Data Governance Council 2022-23 Annual Report

September 14, 2023

Members 2022-23

Kim Berlin, Chair
Diane Bello, Co-Chair
Andrei Antonenko
Ahmed Belazi
Braden Hosch
Cassandra Amadio
David Cyrille
Dawn Medley
Jim Gonzales
Lyle Gomes
Nicholas Prewett
Paula Di Pasquale-Alvarez
Robert Davidson
Sara Lehmann
Susan Agro
Tracey McEachern
Theresa Diemer, *ex officio*
Arielle Markiewicz, *ex officio*

Overview

Stony Brook's data governance system was established in fall 2016 in order to improve Stony Brook's data infrastructure. The Data Governance Council (DGC) oversees the data governance system and began meeting in spring 2017. This is the sixth annual report of the DGC.

Major Accomplishments

- Expanded DGC membership and identified new executive sponsors
- Adopted a new University data access policy
- Completed a pilot project to permanently delete obsolete data
- Filled a new position for a Data Governance Business Analyst
- Continued implementation of the Data Cookbook digital definitions
- Conducted the biennial maturity model assessment

Executive Sponsors and Expansion of the Data Governance Council membership

The two Executive Sponsors for the DGC had been filled by the Senior Vice President & Chief Information Officer and the Vice President of Strategic Initiatives. Both of these positions were vacated during 2022-23 and prompted the DGC to identify new sponsorship. One of the executive sponsors starting 2023-24 will be the incoming Vice President for Information Technology and Chief Information Officer. Pending a formal vote at the next meeting, we have identified the Vice President for Educational & Institutional Effectiveness as the second executive sponsor.

As priorities of the DGC have been moving more toward topics of data risk management and data policy, we considered the membership of the DGC and recognized the need to include VP areas that were not currently represented and who would provide valuable input into these conversations. Three areas that will be added in the 2023-24 year are Vice President for Enterprise Risk Management and Chief Security Officer, Equity and Inclusion, and Campus Planning & Facilities.

Data Governance Framework has been drafted to adjust the executive sponsors and to expand membership of the DGC and to set the DGC membership to VP areas rather than titles. The Stony Brook University Data Governance website will continue to retain the names and titles of council members. https://www.stonybrook.edu/commcms/irpe/about/data_governance/data_governance_council

Adoption of a new Policy on Data and Data Access

The DGC empaneled a committee to review Stony Brook's data access policy and to propose revisions to replace outdated DoIT policies that had been in place since 2007. The committee was composed of Braden Hosch, VP Educational & Institutional Effectiveness; Matthew Nappi, Chief Information Security Officer; Marissa Trachtenberg, Director of Risk Management & Policy Compliance; Dara Goldstein, Chief Privacy Officer, Stony Brook Medicine; and Doug Panico, Assistant Vice President, Audit & Management Advisory Services.

The committee delivered a new University policy that promotes accountability over data assets, responsibilities of data stewards and data custodians, and distinguishes procedures for different types of access. Data Trustees are executive leaders that oversee broad data domains and ultimate responsibility for the data assets in their purview. Data Stewards and Data Custodians are responsible for developing written procedures for how data access is granted for routine access and non-routine access, how

permissions are given for special privileges, how data access requests are evaluated, (such as having a legitimate university business, or advancing mission), the re-release of data, and for how users are routinely removed and reviewed. Data stewards have 6 months from the policy adoption date to fulfill this request and data custodians have 9 months. The collection of this information started in Spring 2023 through a new questionnaire called the "Data Access Management Procedures". This data collection is expected to be ongoing through December 23, 2023.

The policy was approved and adopted by the VP Council on March 23, 2023. It will be reviewed every three years.

The "Policy on Data and Data Access" can be found on the university "Policies" webpage: https://it.stonybrook.edu/policies.

Data asset inventory

The number of data assets in the Data Asset Inventory is the same as last year. The University, excluding Health Systems, maintains 95 data assets. In 2022-23, one new data asset was added, and one was retired. Additional information about data assets is now being collected to comply with the requirements of the new Policy on Data and Data Access.

There are 13 Data Trustees identified and listed below by the number of data assets in their purview.

