Data Governance: A Primer

Braden J. Hosch, Ph.D.

Asst. Vice President for Institutional Research, Planning & Effectiveness

Stony Brook University

AIR
ASSOCIATION FOR INSTITUTIONAL RESEARCH
Data and Decisions for Higher Education

Overview

- Data governance concepts and major aspects
- Selling data governance to senior leadership
- Characteristics of a data governance system
- Maturity models
- Change management in a college or university
- Technological "solutions"

Outcomes for workshop participants

Define data governance as an activity that centers on human behavior more than data

Describe major components of data governance activities

Articulate challenges on their campus and how data governance will address these challenges

Assess their campus culture and organization with a data governance maturity model; select and modify a data governance maturity model for their campus

Identify characteristics of a data governance system; analyze where their own institution has gaps; and create an outline for how data governance could fit into existing organizational structures

Discuss how technology may assist but not perform data governance; describe major functions of data governance software applications or "solutions"

Explain principles of change management in higher education institutions and how they will enable development of data governance on their campuses

Construct an action plan for next steps on their own campus to advance data governance activities

What this workshop will not do

- Design your data governance system for you

- Promote specific technological solutions

- Prescribe specific functions, operations, or organization

- Identify how much money to spend

What is data governance?

Data Governance Definitions (Generic)

"the execution and enforcement of authority over the management of data and data-related assets" - R. Seiner (2014)

"specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, storage, use, archiving and deletion of information" - Gartner IT Glossary

"a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" – Data Governance Institute (2014)

Data Governance Definitions from Universities

"formalizes behavior around how data are defined, produced, used, stored, and destroyed in order to enable and enhance organizational effectiveness" – Stony Brook University (2016)

"adds value to our administrative and academic data systems by the establishment of standards that that promote data integrity and enables strategic integrations of information systems" – Vanderbilt University

"the discipline which provides all data management practices with the necessary structure, strategy, and support needed to ensure that data are managed and used as a critical University asset" – U of Rochester

The 5-second elevator definition

Data governance is …

- a set of guidelines for how people behave and make decisions about data

Master data management is often confused with data governance

Master Data Management (MDM)

- Comprehensive method to link all critical data to a common point of reference
- Example:
  - All screens, documents and systems showing a student's address derive from a common location.

Data Governance

- Formalized system for how people make decisions about acquisition, production, storage, distribution, and analysis of data
- Example:
  - Group decides on a definition for home address and agrees on a common source field

AiR

Important characteristics of DG definitions

Data governance IS	Data Governance IS NOT
• More about people and behavior than data	• IT's responsibility
• A system that requires and promotes shared agreement	• Solved by technology
• Formal (i.e. written down)	• Equally applied across all data assets
• Adds value by supporting institutional mission/goals	

AIR

Activity 1 – What data governance features do you have?

List formal and informal structures you have for promoting data

	Formal	Informal
Policies/Practices		
Groups		
Roles		
Responsibilities		

Why do we need data governance?

Justifications for Data Governance

Justify data governance on your campus based on:

Value	Cost	Risk

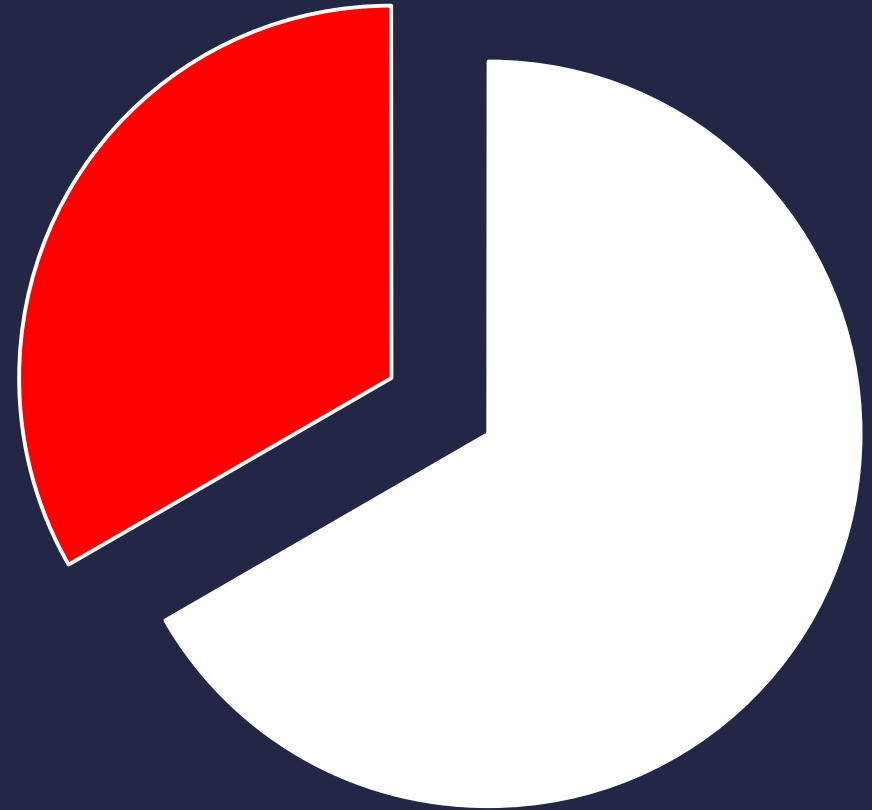# Justifications for Data Governance - Value

Educause identifies significant institutional value to higher education institutions from data governance:

- Official vs. ad hoc data definitions

- Clear responsibilities

- Capacity for analytics

- Competitive advantage

Educause (2015). "The compelling case for data governance," retrieved October 15, 2018 from https://library.educause.edu/~/media/files/library/2015/3/ewg1501-pdf.pdf

Justifications for Data Governance – Cost (1)

A third of Fortune 100 organizations will experience "an information crisis, due to their inability to effectively value, govern and trust their enterprise information."

