U-M Data Governance Revitalization Securing and Sharing Institutional Data

Current State of Data Governance at U-M

Data governance is a collection of practices and processes that help to formally manage the data assets of the University of Michigan. Some of the key areas of focus of data governance are processes for sharing data, ensuring data quality, providing appropriate regulatory compliance, and accounting for IT security, privacy, and appropriate usage.

U-M has a long-standing data stewardship framework that supports access and sharing of administrative data across the university and provides a support model that aligns and coordinates data stewards, data requestors, and central IT system owners. This model is governed by SPG 601.12 which has not been materially updated since its creation in 1994. Despite the lack of updates, some aspects of the existing stewardship model work quite well, especially for data within the PeopleSoft system, which manages U-M institutional data for Student Information, Human Resources, and Finance.

However, many other aspects of data and its management have changed in the last twenty-five years. U-M engages in frequent and varied data collection and sharing practices to advance its research, teaching, and clinical missions. U-M shares large amounts of data across the university, with other universities and consortia, and with external collaborators. The move into the world of cloud computing, external service providers, big data, data science, and analytics is resulting in more diverse and decentralized data. In addition, laws, regulations, along with the privacy and ethical considerations have put an increased focus on appropriately protecting and using data.

Summary of Recommendations

This document is the result of a cross-campus working group that was convened to recommend any changes required to update the existing data governance model to support the current data environment at U-M. The working group found that the existing data stewardship and governance framework must be updated to better support emerging academic and research needs, data-based decision making, and the changing business needs of the institution. The group agreed on a number of recommendations. The most impactful ones are:

- *Review and Revise SPG 601.12*. SPG 601.12 Institutional Data Resource Management Policy is effectively U-M Data Governance policy. The SPG has not been substantially revised since its approval in 1994. While the SPG has stood the test of time in many ways, it should be modernized to account for advances in data science, technology, and a more complex world of privacy, ethics, and regulation. Its focus should remain on better support, broader sharing and use of institutional data, and should take into account the other recommendations found below.
- *Create a Data Governance Advisory Committee (DGAC) Structure*. A DGAC responsible for reviewing and updating top-level data taxonomy and associated data stewards, arbitrating disputes, and making non-binding recommendations regarding data requests. This group will also make recommendations for upgrades to the data governance program, ensuring that it evolves with the changing data needs of the university. The creation of this group will help reduce two issues with the current framework, that requesters have no defined appeal process for denied requests, and the framework has no formal mechanism to keep it up-to-date.
- *Dedicated resource support*. A revitalized data governance program and supporting processes requires some level of dedicated resources to maintain and evolve, including, but not limited to, a Data Governance Coordinator. The amount of coordination, monitoring, and outreach required by an effective data governance process also requires dedicated resources. All of the peer institutions the working group used for benchmarking provide at least one dedicated resource for their data governance process.
- *Create mechanisms to provide systematic logging of all requests*. Today, most requests are managed through emails and verbal communications. There is no common system where requests and their resolutions are logged. A logging and audit system will help U-M to ensure proper compliance and privacy practices are followed, and will allow periodic review of the status and health of shared data sets. This system should be capable of logging all data requests, both those that use the data concierge service, and those that use other channels.
- *Define and provide central services that can support and guide any data request*. A support mechanism with a centralized “concierge” service eliminates the need for requesters to know which data stewards to contact and what process to follow. As the types of data have grown at U-M, the process to request data has gotten more complex, especially with requests that span multiple data sets, and multiple data stewards.

  This service is a convenience to users of the governance framework; requests are not required to use it and can be managed through other channels when appropriate.
- *Reduce the number of, and better define the roles within the framework*. The existing framework has several roles that add a level of specificity not needed in describing the

framework. Removing these roles makes the overall data governance program and processes easier to understand and navigate.

The subsequent sections within this document contain more detail and specifics for each of these items.

The Case for Change

As mentioned above, data governance practices at U-M have, in many ways remained stagnant for decades (noting that Michigan Medicine has significantly advanced data governance processes and practices in some areas). In places where the data structures are foundationally unchanged, such as within the PeopleSoft system and the ITS data warehouse, the practices continue to meet historical needs.

