Data Governance Checklist and Reference Guide

# Introduction

Data is at the center of everything we do at Ohio State, and our systems are critical for creating, storing, and using data. To properly protect our data and ensure that it can support business needs, the data must be properly managed throughout its lifecycle. Implemented well, data governance saves time, reduces risk, improves consistency, increases understanding and enhances data-informed decision support. This document serves as a resource to ensure that all projects incorporate sound governance principles from the outset. While most projects entail implementing new systems or tools, this guide also addresses data governance needs for retiring or decommissioning existing systems.

## Key Concepts:

* **Requirements:** The data that the system collects or provides, including how the data will be used within the system itself and by downstream systems, processes, or reports.
* **Validation:** Methods used to ensure the quality of the data entered into the system at the point of entry and any systems of integration.
* **Integrity:** How data quality and consistency will be maintained within the system throughout the life of the data.
* **Reporting and Analysis:** The reporting needs within the system itself and downstream systems or data repositories including the data catalog.
* **Security and Privacy:** How data will be secured and protected and how access and distribution of data will be managed.
* **Retention and Disposal:** Data retention and disposal of data according to the university Retention Schedules.

# Quick Checklist

## Requirements

Understand the purpose, audience, and data elements, including how the data will be used within the system and by downstream systems, processes, or reports.

☐ Are you using standard terminology and field names if possible? Are there new data elements or metrics that need to be defined and classified? What metadata will be available within the system to support data entry or business process completion?

## Validation

Create a plan to ensure data quality and consistency at point of entry and integrations.

☐ Will data in this system be prepopulated with data from another system or need to sync with another system? Does manually entered data conform to best practices? Do retired data systems need to be addressed?

## Integrity

Ensure data quality and consistency are maintained in the system over time, including plans for ongoing validation, testing, and exceptions.

☐ Will data integrity be measured and managed? What is the plan for audits and testing? Where will user questions about the data be directed?

## Reporting and Analysis

Consider the needs of reports and analyses drawn from the data.

☐ Where will reporting and analysis occur, internally and/or externally? Do reporting roles need to be assigned or created? Is de-identification necessary? How will reports and dashboards be tested and validated?

## Security and Privacy

Secure the data from unauthorized access and ensure sharing and storage follow guidelines based on the data’s classification.

☐ What is the classification of the data in the system? Is a risk assessment is needed? How will the data be secured and monitored? Are there privacy concerns?

## Retention and Disposal

Plan for data retention and disposal according to the university retention schedule.

☐ What is the retention schedule for the system data? Will data need to be archived from the system? What is the ongoing plan for records retention and destruction according to the schedule?

## How to Proceed

Use the following reference guide with detailed questions to guide planning and decisions about data throughout the project lifecycle.

## Who to Contact?

If you have questions or need assistance in using the Checklist and Reference Guide, please contact the Data Governance team at DataGovernance@osu.edu.

For questions about specific data elements and appropriate use, Data Stewards are critical resources. Please consult the [Data Stewards Directory](https://it.osu.edu/data/data-stewards-directory) for the appropriate Data Steward who can answer questions or direct you to the appropriate Subject Matter Expert. If your unit has not been assigned a Data Steward, please contact the Data Governance team and we will connect you with a Subject Matter Expert.

# Reference Guide

This section contains a detailed list of concepts and questions to be addressed throughout the project lifecycle to effectively manage data assets. As stated above, this document serves as a resource to ensure that all projects incorporate sound governance principles from the outset. While most projects entail implementing new systems or tools, this guide also addresses data governance needs for retiring or decommissioning existing systems.

Use this checklist for any project that creates, changes or archives data.

The data governance checklist is organized by topic but also contains a column indicating the corresponding phase of the Software Development Life Cycle (SLDC) as listed below, during which the checklist items should be considered.