Gartner. (2014). "Why data governance matters to your online business," retrieved August 1, 2016 from http://www.gartner.com/newsroom/id/1898914s-why-data-governance-matters-to-your-online-business/

Justifications for Data Governance – Cost (2)

Poor data quality costs the US economy $3.1 trillion every year

$3.1 Trillion

IBM. (n.d.). "Extracting business value from the 4 V's of big data," retrieved October 1, 2018 from
https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

Justifications for Data Governance – Cost (3)

The average financial impact of poor data quality on businesses is $9.7 million per year. Opportunity costs, loss of reputation and low confidence in data may push these costs higher.

Forbes (2017). "Poor-quality data imposes costs and risks on businesses," retrieved October 22, 2018 from https://www.forbes.com/sites/forbespr/2017/05/31/poor-quality-data-imposes-costs-and-risks-on-businesses-says-new-forbes-insights-report

Justifications for Data Governance - Risks

Fines Imposed by Federal Student Aid

Fiscal Year	Clery/Part 86 Imposed Fines	IPEDS Imposed Fines	Other Imposed Fines	Total Imposed Fines
2010	$42,000	$225,000	$48,653,500	$48,920,500
2011	$195,000	$144,500	$4,868,500	$5,208,000
2012	$212,500	$158,500	$624,000	$995,000
2013	$812,000	$56,000	$5,204,137	$6,072,137
2014	$438,000	$111,250	$6,750	$556,000
2015	$500,000	$39,250	$14,130,000	$14,669,250
2016	$307,500	$57,000	$79,462,500	$79,827,000
2017	$2,542,500	$1,500	$382,500	$2,926,500

Source: Postsecondary Education Participants System (PEPS)

AIR

Data as an Asset

By 2020, Gartner predicts that 10% of organizations will have a highly profitable business unit specifically for productizing and commercializing their information assets.

By 2021 companies will be valued on their information portfolios: "Those in the business of valuing corporate investments, including equity analysts, will be compelled to consider a company's wealth of information in properly valuing the company itself."

AIR

Data as an Asset for Universities

Generic Example	At Colleges & Universities

Web sites grant access in exchange for personal data (email address, etc.). These data have value and can be leveraged or even sold.	Data are purchased
	    Names of prospective students
	    Library databases
	    Various datasets (U.S. News, Academic Analytics, etc.)
	Data are sold
	    To vendors for discounts or services
	Lost data carry costs
	    Data breaches

AIR

Who owns the data?

Consider carefully use of the word "ownership" with data
> Often represents assignment of responsibility
> Connotes individual control and property vs. caretaking of shared resource

Institutions own the data

Individuals provide stewardship

Activity 2 - Why do we need data governance?

Identify institution-specific examples that help make the case for data governance

❱ Value – what could you do that you can't do now?

❱ Costs – what costs are you incurring because data are not well governed?

❱ Risks – what risks are you taking because data are not well governed?

Features of Data Governance

Key features of data governance systems

Documents	Groups	Individual roles
- Charter / framework - Principles & values - Purpose & scope - Roles & responsibilities - Written & published policies - Data dictionaries - Communication strategies	- Senior leadership [buy-in] - Policy council - Data steward council(s) - Information security council/program - Positions/office to support DG	- Data stewards - Data custodians/ caretakers - Data users

Principles and values

Establishing principles and values for data governance assists with:

Initial design and implementation	Answering critics	Maintaining focus	Navigating difficult situations

Principles of Data Governance (Generic)

Consistency of data in its sourcing and in its vocabulary, definitions, and taxonomies	Quality which is proactively assessed and standards applied	Ownership and accountability defined across the data lifecycle and recorded in the information asset register
Business alignment which ensures that data is regarded and treated as a key business asset	Access to relevant users, kept secure through access control	Providing trusted insight

Principles and Values – Example University of Wisconsin - Madison

Accountability	Determining who is responsible for the management of data at UW Madison as well as holding them to our outlined standards.
Agility	All of our processes should adapt when necessary
Change Management	New processes demand new and changing staff at UW. We're committed to ensuring smooth transitions and well informed decisions.
Consistency	All decisions made will be applied consistently across campus.
Metrics Driven	We monitor ourselves against measurable goals on a regular basis and use the results to determine courses of action.
Stewardship	Determine formal roles for those in charge of data. This does not mean that everyone on campus is not responsible despite formal roles.
Transparency	We will make it clear how and when decisions are made and when processes are created. We also strive to ensure that decisions and processes are audited to support compliance based requirements.

Principles and Values – Example Stony Brook University

Values			
Shared Assets Data and information are shared organizational resources that constitute valuable assets.	**Stewardship** Employees of Stony Brook University have a responsibility for the curation of data. They serve as caretakers of data to ensure data are collected, stored, and maintained under the premise that others will access and use them over time	**Quality** To ensure data retain value, quality of data is actively monitored and maintained	**Privacy & Confidentiality** Maintenance of individual privacy and confidentiality of educational and personal records represent not only legal requirements but also primary outcomes of data management.

Principles for Data Governance			
Organizational Effectiveness	Transparency	Communication	Compliance
Auditability	Integrity	Accountability	Standards

AIR

Principles and Values – Example Brown University

Guiding Principles of Data Governance at Brown

❯ Institutional data are valuable assets and must be treated as such

❯ Access to accurate and consistent data is essential to informed decision making across the University

❯ Data usage and access rules will be articulated and followed

❯ Data standards can and should be defined and monitored

❯ The security of institutional data is essential, as is appropriate and timely access

❯ The privacy of an individual's information will be protected

Source: https://www.brown.edu/about/administration/data-governance/introduction-data-governance-brown

Connect Data Governance to Mission

Data governance is a system to improve the effectiveness of the organization, not an activity for its own sake

Anchor data governance to mission when justifying need or presenting structure

AIR

Activity 3 - Distill university mission

Data governance should be established to support the institution's mission and/or strategic goals. Colleges and universities have notoriously lengthy mission and goal statements, so it can be a challenge to distill them.