However, the needs of data governance have, and continue to, evolve. Today, there is much more regulatory scrutiny, focus on ensuring IT security and privacy, higher expectations of accurate data, and the requirement of an audit trail in case the history of a data set and how it was shared is needed. The continually decreasing cost of computing equipment and services has made it easier for data requesters to make copies of data sets, merge multiple data sets, augment data sets with additional information, and leverage any number of external service providers.

At present, data governance practices at U-M are decentralized, at times ad hoc, with little-to-no formal, documented, and consistent practices across data domains and between departments. Access to data is in pockets and processes for obtaining and sharing data are unclear. There is a lack of training and education resources for stewards or data users. All of which leads to data sharing and review processes being extremely slow and inconsistent. In addition, there is little-to-no audit trail, so it is impossible to tell what data has been shared, with whom, and for what purpose. All of this complexity encourages data requesters to collect and store redundant data set, increasing risk to the university and costing time and money.

The elements of the existing data governance framework, including SPG 601.12, must be enhanced and updated to support the current environment. The framework must also be positioned to be more programmatic in order to continually evolve in lock-step with the changing data environment going forward.

The working group, and this document, focuses only on the data governance process. Related processes such as the physical effort of sharing data, and the cataloging of data sets and data flows at U-M are outside of the scope of this group.

Role Redefinition

Today there are myriad roles in the data governance framework. While these may have made sense at one time, they have, in some cases, become ill-defined in the current environment, over-prescriptive, and it's sometimes confusing to understand the responsibilities of each role. The updated framework will reduce the number of formal roles and better clarify those that remain.

The changes to the defined roles are described in the following table.

Current	Today	Proposed
Executive Data Steward	Does not exist in the current framework	A high-ranking officer at the university that retains ultimate authority over a specific area of data. The data stewards for an area will typically belong to the executive data steward's organization.
Data Steward	This is a high-ranking officer at the university. All data stewards delegate the day-to-day stewardship to delegated data stewards	The current role is effectively replaced by the existing delegated data stewards. The specific responsibilities of data stewards are detailed in their own section of this document.
Delegated Data Steward	This role provides policy-level stewardship for a specific data area.	Remove Role. The responsibilities of this role are moved to the data steward role. Data stewards are free to assign individuals to help them steward the data, if necessary.
Data Manager	Data managers have operational responsibilities for a specific data area.	Remove Role. While no longer a formal part of the data governance framework. Data stewards are free to assign individuals to help them manage the data, if necessary.
Data Management Integration Coordinator	University staff who represent the technical aspects of data sharing.	Remove Role. ITS and HITS will provide technical support where necessary. Usually, this is outside the scope of data governance, although some responsibilities of the technology providers are detailed in a section of this document.

Data Custodian	Not defined in the current framework. Some responsibilities of the custodian are managed by the delegated data stewards and data managers.	The person or area that 'creates and/or maintains' the data. This may be an IT group, unit/department or researcher/faculty member. The choice of the word custodian is to reinforce that U-M owns the data.
Data Governance Coordinator	Does not exist in the current framework.	A role dedicated to the on-going operation and coordination of the data governance program.
Data Governance Advisory Committee member	Does not exist in the current framework	This new committee is described in a section below.

Each of these roles is described in more detail below.

Data Steward

Data Stewards are responsible for a specific area of UM institutional data. Data stewards are typically senior-level university officials who are empowered to oversee a specific area of institutional data. Often, but not always, data stewards are also data custodians for some or all of the data in their area. Data stewards may appoint backups and assistants to monitor and maintain their data domain. The data governance coordinator will work with data stewards to keep the list of authorized stewards and assistants current.
Responsibilities of data stewards include:

- Maintain knowledge of applicable laws, regulations, UM policies and best practices relevant to a particular data set in order to help ensure compliant practices..
- Provide and manage requirements and restrictions for access to institutional data and approve or decline data access requests. Manage data access in a way that is consistent with University roles and responsibilities and the institutional data access philosophy.
- Cooperate with other data stewards and custodians to resolve requests for data sets, or combinations of data sets, that span the areas of multiple data stewards.
- Work with data custodians to establish procedures to ensure completeness, accuracy, and integrity of data, and to resolve any inconsistencies in the data.