Data Trustees	# Data Assets
Provost	23
Vice President Finance & Administration	23
Chief Information Officer	13
Vice President Student Affairs	9
Vice President for Educational and Institutional Effectiveness	7
Vice President for Research	4
Vice President of Human Resource Services	3
Executive Vice President Health Sciences	3
Vice President for Advancement	3
Director of Athletics	2
Vice President for Enterprise Risk Management and Chief Security Officer	2
Vice President for Facilities and Services	2
Vice President for Marketing and Communications	1
Grand Total	**95**

Data trustees and data steward organizations are identified for all University data assets (Appendix B). The Data Asset Inventory contains information about basic contents, storage location, sensitivity level, data acquisition, data integration, linkage data access, reporting and the availability and type of analytics (see Appendix C).

Data Disposition

A pilot project was successfully conducted this year to permanently delete obsolete data and to begin a repeatable, annual process. This ongoing process will reduce risk to the university by removing sensitive data that has no practical value to the university. The Financial Aid Office, led by Nicholas Prewitt, worked with the Information Technology Enterprise Applications & Integration team led by Jim Gonzalez to identify financial aid Institutional Student Information Record (ISIR) records for past applicants who never attended the university. The teams identified ~84,000 records between 2003-2016 that contain

applicants of the university who never attended and have no financial records. The process to delete these data was more manual than automated but was written and stored in a way to be repeatable once a year. The Financial Aid office plans to repeat the purge each January for the following year of data (next will be 2017). It is expected that approximately 5,000-7,000 ISIR records will be deleted each year.

A DGC priority for 2023-24 is planned to scope another area for data disposition. Two areas being considered are application data for prospective students who did not enroll, and application data for prospective employees who were not hired.

Business Analyst for Data Governance

The first full time position dedicated to Data Governance activities was filled this year. The new position of Business Analyst for Data Governance was filled by Arielle Markiewicz in January 2023. She reports to the Assistant Director for Data Governance and Management, Theresa Diemer, in the Office of Institutional Research, Planning & Effectiveness. This new position is an important role for the long-term support and effective continuation of data governance activities and is making an impact to advance the priorities and initiatives of the DGC.

Continued implementation of the Data Cookbook

The Data Cookbook was acquired at the end of 2017-18. This metadata management tool provides a digital repository for data definitions and other metadata that will be integrated with existing reporting and analytics to improve the understanding about the data. The new Data Governance Business Analyst, Arielle Markiewicz, has greatly advanced this work and added 59 definitions for a total of 198 and added a number of new report profiles for a total of 68. Reference data was added by the Enterprise Data & Analytics office to more than 75 definitions. Reference data provides data profiles on allowable values, data types and ranges. It is planned in 2023-24 to make this information available to end users when viewing reports and analytics in our effort to improve communication and understanding about data.

Maturity Assessment 2023

Every two years we conduct an assessment of the maturity level of our data governance topics across the most commonly used data domains. The data governance maturity assessment considered data domains along five key topics:

- Organizational Structures
- Roles and Responsibilities
- Data Quality
- Communication
- Culture

The most common data domains, financial, student/academic, human resources, research, and facilities are provided in the assessment and respondents also had the opportunity to rate "other" data domains.

Maturity was rated on a 5-point scale (1-informal, 2-developing, 3-adopted & implemented, 4-managed & repeatable, 5-integrated and optimized). Members of the DGC, data stewards of assets in the university data asset inventory, and selected university leaders with titles such as Assistant/Associate Vice President/Provost were asked to respond. We received 28 responses out of 95, or about 30% participation in this years survey.

Maturity Assessment 2017-2023

The maturity level between 2017-2023 has progressed from the range of "developing" stage (2.0-2.9) to "adopted & implemented" stage (3.0-3.4) for most of the data domains and most of the data governance topics. Since we initiated Data Governance, "Organizational Structures" are much more established with