Summarize the main points of your institution's mission, preferably so that it fits on a slide.

Example

Stony Brook's framework for data governance outlines a set of principles, structures, roles, and responsibilities to improve the data infrastructure and to advance institutional goals

Stony Brook has a five-part mission to provide & carry-out:
- Highest quality comprehensive education
- Highest quality research and intellectual endeavors
- Leadership for economic growth, technology, and culture
- State-of-the-art innovative health care, with service to region and traditionally underserved
- Diversity and positioning Stony Brook in global community

Source: https://www.stonybrook.edu/commcms/pres/vision/mission.php

Structure – Generic Example
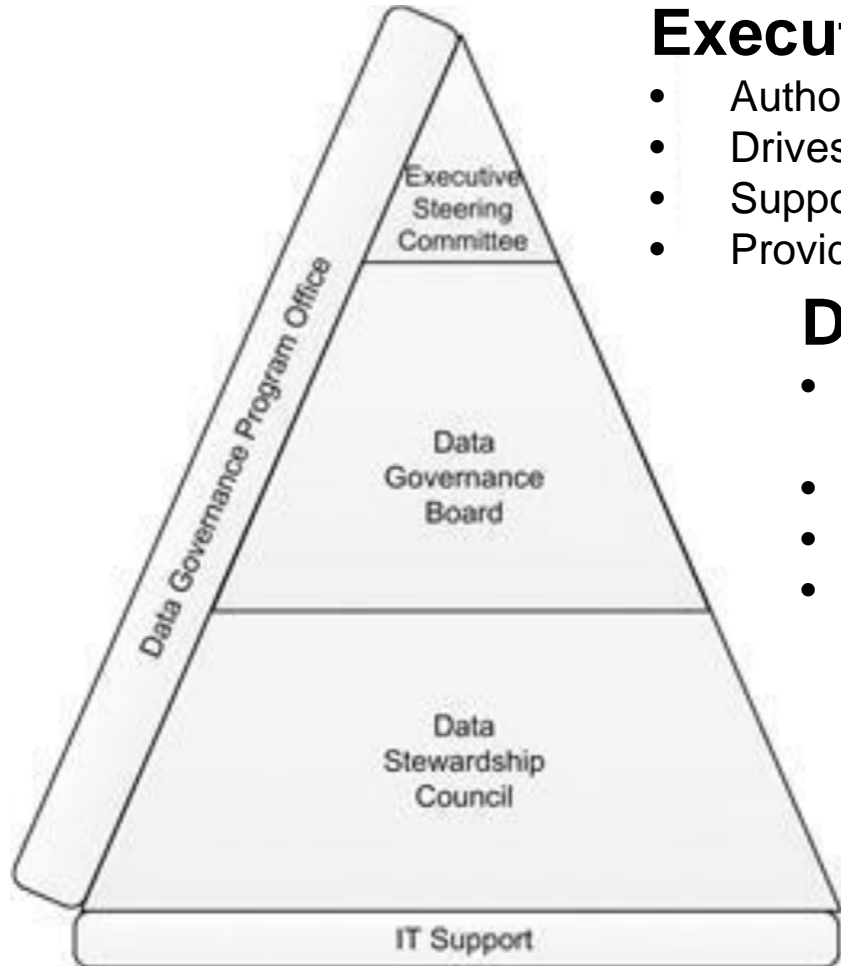
Executive Steering Committee
- Authorized to change the organization
- Drives cultural change
- Supports the program enterprise-wide
- Provides funding for the Data Governance Program

Data Governance Board
- Made up of high-ranking representatives of data- owning business functions who can make decisions about data for the company
- Assign members of the Data Stewardship Council
- Approve decisions of the Data Stewardship Council
- Approve data-related policies

Business Data Stewards
- Experts on use of their data domain data
- Able to reach out to SMEs to gather information and make decisions
- Typically someone who others come to as the most knowledgeable about the meaning of the data (and how it is calculated)
- Makes recommendations on data decisions and write data-related procedures

Plotkin (2014). Data stewardship: An actionable guide to effective data management