- Work with data custodians and Information Assurance to classify the data for sensitivity and criticality, and establish and maintain a data catalog for their area.
- Promote relevant training, education, and awareness. This includes communicating business rules and definitions.
- Coordinate with the Data Concierge service to monitor the logs for inappropriate sharing or usage of data. Ensure corrective action is taken, including modifying or revoking access rights.

Some responsibilities assigned to data stewards and data managers in the current framework will be reassigned to other areas in the new framework. For the most part, these responsibilities have always been managed in the proposed areas and the changes to the framework simply reflect the existing process. Some of these changes are:

- Ensure accurate, valid, and timely collection of data - the responsibility of the data custodian and the technology provider.
- Set policies about storage and protection of data - the responsibility of the technology provider and Information Assurance.
- Ensure disaster recovery and business continuity plans are developed and implemented - the responsibility of the technology service provider, with appropriate guidance from Information Assurance can provide guidance.
- Log and audit user activity - the responsibility of the Data Concierge service and technology provider.
- Manage data extract schedules - the responsibility of the technology provider.

Data Custodian

The data custodian is the person or area that ‘creates and/or maintains the data. This may be an IT group, unit/department or researcher/faculty member. The choice of the word custodian is to reinforce that UM owns the data.

When the existing framework was implemented, most large institutional datasets were managed centrally. So the existing framework has no provision for data sets that fall into a data steward’s domain but aren’t directly managed by the steward. In today’s environment, there are a growing number of distributed data sets created and managed by units, departments, researches, and third parties that fall within the scope of a data steward, yet the day-to-day management of the data falls outside the control of the steward. In the revised framework, the custodians of these data sets must coordinate with the data stewards to strive to ensure that data stewardship covers all university data.

Data stewards and data custodians must work together to ensure all data steward responsibilities, and policies governed by SPGs, are fulfilled for all data sets within the data steward's domain, including data sets not within the direct control of the data steward.

Responsibilities of data custodians include:

- Work with data stewards to establish procedures to ensure completeness, accuracy, and integrity of data.
- Work with data stewards and Information Assurance to classify the data for sensitivity and criticality, and establish and maintain a data catalog for their area.
- Work with data custodians to establish procedures to ensure completeness, accuracy, and integrity of data, and to resolve any inconsistencies in the data.
- Work with technology providers to ensure accurate, valid, and timely collection of data.

Central Service Support for Data Requests (Data Concierge service)

ITS will provide a Data Concierge (DC) service, which can provide support and guide all data requests, ensuring simplicity and transparency for users of the data governance process. This group will manage requests for data sharing from the time of the request until its resolution. The DC service eliminates the need for requesters to know which data stewards to contact and what process to follow, ensures requests are properly logged and provides consistent communication and documentation throughout the process.

This service is a convenience to users of the governance framework; requests are not required to use it and will still be able to be managed through other channels when appropriate. Responsibilities of the data concierge service include:

- Coordinate with the appropriate data stewards, data custodians, and technology areas to resolve requests to share data.
- Manage the logging process to ensure all requests have an audit trail.
- Work with data stewards to ensure the data catalog is accurate and up-to-date.
- Coordinate with technology owners and data custodians to ensure data retention processes are consistent with legal or University policies.

Data Governance Advisory Committee

The data governance advisory committee (DGAC) is responsible for making recommendations regarding data requests and for upgrades to the data governance program, ensuring that it evolves with the changing data needs of the university. The creation of this group will help reduce two issues with the current framework, that requesters have no defined appeal process for denied requests, and the framework has no formal mechanism to keep it up-to-date.

Responsibilities of the DGAC include:

- Review new data types and data stewards, arbitrate disputes and make non-binding recommendations regarding data requests.
- In cases where the requester and the data steward cannot come to an agreement, offer new insights and possible solutions to the impasse. The DGAC may recommend specific requests be escalated to the UM Executive Officers.
- Coordinate data integrity discussions and resolution, when agreed by involved parties.
- Review and advise on which data requests should be communicated to other areas of the university, including the VPIT/CIO, the Chief Information Security Officer, executive officers of the university, specific campus departments, or general outreach and communications.
- Make recommendations for upgrades to the data governance program, ensuring that it evolves with the changing data needs of the university.

