Establishing Data Stewardship Models

ECAR

ECAR Working Group Paper | December 18, 2015

ECAR working groups are where EDUCAUSE members come together to create solutions to today's problems and provide insight into higher education IT's tomorrow. Individuals at EDUCAUSE member institutions are invited to collaborate on projects that address core technology challenges and advance emerging technologies important to colleges and universities. More information can be found at the ECAR working groups website.

Introduction

In the current world of enterprise-wide systems, business intelligence, and metrics—as well as the security and privacy concerns that accompany them—administrative data are increasingly recognized as an institutional asset that, because of their strategic importance, need to be managed carefully and used appropriately. Data stewardship provides higher education institutions with a way to carry out this task in a coordinated fashion and to serve their students, faculty, and staff well. Just as physical assets should be managed effectively to provide the best learning and working environments for a campus, so too should the data available to the institution about its services, programs, operations, finances, and facilities be managed to improve understanding, increase efficiency and effectiveness, inform decisions, and support change.

A data stewardship program speaks directly to the university about the responsibilities of data curation and access; it is the bedrock of any good data governance program.[1] Although it may be possible to have data stewards and even a data stewardship program without data governance in place, true data governance cannot exist without data stewards. Indeed, the formation of data governance often begins with stewardship.

This paper provides guidance on establishing a data stewardship program for administrative data. It clarifies the different types of data stewards and managers and their roles and responsibilities, where they reside in an organization, how they work with colleagues to ensure that the data are maintained, and what special skills or training are needed to meet both university responsibilities and best practices.

Defining the Data Steward

You can usually tell a data steward by the kinds of questions they might ask in a meeting, such as "Why are those data important to the institution?" or "How long do you think we should keep that record?" or "What can we do to improve the quality of that analysis?" All of these questions point to someone who is interested in the data environment and the processes that go into capturing, storing, processing, and maintaining the data that are the lifeblood of the university.

One hallmark of a data steward is being the "go to" person with the "tribal knowledge" of how data are collected, maintained, and interpreted. As data stewardship programs mature, formalize, and achieve

EDUCAUSE

alignment with (or spawn) data governance programs, data stewards' expertise is translated into documentation and business rules that facilitate transparency and allow authorized consumers greater self-service to use data as intended, congruent with the data steward's oversight and guidance.

Good data stewards share some common characteristics. For one, stewards listen to and understand issues related to the data processes and quality. They want to do what they can to improve everything that goes into the area of data management. They want to find solutions to improve the overall data landscape, and as such they are willing to guide individuals or groups in their pursuit of better data quality. They understand which resources can best assist with or provide answers to data issues and questions, and they are willing to direct anyone to those resources in the interest of data improvement. As they learn of or identify data-related issues, they look to find a resolution path, whether guided individually or through others. They also understand that not all data issues can be resolved by quick fixes. Sometimes systemic change is required, but a data steward believes that the long-term view of a better data landscape is worth the effort it takes to make the needed changes. The best data stewards possess the ability to see the "big picture" to assess the value of data and the relationships of the data to processes, to other data, and to decision making and outcomes. They also can be extremely detailed oriented, intimate with every data element and how terminology is used to represent the relevant concepts.

Data stewardship is still fairly new as a broad concept in higher education, and the term *data steward* is used inconsistently from one institution to another. In addition, a data steward's responsibilities may vary depending on institution size and type, as well as data maturity (e.g., whether data governance is in place or a formal stewardship program exists). For example, the term *data steward* may refer to an IT staff member who is responsible for maintaining control of access to data, to a staff member who updates particular data elements, to a department director who manages her staff's use of the data, or even to a committee that approves new data values. In this paper we focus on what is essential to responsible data management in this definition of *stewards*:

> **Data Stewards:** Individuals who are responsible for promoting appropriate data use through planning, policy, and protocols at your institution. Data stewards provide university-level knowledge and understanding for a specific data area (e.g., student data, financial data, HR data, or alumni development data). Data stewards are responsible for data quality and data integrity, including consistent data definitions and their application throughout connected systems. They collaborate with other stewards to ensure that overlap areas (e.g., student employees and employees who are students) work across the board and that system updates are scheduled reasonably and tested appropriately. Data stewards work with security, privacy, and compliance officers to ensure that data are classified appropriately and that appropriate training is provided to users who will interact with data. Example stewards are the registrar and HR director.

Data stewards' responsibilities can be grouped into four main areas: operational oversight; data quality; privacy, security, and risk management; and policies and procedures.