Structure – Stanford University
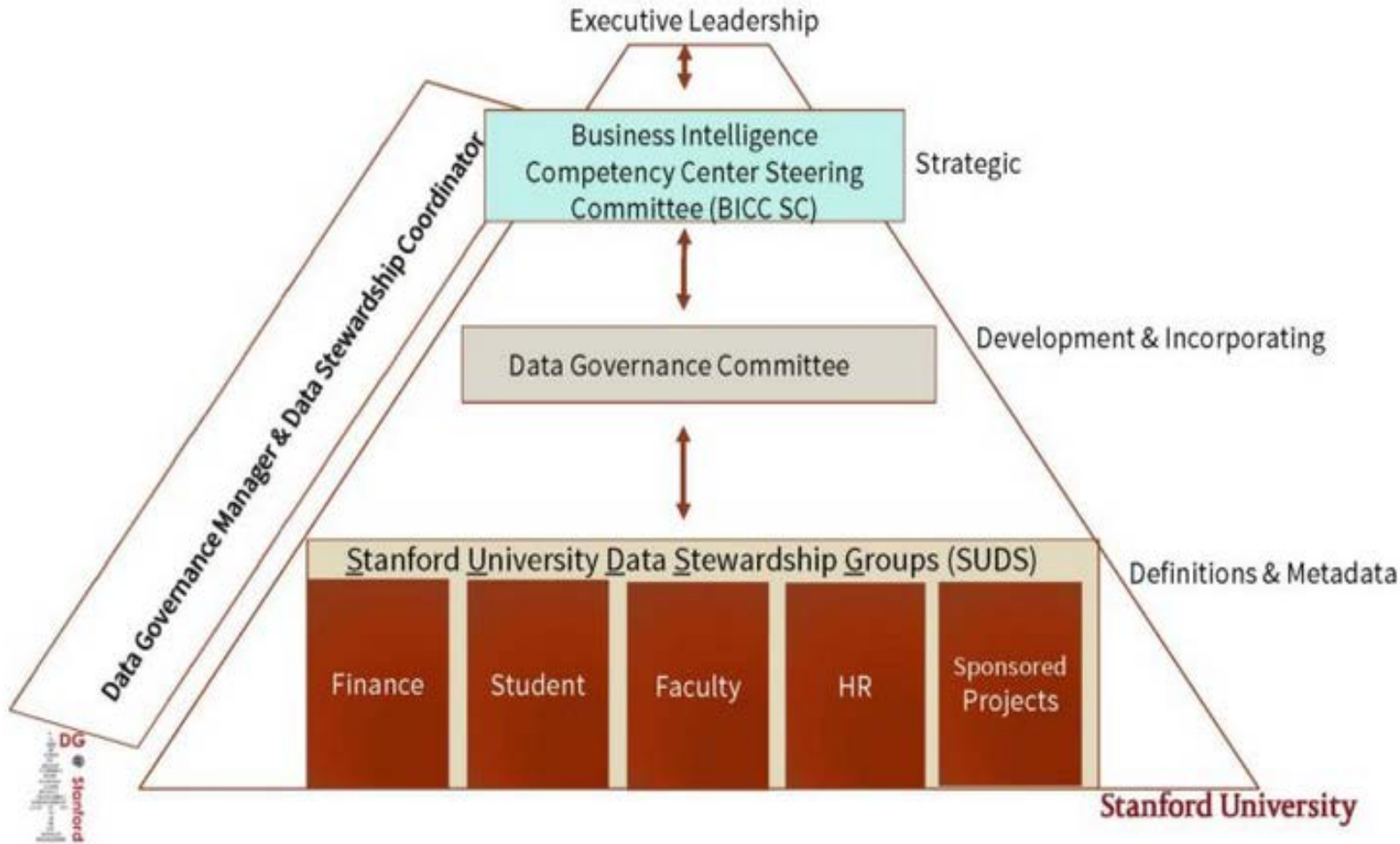
BI Competency Ctr. Steering Committee
- Cross-functional oversight & communicates long-term value of BI program
- Achieves peer buy-in, and effects change in business process and data quality
- DG adopters and champions
- Ensures alignment of DG with university goals

Data Governance Committee
- Sets & incorporates DG policies, standards, procedures, roles & responsibilities
- Includes lead steward from each of the data steward groups, plus reps from additional units

Data Stewardship Groups
- Provide metadata infrastructure to support improved decision-making university-wide
- Ensure information integrity
- Build data knowledge
- Meet compliance requirements
- SMEs who define reporting terms and gather metadata associated with their reporting environment

# Structure – University of Wisconsin-Madison
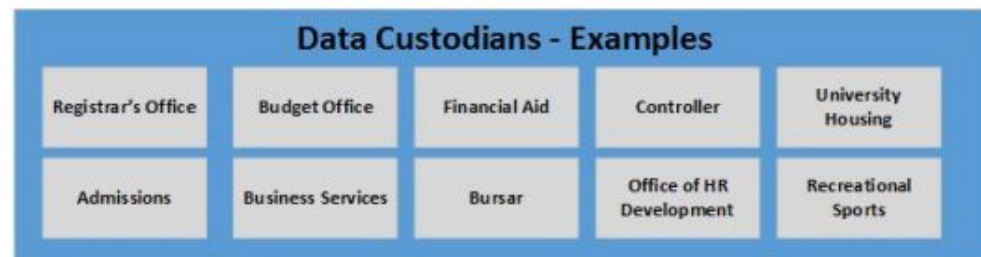
Data Governance Steering Committee
- provides executive level guidance to the program
- promotes Data Governance across UW-Madison
- allows for / facilitates data-driven decision making
- determines priority and budget of major data-related projects.

Data Stewardship Council
- determines operational structure of the program
- drafts, communicates, and recommends approval of data-related policies
- implements, budgets, and monitors data-related programs across UW-Madison.