The DGAC will consist of a combination of representatives, including, but not limited to: data stewards, appropriate campus representatives, staff from Information Assurance, and the data governance coordinator.

Data Governance Coordinator

The Data Governance Coordinator ensures all aspects of the data governance process are working as expected, ensures the various roles are continually educated on their duties and best practices and provides outreach and visibility to the UM community. The coordinator does not manage data requests; these are the responsibilities of data requesters, data stewards, and the concierge service, if engaged. However, the coordinator may report on the status of requests to ensure they are managed in a timely fashion.

Currently, on campus, there are no dedicated resources to ensure that the data governance process operates properly and efficiently. The new framework strongly recommends that some level of FTE be dedicated to ensuring all of the various pieces of the framework and program are running and meeting goals. Michigan Medicine has an FTE, and supporting committees

focused on Data Governance, and all benchmarked institutions have dedicated staff for data governance.

Responsibilities of the DG coordinator include:

- Help build and sustain relationships between data stewards, themselves, and with other roles in the data governance framework, in order to help facilitate cross-steward cooperation when needed for data quality or multi-dataset requests.
- Ensure data stewards have appropriate education and awareness to perform their responsibilities.
- Maintain strategic direction and goals for the data governance program.
- Monitor and publish metrics and key performance indicators for the governance program.
- Provide information to ITS Communications for outreach to the UM community at large.
- Monitor requests to ensure they are resolved in a timely manner.
- Review logs and ensure any anomalies are resolved with the help of the appropriate data stewards
- Coordinate with related data functional areas to help the entire data ecosystem work together. For example, working with the Business Intelligence area to help with data consistency.
- Coordinate the DGAC meetings and distribute action items to assigned parties.

Technology Provider

While not a formal role in the data governance framework, there are certain activities, managed at the technology level that are necessary for data governance to properly function. For the most part, the roles in the framework should not have to directly interface with the technology providers - the Data Concierge service will coordinate communication between the technology providers, data stewards, and requesters.

Some responsibilities of the technology provider include:

- Work with data custodian to ensure accurate, valid, and timely collection of data
- Ensure appropriate IT security and compliance requirements are implemented and maintained in consultation with Information Assurance.
- Ensure appropriate disaster recovery plans are developed and implemented.
- Implement data retention requirements consistent with legal or University policies
- Log request and user activity and communicate logged information to data stewards for auditing.
- Manage data extract schedules.

Some benefits of the new structure

- Reduced need for requesters to know which data stewards to contact and what process to follow, through a concierge service
- Better communication and education of the data governance program
- Better training and awareness for data stewards and other people in the data governance program
- An established process to keep the data governance framework and program up-to-date, to classify and govern new institutional data as it's created, and to keep associated data categories and data governance roles current
- An ongoing commitment to maintain and update policies and principles for data stewardship
- Updated policies, principles, and practices that balance the imperative to share data AND a deeper focus on privacy and IT security
- An appeal process facilitated through an advisory committee
- Reporting or auditing on the governance process and its effectiveness

Working group members and benchmarked institutions

2019 Working Group:

- Laurie Alexander, Libraries/Academic Affairs
- Sol Bermann (co-chair), ITS Information Assurance
- Wendy Bezotte, Michigan Medicine/Medical School
- Becky Chadwick, UM Dearborn
- Chris Eagle (co-chair), ITS Strategy and Planning
- Patrick Franklin, ITS Information Quest
- Susan Gelman, LSA/Faculty
- Vikki Hamilton, Office of University Development
- Rebecca Hulea, Michigan Medicine/Compliance
- Mark Nogueira, Office of General Counsel
- Tracy Pattok, Office of Budget and Planning/Provost
- Paul Robinson, Office of Enrollment Management
- Cindy Shindledecker, UM Office of Research
- Denise Stegall, UMHR/Business and Finance

The working group leveraged the work documented in the 2016 Data Governance Working Group Exploratory Findings

In addition, benchmarking was done with the following peer universities:

- University of California-Los Angeles (UCLA)
- University of Colorado
- University of Illinois
- Indiana University
- University of Maryland
- Ohio State University
- Penn State University
- Purdue University
- Stanford University
- University of Texas-Austin
- University of Virginia
- University of Washington
- University of Wisconsin-Madison