

Planning a Data Management & Governance Vision or Strategy

Introduction

In this activity you will be guided through a worksheet with the goal of identifying key data management and governance challenges your organization is currently facing, as well as some immediate action steps for addressing those challenges.

This activity is not intended to provide an end-to-end roadmap for implementing effective data management practices. However, our hope is that it provides a useful jumping off point for you to explore your own organization's unique data needs and solutions.

Each section below provides instructions and reflection questions related to a different section of the following worksheet.

Do You Know Your Data?

Pick a dataset or data source that is mission-critical for your organization, or that gets a lot of use—that is, one of your best assets and answer the following questions:

- Do you have a data dictionary for the dataset?
 - If so, does it contain:
 - A data table description (or one for each table, if there are multiple) including:
 - What *type of entity* it describes (i.e., students, courses, schools, etc.)
 - The data's *provenance* (i.e., how the data is being generated, or where it is being sourced from)
 - The name of each data field (including any abbreviation used for data storage)
 - What each field represents
 - The type of data each field contains (numeric, text, etc.) and a range of possible values
 - Whether a field is calculated from other fields and, if so, the formula
 - If not, do you or someone in your organization have the information you'd need to develop one?
- How is your data being stored—in a text file format like CSV (or TSV, etc.), in a spreadsheet, or in a database?
- How is your data structured—in a tabular format or unstructured? If the former, is it in a long or a wide format?
 - [Tabular Data](#)
 - [Unstructured Data](#)
 - [Long vs. Wide Data](#)
- Does your data follow some of the best practices we've discussed?
 - If tabular, is it in a rectangular shape (i.e., the data is arranged in complete columns and rows, with missing values specified/represented by NA or some other value)?
 - Are your variable names meaningful/descriptive and consistent?

- Are the values of any categorical variables consistent (i.e., there aren't any duplicates due to misspelled values, etc.)?
- Does each table in your dataset only refer to one type of unit/use a consistent level of detail (i.e., no rows are aggregates, student/school records aren't in the same table)?
- If the data is in a spreadsheet format, colors aren't being used to represent data?
- If the data is in a database, has it been [normalized](#)?

Challenges?

Based either on your responses to the previous questions (i.e., where your dataset doesn't meet an indicator of quality/consistency), or on other cases where you have faced challenges deriving reliable insights from the data, what do you see as some of the core challenges you are facing with this dataset? Think of [fitness for use/purpose](#)—when is your data not fit for use in a scenario where it is needed, or when has it failed to meet its purpose? Why? What would you need to change to make the data more fit?

After you've identified some of your key challenges (no need to fill the entire table if you don't need to!) try to prioritize these in order of their importance or impact on your organization's functioning. For example, if there is a challenge that is preventing you from deriving any useful or actionable insights from the data, rate that a "1"—any challenges that prevent the data from being as useful as it could be, but do not prevent its use entirely, rate lower.

Data Management Domains of Focus

In some of these cases, the potential "solution" to the challenge may seem obvious. If inconsistent categorical values are preventing you from getting accurate aggregations of data for that variable, the obvious solution might be to just edit the incorrect values. This may not address the underlying issue, however—are those values inconsistent due to data entry errors, or because the categories being used are unclear/not mutually exclusive? Is this an issue you'll only need to fix once, or that you'll need to keep fixing?

One way of identifying the sort of solution that will actually address a challenge is to determine which data management "domain" it would fall under. *Using the definitions provided below, try to categorize each of the challenges you identified based on the domain that addresses each.* Are there any that are recurring or higher priority depending on your earlier ranking?

Domain	Description
Data Governance	Planning, oversight, and control over management of data and the use of data and data-related resources.
Data Architecture	The structure of data and data-related resources.
Data Modeling and Design	Analysis, design, building, testing and maintenance of statistical models.
Data Storage and Operations	Physical data storage deployment and management.