Structure – Stony Brook University
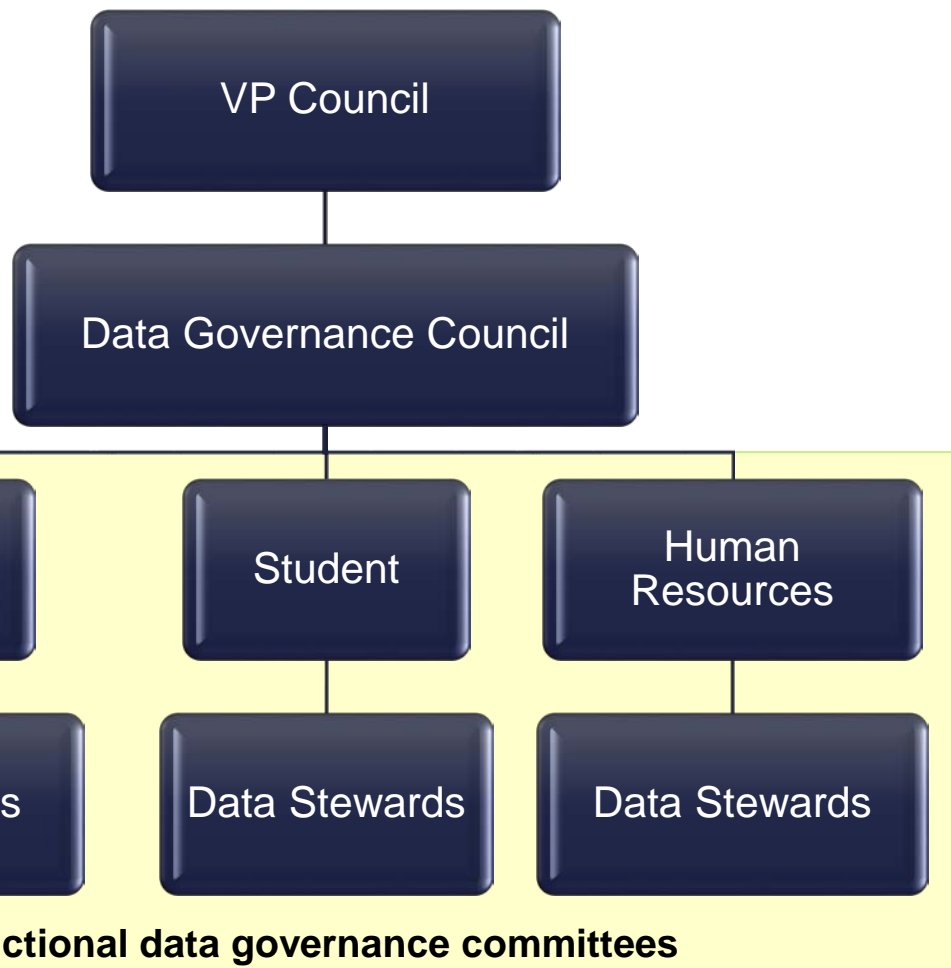
VP Council (Project 50 Forward SteerCo)
- Executive sponsors of project
- Establishes authority and purview of data governance system

Data Governance Council
- Recommends and implements institutional policy for data governance
- Sets priority for

Functional Data Governance Committees
- Implements institutional policy for data governance
- Recommends solutions to specific data issues
- Considers and approve changes to code sets, additions to tables
- Develops solutions to data governance issues
- Communicates with data caretakers in their areas

Policy-Making Body - Data Governance Council

Prioritizes decisions regarding data to address most relevant needs of organization

Reviews, evaluates, and reports on data governance performance and effectiveness

Ensures that annual performance measures align with data governance and business objectives

Reviews and approves data governance policies and goals

Ultimately is accountable for business data use, data quality, and prioritization of issues

Makes strategic and tactical decisions

Defines data strategy based on business strategy and requirements

Plotkin (2014). Data stewardship: An actionable guide to effective data management

Data Governance Council Membership Examples

UW-Madison	Stony Brook
Chief Data Officer	Chief Institutional Research Officer
Director of Univ. Communications	Analytics and Enterprise Data Officer
VP for Teaching & Learning	University Controller
VP for Diversity	Chief Enrollment Management Officer
AVC Business Services	University Registrar
AVC Legal Affairs	Chief Financial Aid Officer
Assoc. Dean Biomedical Informatics	Provost's Office designee
VP Libraries	VP Student Affairs designee
CISO	VP Administration designee
Campus Records Officer	VP Human Resources designee
Assoc. Dean Education	VP Information Technology designee
Faculty/Dean Representation	VP Research designee
	SVP Health Sciences Designee
	University Senate designee
	Chairs & Vice Chairs of FDGCs (6 people)

Data Stewardship Definitions

Data stewardship is the most common label to describe accountability and responsibility for data and processes that ensure effective control and use of data assets. – Knight (2017)

Data stewardship is the operational aspect of an overall Data Governance program, where the actual day-to-day work of governing the enterprise's data gets done. – Plotkin (2014)

Data Stewardship is concerned with taking care of data assets that do not belong to the stewards themselves. Data Stewards represent the concerns of others. Some may represent the needs of the entire organization. Others may be tasked with representing a smaller constituency: a business unit, department, or even a set of data themselves. – Data Governance Institute (n.d.)

Types of Data Stewards

Business Data Steward

- Accountable for data owned by business area
- Work with stakeholders to make recommendations on data issues
- Manage metadata for their data
- Champion data stewardship for their areas

Technical Data Steward

- Provide expertise on applications, ETL, data stores, and other links in information chain
- Assigned by IT leadership to support data governance

Domain Data Steward

- Business steward for widely shared data
- Work with business stewards as stakeholders to achieve consensus
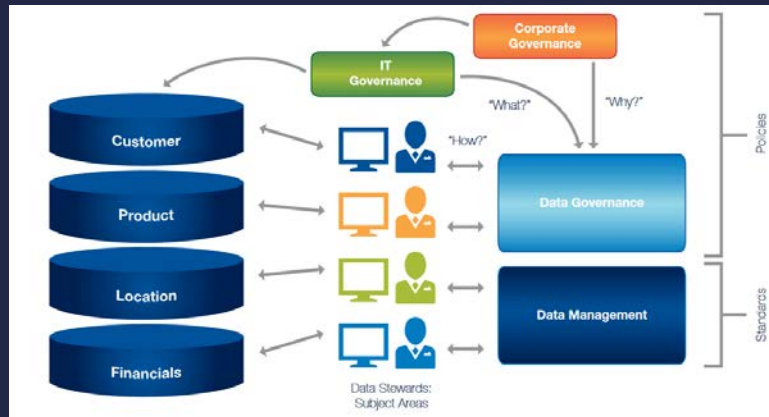
Project Data Steward

- Represent data stewardship on projects
- Funded by projects
- Work with business data stewards to obtain info and make recommendations about data stewarded by business stewards
- Notify business data stewards about data issues raised by the project

Operational Data Steward

- Provide support to business data stewards
- Recommend changes to improve data quality
- Help enforce business rules for the data they use