the advent of the DGC, executive sponsors, a newly dedicated data governance position, and the naming of data trustees, data stewards and data custodians for University data assets. The new policy for data and data access supports the understanding of what the "Roles & Responsibilities" are for those who oversee a data domain. "Communication" tends to be the lowest rated and getting information about data systems, policies, definitions circulated is a challenge. The Financial data shows a steady rise each year but other data domains show a slight decline from 2021. This could be a number of reasons – changes in personnel, less participation due to lack of focus and bandwidth and possibly anomalous results in 2021 that set the stage too high. Despite the slight declines, much progress has been made in several areas and we feel the overall direction continues to improve.
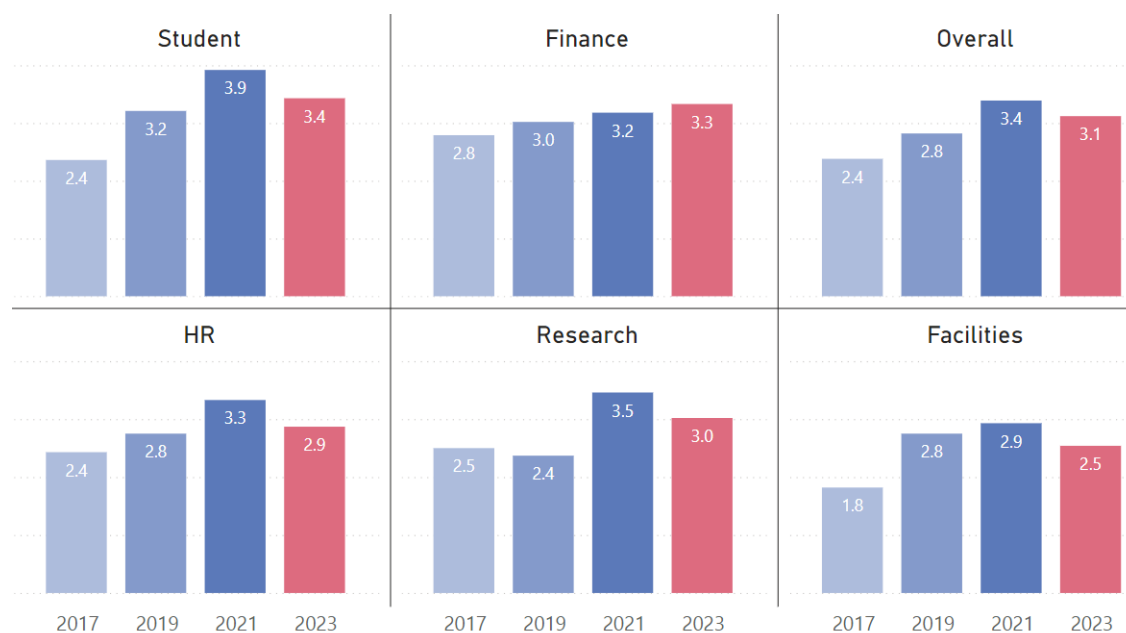
Data Governance Topics across all data domains 2017-2023
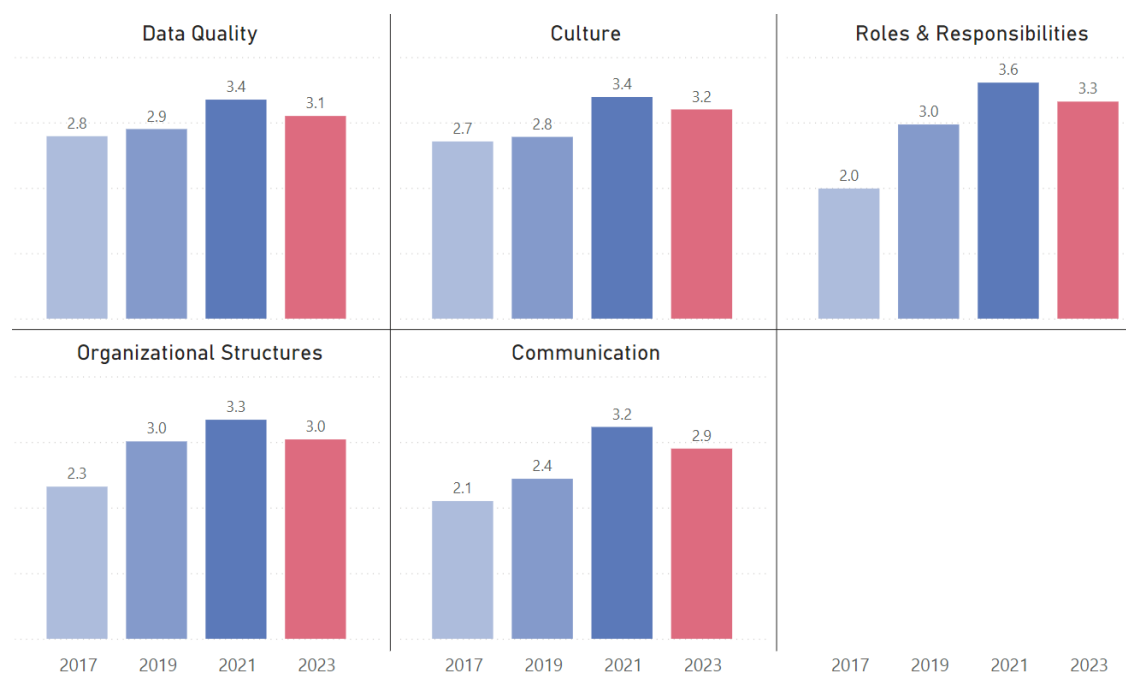
The Table below shows the change in the average and median responses between 2021 and 2023 for each data governance topic and across each of the data domains. The number of responses for 2023 is shown.