Domain	Description
Data Security	Maintaining data privacy, confidentiality, and controlled access.
Data Integration and Interoperability	The movement and processing of data between systems and tools, from acquisition to delivery.
Documents and Content	Storing and organizing and maintaining access to data found in unstructured sources (electronic files) and integrating with structured (database) data.
Reference and Master Data	Management of shared data to reduce redundancy and ensure better data quality through standardized definition.
Data Warehousing and Business Intelligence	Management of analytical data processing and enabling access to data for further analysis and reporting.
Metadata	The collection, management, and delivery of information about datasets.
Data Quality	Defining, monitoring, and managing data quality/data integrity

Source: Adapted from Table 7.1 in *Modern Data Strategy*

Develop a Vision Statement

Now that you have identified a few key data challenges your organization is facing, let's try and come up with a vision statement that encompasses and motivates these various strategic items.

Remember, a vision statement sets out an ideal *future* state. Some tips for writing a vision statement can be found here: <https://www.cascade.app/blog/write-good-vision-statement>.

1. Define what you *do* as an outcome—what is the ideal end state your organization hopes to see with regards to managing your data? Remember, it's not precisely about what you *do* but what you hope your successful work will *achieve*.
2. Identify the unique twist your organization brings to make this outcome appear. For more mission-driven organizations this will read slightly differently than a for-profit institution, but the point remains that your organization brings *something* to the mission you believe in. What is that thing?
3. Determine a high-level quantifiable data management goal that addresses the broad challenge you identified earlier. It need not be specific, in fact it's better if it is not too specific, but having a goal here can be useful.
4. The final tip is to make the statement sound relatable and attractive. Your vision statement can be inspirational or rousing. It need not be a mere formality (though that is fine also)

Develop a Strategy

Looking at the challenges you identified, as well as their domains, are there a few that are overarching, interconnected, or that apply to other data that your organization works with? Try distilling these into a couple of core data challenges your organization might be facing.

Taking into account what we've discussed so far, what are some immediate next steps your organization might take to address each challenge? Start by further investigating the domain associated with each challenge and any potential actions or strategies associated with improvement in each (the first line already sets this up for you). As you identify next steps, who are the people—in your organization or outside it (stakeholders, consultants, etc.)—that are going to carry out this work? What resources will they need to be successful?

What next?

Now that you have a high-level vision and strategy in mind for improving your organization's data management and governance, next comes the hard work of putting it into practice! If you are looking for a larger overview of the concepts and principles behind data management and governance, check out our eLearning. For more information, contact Tucker Plumlee (tplumlee@crec.net).

Fundamentals of Data Management and Governance

Data is just as valuable an asset as an organization's physical infrastructure—how can it be managed to ensure efficient and effective deployment to meet the organization's goals? This course provides an introductory understanding of key principles and best practices in data management and governance. In addition to better understanding their own data infrastructure and practices, participants will identify potential opportunities to access and share data held by federal, state, and local partners.



Pre-Requisites: The material does not assume technical expertise with data management systems or tools beyond basic Excel and is appropriate for those with varying levels of data handling experience.

Topics Include:

- Handling different data structures (long vs. wide) and types (text vs. numeric).
- Good practices for using spreadsheets and determining when a database is more appropriate for data storage.
- Thinking of data as an organizational asset rather than a tool.
- Frameworks for understanding the different aspects of data management.
- Developing a data governance vision, strategy, and standard operating procedures.

Dataset:

Do You Know Your Data?

Dictionary: Yes / No

Storage Type: CSV / Spreadsheet / DBMS

Shape: Tabular / Unstructured || Long / Wide

Data Quality Factors:

Variable names meaningful/consistent

Missing values specified

Consistent categories

Rectangular format

Each table uses one unit/level of detail

If spreadsheet, no colors

If database, normalized

What Are Your Data Challenges?

Priority	Challenge	Domain

Develop a Data Management Vision Statement

Develop a Data Management Strategy

Challenge	Next Steps	People/Resources
	1. Investigate _____. 2. 3. 4.	
	1. Investigate _____. 2. 3. 4.	
	1. Investigate _____. 2. 3. 4.	