Plotkin (2014). Data stewardship: An actionable guide to effective data management

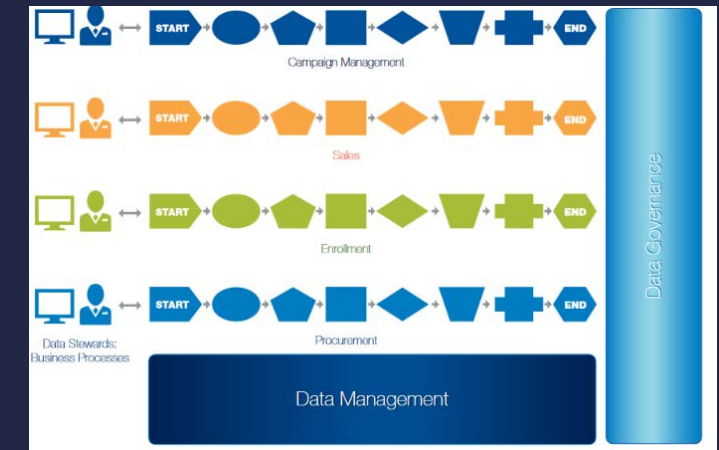# Alternative models for types of data stewards

Data Stewards by Subject Area

Data Stewards by Function

Data Stewards by Process

Data Stewards by System

Data Stewards by Project

Dyché & Polsky (2014). "Five models for data stewardship" SAS Whitepaper
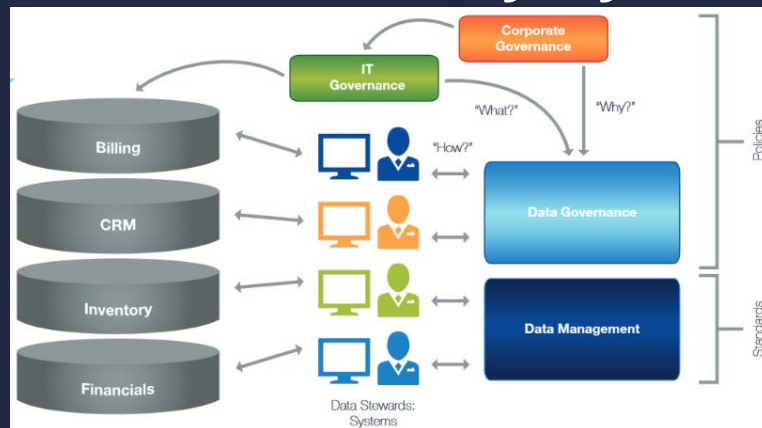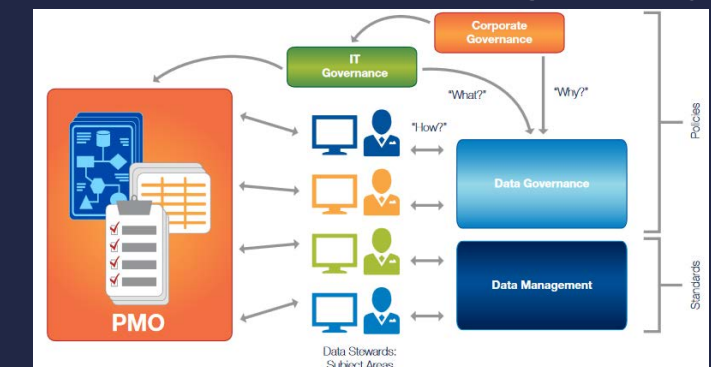
Data Steward Responsibilities

Oversee management of selected data assets

Participate in data governance and carry out decisions

Assist in creation and maintenance of data dictionaries, metadata

Document rules, standards, procedures, and changes

Ensure data quality and manage specific issues

Communicate appropriate use and changes

Manage access and security

Functional Data Stewardship Council/Committees

Coordinate data stewards in related area

Set and review definitions, data quality rules, creation/usage rules, determines official version of metadata

Consider and approve changes & additions to code sets

Ensure dictionary standards are followed in area

Review data quality in functional area; identify practices promoting data quality identify areas for improvement and monitor improvements

Respond to inquiries about process, content, limitations and uses of data, especially in cross-functional settings

Elevate issues that require resolution

Communicate proceedings, including notice of changes and decisions

AIR

Stony Brook Roles and Responsibilities Matrix

	Data Governance Council	Functional Data Governance Cmtes	Data Stewards
Standards and Policies	Define, Establish, Monitors, Audit, Verify, Develop, Revise	Cross functional implementation, coordination	Functional implementation
Data Quality	Identify, Adopt enterprise-wide DQ tool Big picture	Prioritize levels Monitor area Identify needs	Review audit reports, Coordinate clean-up, Initial prioritization
Metadata	Establish standards	Ensure cross-functional alignment	Implement Maintain
Metrics	Review, Identify, Monitor	Monitor area Identify area priorities	Monitor Remediate

B

AIR

Data users

Often not considered in data governance systems (but should be).