Areas and Topics	Mean Maturity Level					Median Maturity Level					n
	2017	2019	2021	2023	Change 2021-23	2017	2019	2021	2023	Change 2021-23	
Financial Data	**2.79**	**3.02**	**3.18**	**3.33**	0.05	**3.0**	**3.0**	**3.0**	**3.0**	0.0	**16**
Communication	2.50	2.50	3.00	3.07	0.02	2.5	3.0	3.0	3.0	0.0	
Culture	3.25	3.20	3.21	3.47	0.08	3.0	3.0	3.0	4.0	0.3	
Organizational Structures	2.67	3.25	3.09	3.25	0.05	2.0	3.5	3.0	3.0	0.0	
Data Quality	3.43	3.09	3.15	3.33	0.06	3.0	3.0	3.0	3.0	0.0	
Roles & Responsibilities	2.13	3.08	3.43	3.50	0.02	2.0	3.0	3.0	3.5	0.2	
Student Data	**2.36**	**3.21**	**3.92**	**3.43**	-0.13	**2.0**	**3.0**	**4.0**	**4.0**	0.0	**19**
Communication	2.00	2.71	3.91	3.39	-0.13	2.0	3.0	4.0	4.0	0.0	
Culture	2.88	3.38	4.00	3.44	-0.14	3.0	3.0	4.0	4.0	0.0	
Organizational Structures	2.22	3.46	3.80	3.37	-0.11	2.0	4.0	4.0	3.5	-0.1	
Data Quality	2.71	3.15	3.95	3.35	-0.15	3.0	3.0	4.0	3.0	-0.3	
Roles & Responsibilities	2.00	3.36	3.96	3.61	-0.09	2.0	3.5	4.0	4.0	0.0	
Human Resources	**2.43**	**2.75**	**3.33**	**2.87**	-0.14	**2.0**	**3.0**	**3.0**	**3.0**	0.0	**15**
Communication	2.20	2.25	3.16	2.43	-0.23	2.0	2.0	3.0	2.5	-0.2	
Culture	2.83	2.82	3.44	3.07	-0.11	2.5	3.0	4.0	3.0	-0.3	
Organizational Structures	2.29	3.09	3.25	2.87	-0.12	2.0	3.0	3.0	3.0	0.0	
Data Quality	2.83	2.67	3.11	2.85	-0.08	2.5	3.0	3.0	3.0	0.0	
Roles & Responsibilities	2.00	2.91	3.70	3.15	-0.15	2.0	3.0	4.0	3.0	-0.3	
Research	**2.50**	**2.37**	**3.46**	**3.02**	-0.13	**3.0**	**2.0**	**3.5**	**3.0**	-0.1	**10**
Communication	2.20	2.17	3.08	3.00	-0.03	2.0	2.0	3.0	3.0	0.0	
Culture	2.80	2.00	3.36	3.10	-0.08	3.0	2.0	3.0	3.0	0.0	
Organizational Structures	2.71	2.60	3.60	2.90	-0.19	3.0	3.0	4.0	3.0	-0.3	
Data Quality	2.80	2.60	3.42	3.00	-0.12	2.0	3.0	3.5	3.0	-0.1	
Roles & Responsibilities	2.00	2.50	3.79	3.11	-0.18	2.0	2.5	4.0	3.0	-0.3	
Facilities	**1.82**	**2.75**	**2.93**	**2.54**	-0.13	**2.0**	**2.0**	**3.0**	**2.0**	-0.3	**8**
Communication	1.60	2.57	2.83	2.38	-0.16	2.0	2.0	3.0	2.5	-0.2	
Culture	1.80	2.50	2.64	2.63	0.00	2.0	2.0	2.0	2.0	0.0	
Organizational Structures	1.71	2.67	3.00	2.38	-0.21	2.0	2.0	3.0	2.0	-0.3	
Data Quality	2.20	3.00	3.08	2.63	-0.15	2.0	2.5	3.0	2.5	-0.2	
Roles & Responsibilities	1.80	3.00	3.08	2.71	-0.12	2.0	3.0	3.0	3.0	0.0	
Overall	**2.38**	**2.82**	**3.39**	**3.12**	-0.08	**2.0**	**3.0**	**3.0**	**3.0**	0.0	**28**
Communication	2.10	2.44	3.23	2.90	-0.10	2.0	3.0	3.0	3.0	0.0	
Culture	2.71	2.78	3.39	3.20	-0.06	3.0	3.0	3.0	3.0	0.0	
Organizational Structures	2.32	3.01	3.34	3.04	-0.09	2.0	3.0	3.0	3.0	0.0	
Data Quality	2.79	2.90	3.35	3.10	-0.07	2.0	2.0	3.0	3.0	0.0	
Roles & Responsibilities	1.99	2.97	3.61	3.32	-0.08	2.0	3.0	4.0	3.0	-0.3	