- **Operational Oversight**: One hallmark of data stewards is their key role in overseeing the life cycle of a particular set of institutional data. Specifically, data stewards are responsible for defining and implementing policies and procedures for the day-to-day operational and administrative management of systems and data, including the intake, storage, processing, and transmittal of data to internal and external systems. To ensure compliance with data policies and procedures, stewards provide training and documentation for employees with data-entry and maintenance responsibilities. As part of the

oversight for institutional data, the data steward must be accountable to define and document data and terminology in a business glossary.[2] This includes ensuring that each critical data element has a clear definition and is still being used—or retiring those that are not—and that adequate documentation is developed, maintained, and distributed appropriately. Often, operational oversight accompanies being a process owner; in cases where there are multiple process owners, a high degree of coordination is required to ensure that the processes are aligned with data policies and guidelines.

- **Data Quality:** Data stewards are ultimately responsible for establishing data-quality metrics and requirements, including defining the values, ranges, and parameters that are acceptable for each data element. Data stewards work with custodians to establish procedures for detection and correction of data-quality issues and collaborate with process owners to establish policies, procedures, and internal controls affecting the quality of data. In addition, data stewards engage in the ongoing and detailed evaluation of data quality, the identification of anomalies and discrepancies, and the contribution of expertise to understand the root cause and implement corrective measures. This requires tactical resolution of individual unit records such as invalid field values, duplicate person records, incorrectly merged person records, and incorrect identifiers and attributes.
- **Privacy, Security, and Risk Management:** Data stewards are responsible for overseeing privacy, security, and risk management pertaining to data. In one of the more challenging aspects of protecting the data, stewards must establish guidelines and protocols that govern the proliferation of data to ensure that privacy controls are enforced in downstream systems and processes. To be effective, the data steward must compile retention, archival, and disposal requirements and ensure compliance with institutional policy and regulations. Accordingly, the data steward will establish and implement data-curation practices to ensure that the life span of data is commensurate with requirements.[3]

  However, data stewards must protect data while striking a balance between transparency and privacy. This requires establishing information security requirements, including data classification and identification.[4] In addition, data stewards must be knowledgeable in regulatory and compliance requirements relevant to the data domain to evaluate risks to the confidentiality, integrity, or availability of the data based on an in-depth understanding of processes and the likelihood and impact of adverse outcomes.
- **Policies and Procedures:** Data stewards define policies and procedures for access to data, including the criteria for authorization based on role and/or the individual. Responsibilities include oversight of the access request, approval, provisioning, and deprovisioning processes to ensure they are appropriate and commensurate with risk. Working closely with data custodians to establish incident-detection controls, stewards evaluate any suspected or actual breaches or vulnerabilities in confidentiality, integrity, or availability and report them to management or information security personnel.

Regulatory Statutes

A data steward may need to be familiar with regulatory statutes, depending on the area with which he or she is associated. Two common statutes are:

- **FERPA:** The Family Educational Rights and Privacy Act serves as the basis for protecting educational records.
- **HIPAA:** The Health Insurance Portability and Accountability Act is key to working with clinical and health insurance data.

In summary, effective data stewards make data available to the institution, thereby playing a key role at the heart of collaboration, supporting institutional research, assessment, and analytics efforts that involve the domain data. In fact, because this role may also be critical to the success of business intelligence and analytics programs, the resource needs for data steward support of various initiatives should be carefully considered, evaluated, and aligned with institutional priorities.

Data Stewardship Skills

In addition to professional traits that a good data steward should possess, other skills are needed. Some of these are common to all domains, and some are specific to certain areas of expertise.

- **Relationship Management:** A data steward must manage and maintain relationships with numerous data stakeholders across the institution.
- **Facilitation:** Stewards should have the ability to successfully facilitate meetings, reviews, and working sessions.
- **Communication:** Data stewards facilitate communication regarding business process changes that may affect downstream systems or analytics relating to specific data elements. The ability to communicate well with individuals and groups, both in writing and verbally—and with both business and technical colleagues—is essential in conducting good data stewardship.
- **Process Definition:** The ability to understand, define, and document process as it relates to the data is a core skill for data stewardship.
- **Change Management:** To make improvements to data quality, change is often in order. The ability to understand and facilitate that change is necessary.
- **Problem Solving:** The ability to understand the root cause of data issues requires an understanding of the associated problems. Being able to think through the issues and devise a solution is a key asset for a data steward.
- **Policies:** The ability to develop, apply, and communicate institutional policies that are specific to the area that the steward governs should be second nature.

Finding the right person with the appropriate skills will go a long way in making the steward as effective as possible. Given the proper skill set, a data steward can be a valuable asset.

Where Data Stewards Are Found

Although the person who is responsible for the overall data stewardship program may have a title to reflect that role, the term *data steward* is typically a designation given to managers and key staff with other job titles and responsibilities.

Data stewards can represent many areas within the institution, and responsibilities may vary by area depending on policy, legal, and business implications. Recognizing this diversity and flexibility is key to making certain that the institution has sufficient coverage to meet its stewardship goals. The following outlines the various areas where data stewards may be needed and what kinds of data are typically managed in those areas.