Example formal responsibilities (Stony Brook)

Recognize that institutional data and information derived from it are potentially complex. Make efforts to understand the source, meaning and proper use of the data through training sessions, utilizing data dictionaries and knowledge of supporting system processes.	Include information about the data source and criteria when distributing data, reports and ad hoc analytics to guard against misinterpretations of data.	Respect the privacy of individuals whose records they may access. Unauthorized disclosure or misuse of institutional information stored on any device is prohibited	Ensure that passwords or other security mechanisms are used for sensitive data	Report data quality issues to appropriate data steward

AIR

Administrative Office / Positions Supporting Data Governance

In general, offices and positions dedicated to supporting data governance are still emerging in higher education

Chief Data Officer	Data Governance Program Manager
• Purdue University • University of Florida System • University of South Carolina – Columbia • University of Rochester • University of Wisconsin - Madison	• Purdue University • Stanford University • Yale University

Pomerantz, J. (2017) *IT leadership in higher education: The chief data officer*. *Educause. Research Report. Louisville, CO:ECAR.*

Maturity Models

Assess your current state of data governance

Formal assessment of current data governance practices

- Assists with senior leadership buy-in

- Identifies gaps and important implementation considerations

- Extends beyond the informal list we made in Activity 1

- Uses a maturity model to quantify the existing state; allows for measurement of progress in a future state

AIR

Activity 4: Data Governance Maturity Model

	Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**
	Informal	**Developing**	**Adopted and Implemented**	**Managed and Repeatable**	**Integrated and Optimized**
Data Governance	Attention to Data Governance is informal and incomplete. There is no formal governance process.	Data Governance Program is forming with a framework for purpose, principles, structures and roles.	Data Governance structures, roles and processes are implemented and fully operational.	Data Governance structures, roles and processes are managed and empowered to resolve data issues.	Data Governance Program functions with proven effectiveness.
Culture	Limited awareness about the value of dependable data.	General awareness of the data issues and needs for business decisions.	There is active participation and acceptance of the principles, structures and roles required to implement a formal Data Governance Program.	Data is viewed as a critical, shared asset. There is widespread support, participation and endorsement of the Data Governance Program.	Data governance structures and participants are integral to the organization and critical across all functions.
Data Quality	Limited awareness that data quality problems affect decision-making. Data clean-up is ad hoc.	General awareness of data quality importance. Data quality procedures are being developed.	Data issues are captured proactively through standard data validation methods. Data assets are identified and valuated.	Expectations for data quality are actively monitored and remediation is automated.	Data quality efforts are regular, coordinated and audited. Data are validated prior to entry into the source system wherever possible.
Communication	Information regarding data is limited through informal documentation or verbal means.	Written policies, procedures, data standards and data dictionaries may exist but communication and knowledge of it is limited.	Data standards and policies are communicated through written policies, procedures and data dictionaries.	Data standards and policies are completely documented, widely communicated and enforced.	All employees are trained and knowledgeable about data policies and standards and where to find this information.
Roles & Responsibilities	Roles and responsibilities for data management are informal and loosely defined.	Roles and responsibilities for data management are forming. Focus is on areas where data issues are apparent.	Roles and responsibilities are well-defined and a chain of command exists for questions regarding data and processes.	Expectations of data ownership and valuation of data are clearly defined.	Roles, responsibilities for data governance are well established and the lines of accountability are clearly understood.

Stony Brook Data Governance Maturity Model Initial Results – Spring 2016

Stony Brook Data Governance Maturity Model 2.0 Results Summary

BY FUNCTION

OVERALL (Area Averages)	DATA GOVERNANCE	CULTURE	DATA QUALITY	COMMUNICATION	ROLES & RESPONSIBILITIES
Level 1-Informal	11%	18%	8%	14%	10%
Level 2-Developing	41%	30%	34%	39%	32%
Level 3-Adopted and Implen	11%	14%	9%	14%	3%
Level 4-Managed and Repea	3%	3%	6%	2%	7%
Level 5-Integrated and Optir	0%	0%	2%	0%	0%
Not enough information	34%	35%	41%	32%	48%

Baseline
Dimensions

Current 2015

Target 2017

Maturity

Data Governance
Culture
Data Quality
Communication
Roles & Responsibilities

Integrated & Optimized
Managed & Repeatable
Adopted & Implemented
Developing
Informal

K

51

Elements to change management

Process Representation Deliberation	Executive sponsors Mission alignment Project mgmt./timeline
Initiative	
Problem statement Research/environ. scan Ideas for solutions	Interested allies Interest mapping Advocacy from others

Activity 5 – Assemble your group

Data governance requires support of senior leadership and functional leadership

Identify

> ❯ Senior leaders who will sponsor

> ❯ Functional leaders and their potential for collaboration (includes available bandwidth, interest, capability, willingness)

Case Study – Stony Brook University

Initiative to strengthen university data infrastructure (Jan 2015-Sept. 2016). Charge to examine:

Data governance	Data quality	Communication

Charge for data governance (first 9 months)

Examine existing governance structures

- active and inactive groups and lines of responsibility
- existing processes, practices and procedures that significantly impact data management and stakeholders.

Identify and articulate

- Roles of cross-functional groups
- Functional roles in business units (e.g. data owner, data custodian, report owner) will also be identified and articulated.

Cross functional review group

Draft formal governance structure for data management

- Principles, mission, and goals
- Post on a website to codify roles and responsibilities.

