is there data literacy training at your institution? (check all that apply)

21 responses



# Survey: Distribution of Enterprise Data

How is commonly used institutional/enterprise data typically made available at your institution? (check all that apply)

21 responses



# Footnote to the survey question above: Insights from Brent Kremer, Director of Data Management and Analytics, Enterprise Systems, Division of IT about ways in which Virginia Tech data experts typically access enterprise data

The solution typically varies depending on the data domain. Finance, HR, Sponsored Programs, Foundation are very mature where student is much more decentralized and less mature.

**Direct access to the system of record** – Most users (maybe 90%) that need access to data in an application will have some sort of permission to that app where they will be able to run reports within the application. They work directly with data stewards to get that access.

**Central/enterprise database access** – we have very few people who access to the system of record databases... but we do have a lot of people who access the data warehouse directly. We have a mature and robust access policy that gives people access to the data warehouse. There are several tools (PowerBI, Tableau, Qlik, Microsoft Excel/Access, etc) that connect to the warehouse and pull their own reports from the warehouse. Central IT does not have control of those apps so we don't know all the apps used or who is using them. We do know about 500 people do this. Many of those people probably don't know they are doing this, they are using an app created by a decentralized IT person.

**Central/enterprise APIs** – most of our data is not behind any API. We have not done more of this because partners are not asking. In theory we can use MicroStrategy to do this but VTTI is the only grouped that asked. We have setup a few Oracle APIs for specific purposes coming out of banner. For example one group wanted to check a fund balance before allowing people to charge a service they provided to it.

#### Flat file exports from System of Record - very few

#### Flat file data feeds - very few

**Reports** – most people doing analytics get the majority of their data this way. They use a variety of systems to get it.

- MicroStrategy has 1,300 active monthly users who run 200k+ reports each month.
- Web Reports is integrated with banner and uses legacy scripting tools. In many cases people think this is reporting within banner.
  - Student has reporting tools but many people still go through a centralized request process as users have a hard time figuring out how to run those tools. This is because of the type of users. We don't have this same issues with finance, HR or other domains.
  - I can't tell the difference between users from web reports and users from PowerBI. So that earlier
     500 number is shared between these different methods.
- Slate and Blackbaud have very robust warehouse / analytics functionality that is vendor owned.

When doing a complex analysis people will use multiple systems to get the data then use either excel or MicroStrategy to combine it in a single place.