1. Planning
2. Analysis
3. Design
4. Implementation
5. Testing and Integration
6. Maintenance



## Data Requirements

This section addresses the data that the system collects or provides, including how the data will be used within the system itself and by downstream systems, processes, or reports.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 1. | * What is the purpose for the system and for the data gathered by it?
* Who is responsible for decisions about system data?
 | 1. Planning |
| 2. | * What is the intended use of the data?
* What data is most critical? Recognizing that not all data is equally important, what is highest priority to govern for this system?
 | 2. Analysis |
| 3. | * What is the intended use of the data elements being collected?
* Consider how institutional and common definitions can be leveraged as much as possible during this stage. Inconsistent use can confuse users and impact data integrity.
* What new data elements or metrics need to be defined and classified?
 | 3. Design |
| 4. | * Can the system be configured or customized to leverage standard field names for clarity and transparency?
* Are the labels, descriptions, and categories consistent within the system and compared to other systems?
* Create data element and term definitions in the Implementation phase, confirming work from Design Phase
	+ Source of data
	+ Relevant calculations
	+ Intended use of data
	+ Data element classification
	+ Definition of field
	+ SME / contact for questions
* What metadata will be available within the system to support data entry or business process completion?
 | 4. Implementation |

* If project is system retirement, consider the following:
	+ Will the data from the system need to be preserved after the system is taken down?
	+ Will it be used to populate a replacement system?
	+ Will it be extracted and stored in a data repository for storage? If so, how will users access the data?

## Data Entry and Validation

This section addresses methods used to assure the quality of the data entered into the system at the point of entry and any systems of integration.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 3. | * Will data in this system be prepopulated with data from another system?
* Will data in this system need to sync with data in another system? What are the rules for data synchronization?
* For data that is manually entered into this system, consider the following:
	+ Field-level validation such as dates, input masks, range of valid values
	+ Drop-down lists rather than freeform text entry
	+ Making fields required
	+ Careful use of null values and default values
	+ Derived fields
	+ Audit reports
 | 3. Design |

## Data Integrity

This section addresses how data quality and consistency will be maintained within the system.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 3. | * How frequent are exceptions to the main business processes of the system? How will exception affect data integrity?
 | 3. Design |
| 4. | * How will data validation occur for direct data entry or data inserted via integrations?
* How will data integrity be measured and managed over time? What audits are needed?
 | 4. Implementation |
| 5. | * What is the system test plan?
* How will data conversions be validated?
* How will data validation controls be tested?
 | 5. Testing and Integration |
| 6. | * Where will user questions about the data be directed?
 | 6. Maintenance |

## Reporting and Analysis

This section addresses the reporting needs within the system itself and downstream systems or data repositories including the data catalog.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 3. | * Where will reporting and analysis occur? Are reporting roles needed in the system?
* Will data be shared externally? Is a use or sharing agreement needed?
* Will the data flow to a data lake or warehouse? Are integrations needed in or out?
	+ What views will be needed to support analysis and reporting?
	+ What analysis is needed such as trends over time, period-over-period, internal vs. external benchmarks, actuals vs. goals, etc.)
	+ Will data be de-identified for analysis purposes? How will de-identification be accomplished?
* As reports are designed, plan for certification for those that will be widely used or are enterprise in scope and collect general report metadata such as purpose and description - the "what" and the "why" - during this stage.
 | 3. Design |
| 4. | * As reports are developed, create metadata, and populate catalog template (unless system will populate data catalog via integration)
 | 4. Implementation |
| 5. | * How will reports and dashboards be tested and validated?
 | 5. Testing and Integration |

## Data Security and Privacy

This section addresses the ways that data will be secured and protected and how access and distribution of data will be managed.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 2. | * What regulated data will be in the system such as PII, PCI, etc.?
* What is the classification of the data in the system?
* What risk assessment is needed?
 | 2. Analysis |
| 3. | * How will the data be secured?
	+ What access processes will be used to request system access?
* What system roles will permit updates to system data?
 | 3. Design |
| 4. | * How will system use be monitored? What logs will be created and how will they be managed?
	+ Do they identify user activity? What is the privacy impact?
 | 4. Implementation |
| 6. | * How will system access be reviewed on a regular schedule?
 | 6. Maintenance |

## Data Retention

This section addresses data retention and disposal of data according to the university Retention Schedules.

|  |  |  |
| --- | --- | --- |
| SDLC # | Data Requirement Elements | Corresponding Software Development Life Cycle Phase |
| 6. | * Plan for data retention and disposal when retention limits are reached:
	+ What is the retention schedule for the system data?
	+ Will data need to be archived from the system? How will that occur?
 | 6. Maintenance |

* If project is system retirement, consider the following:
	+ Has any of the data in the system reached its retention limit?
	+ If so, how will that data be destroyed following approval of a destruction certificate by the University Records Manager?
	+ What is the ongoing plan for records retention and destruction according to the schedule?