Formalize a process for prioritization

AIR

Charge for data quality improvement

Examine existing practices for ensuring data quality within

- PeopleSoft
- data warehouse
- other functional systems

Articulate and publish practices for developing, maintaining, and communicating

- data definitions (such as robust data dictionaries)
- transparent source information
- update schedules
- error check practices and
- clean-up procedures
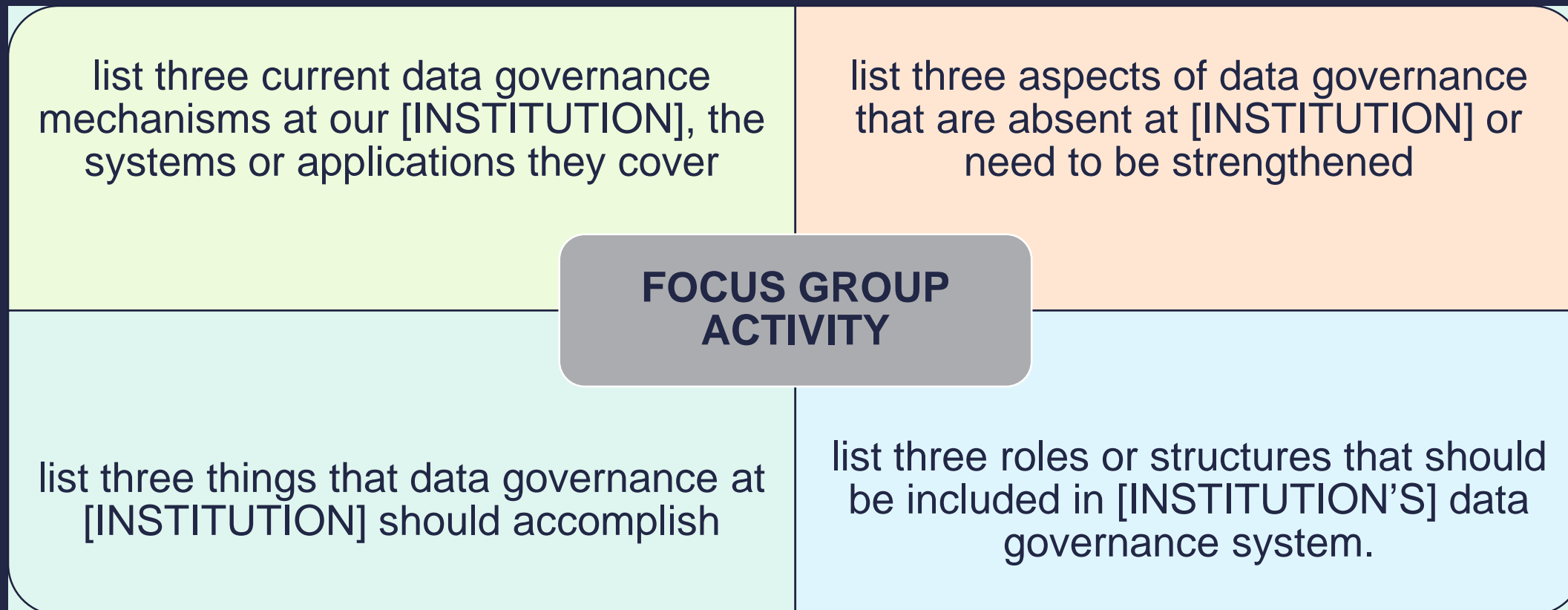
AIR

Charge for training and communication

Develop a communication plan for

- How new capabilities for business intelligence go beyond initial reporting functionality
- Availability
- Use limitations, and opportunities
- including needs identification for documentation, training, workshops, etc.

Develop, document, and adopt reporting standards

Example initial process for data collection

With broadly representative planning group (~20 people), conduct focus group with notecards and flipchart

list three current data governance mechanisms at our [INSTITUTION], the systems or applications they cover

list three aspects of data governance that are absent at [INSTITUTION] or need to be strengthened

FOCUS GROUP ACTIVITY

list three things that data governance at [INSTITUTION] should accomplish

list three roles or structures that should be included in [INSTITUTION'S] data governance system.

AIR

Activity 6 – Draft input for planning process

Using the framework below, draft useful responses to be incorporated into local planning

> list three current data governance mechanisms
> list three aspects of data governance that are absent
> list three things that data governance should accomplish
> list three roles or structures that should be included

[Anticipate responses that may be counterproductive]
> E.g. "IT should control data governance"

Technology applications for data governance

Technology can support data governance	Data dictionary management
	Data quality analysis
	Master data management
	Issue and process management
Technology will not	Build organizational structures, responsibilities, accountabilities
	Mend dysfunctional organizations
	Implement organizational or cultural change

AIR

Selected Data Governance Applications

- Axon (Informatica)
- Collibra
- Data Cookbook (iData)
- Melissa Data
- Oracle Data Quality Middleware
- SAS Data Governance

Issues to consider when selecting technology

Alignment with DG needs

- Metadata management
- Integration w/ reporting tools
- Data quality
- Security/user roles

Initial cost and annual cost
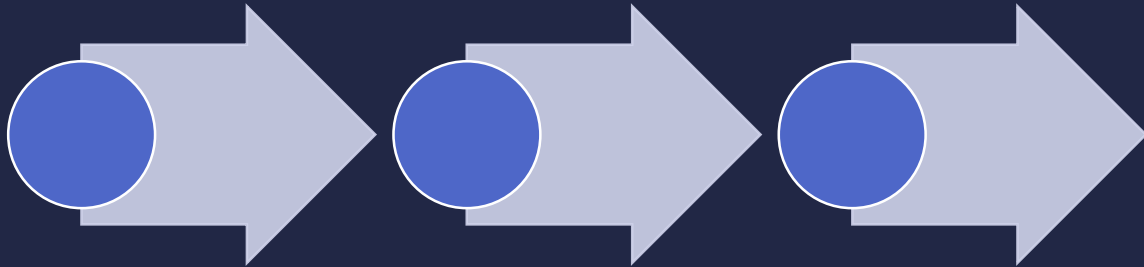
User Community

Ease of implementation and impact on IT

Final Thoughts

Data governance is not a project, it's a process

Project Model
- Linear
- Implies conclusion

Process Model
- Cyclical
- Ongoing

Data governance is only one part of a data strategy

A data strategy is a larger vision for how your organization will work with data.

- Data acquisition
- Data governance
- Data quality
- Data access

- Data usage & literacy
- Data extraction & reporting
- Data analytics

Hosch, B. (forthcoming 2019). "Key elements of a data strategy" in K. Powers ed. *Data strategy in colleges & universities.* New York: Routledge.

Takeaways

- Data governance is more about people than data
- All higher ed change management principals apply
- Process and written documents are essential
  - Leadership support
  - Broad-based consultation, including faculty
  - Opportunity for consultation
  - Representation
- Software can help, but it won't fix broken processes or organizations
- Starting data governance is hard work; sustaining it is harder