(1-informal, 2-developing, 3-adopted & implemented, 4-managed & repeatable, 5-integrated and optimized)

Appendix A – Progress on 2022-23 Project Priorities

In September 2022, the Data Governance Council identified and ranked a set of priorities for the 2022-23 year. Below is the list of priorities, notes and status if it was initiated this year

#	Priority 2022-23	Type	Notes	Status
1	**Data purging pilot Financial Aid ISIR records**	Policy/Project Issues	Pilot data purging for Financial Aid ISIR data. Applicants and associated data for those who never attended SBU were permanently deleted. Data purge will be repeated annually.	Complete
2	**Data Access Policy**	Policy/Project Issues	Assess current data access policies and develop new data access policy	Complete
3	**Administrative Post maintenance**	Data Issue	Identify and store faculty administrative roles in PS	In progress
4	**Continue populating the Data Cookbook**	Policy/Project Issue	Write definitions and data profiles for common reporting terminology	In progress
5	**ERP Migration to cloud**	Policy/Project Issue	Identify and discuss structural and process changes that need to be done prior to migration to Oracle Cloud. Identify a "hit list" of changes that would improve our data into the new system	In progress
6	**Definition of "Faculty"**	Data Issue	a definition of "faculty" that includes the various types and situations, such as joint-appointments, ability to vote in dept., salary allocations, courtesy appts. How can we better categorize and store	In progress
7	**Role-based access**	Policy/Project Issues	Assign access to data via administrative roles at the university. This project is dependent upon priority #3, Administrative Posts	In queue
8	**Data policy usage and storage**	Policy/Project Issue	University policy and training for employees who have access to sensitive data on the proper use and handling of data.	
9	**Formalized communication plan**	Policy/Project Issue	Formal communication to data stewards and managers regarding the roles and responsibilities	
10	**Published list of data stewards**	Policy/Project Issue	Curate and publish a website of data stewards	
11	**Consistency of local titles in PeopleSoft**	Data Issue	Consistency of local titles for employees. Staff assoc titles have multiple actual jobs holding that title	

#	Priority 2022-23	Type	Notes	Status
12	**Primary Work Location**	Data Issue	maintenance of a primary work location for all State and RF employees	
13	**Clean Address software implementation**	Data Issue	Software has been acquired but not implemented. Software will integrate with PeopleSoft to validate incoming addresses	
14	**Distribution Lists**	Policy/Project Issue	ability to generate lists of individuals associated with SBU (students, employees, affiliates, etc..) types or populations. Identifying roles and permissions for who can pull population contacts for distribution.	
15	**Limiting access to transcripts**	Policy/Project Issue	Access to student's transcripts to be limited to currently enrolled.	
16	**Import High School transcript data**	Data Issue	Import additional fields for classes taken from HS transcript. Extracting data that comes in from a scan.	
17	**Identifying the owner(s) for Affiliate data**	Data Issue	Having a data trustee for data asset - in this case, a group of people who are "affiliates"	
18	**Centralized location for employee training**	Data Issue	ability to track what training has been accomplished. Some employee training is stored in individual side-systems	
19	**RF job data, status, and work location**	Data Issue	Research foundation employees are not coded with the same level of detail as State employees	
20	**CBASE data**	Data Issue	Medical and Dental class schedule integration with ERP system to understand more about what courses are taught, instructors	