Student Data

- **Admissions:** Admissions data deal with prospective students, individuals who have submitted applications to the institution, and persons who have been accepted for admittance to the university. Data in this area may have confidentiality issues and include familial relationships, financial information, previous academic history, and other details that allow the institution to make choices about which applicants to admit.
- **Housing:** Housing data generally deal with dorms or other facilities where students live on or off campus. University policy tends to govern this area, given that ensuring the privacy and safety of students and others who may be staying in campus housing is critical. Stewards who work with these data may also work with information tied to other campus services but related to housing, including meal plans and options.
- **Registrar:** Registrar data includes information related to enrollment and registration, such as programs, curricula, classes, classrooms, grades, completions, requirements for graduation, and student records. FERPA requirements are heavily applicable, as are personal privacy regulations.
- **Student Health:** This area of stewardship tends to be centered around student medical records. The need for confidentiality is extremely high, and HIPAA and FERPA regulations are enforced.
- **Bursar:** This area focuses on payment information, generally relating to student accounts. Since these are financial transactions, stewardship of these data is important to the institution. Many of the transactions will be conducted electronically, so PCI[5] (payment card industry) regulations are important. Stewards of this data should be familiar with these types of guidelines and be able to ensure that audit procedures related to the data are enforced.
- **Financial Aid:** Eligibility, applications, awards, and appeals are just some of the types of data contained in this area. The need for stewardship for financial aid data is extremely high, given the scrutiny and audit that accompany this function. In addition, frequent changes to financial aid guidelines and reporting requirements necessitate that data are maintained at the highest quality levels.
- **Athletics:** Depending on the institution, data maintained in this area can range from informal intramural activities to highly competitive and regulated divisional sports. Scholarships, sponsorships, facilities, and professional stipends may all be tracked, and the resulting data can be voluminous, auditable, and, in some cases, public. Stewardship of these data is critical, particularly given the visibility and importance that athletics may have on the reputation and marketability of the institution.
- **Learning Analytics:** Learning analytics includes data from learning management systems and may also integrate data from other sources (e.g., student aptitude and demographic data). Stewardship in this area might need to address what data may and may not be used and by whom, as well as specific compliance issues, among others.[6]

Academic Data

- **Provost Office:** Data in this area relate to professional and academic decisions regarding faculty. Included in this data domain are tenure considerations and decisions, promotions, academic discipline, and hiring information. These data may be highly confidential in some institutions, while in others they might be a matter of public record.

- **Faculty Affairs:** Divisions, schools, and other distributed units often manage the academic appointment process, oversee recruiting efforts for new faculty, and provide support to faculty in the development of their academic and professional careers. These data are highly confidential and are managed with a high regard for personal and professional standards.
- **Academic Administration:** Data in this area encompass a broad swath, including administrative and academic data related to colleges, schools, divisions, affiliates, and other units of a university. Since this data may include information about students, faculty, programs, and related academic information, FERPA rules apply. In addition, academic administration may include personnel data, so PII policies are enforced. Academic policies relevant to the individual institution would be brought into focus as well, so stewards in this area need to have a broad knowledge of data governance.
- **Library:** Responsibilities for this area can vary widely, so stewardship activities will need to be adjusted accordingly. At a base level, data about the institution's print, digital, audio, and other holdings are maintained, along with information about usage of those assets. Assuming the library is providing an archival function, data about items contained in the repository would also be available. Beyond those functions, additional data may be maintained on interlibrary assets, research activities, special collections, and activities that the library sponsors or is collaborating with. Much of the data contained are expected to be public facing; however, some may be confidential (e.g., patron logs[7]) or reserved for institutional purposes.

Alumni Data

- **Alumni Relations and Development:** Data in this area deal with development activities for the institution. Confidential information is prevalent, since these data are heavily concerned with prospects, donors, and gifts. Data policy in this area tends to be focused on the care of these highly sensitive data, and institutions take the handling of these data very seriously.

Administrative Data

- **Financial Services:** This area of responsibility can vary by institution but generally involves data related to administrative financial transactions, statements, and balances. Included in this are general-ledger, accounts-payable, accounts-receivable, and fixed-assets data. It may also include payroll data in some institutions. Given the sensitive nature of this type of data, particularly when payroll data are included, stewardship is extremely important. There are many legal and policy statutes around these areas, and, given audit and accountability concerns, data consistency and accuracy are paramount.
- **Institutional Finance:** Data in this area are primarily focused with budgets, investments, and other strategic financial areas of the institution. Since these data can be highly confidential and regulated, stewardship is essential. Knowledge of regulations and transmission protocols is important.
- **Institutional Research and Assessment:** The stewards in this area focus on data that get added to official reports and aggregate counts (e.g., factbooks), not any individual departmental data. Reporting may be cross-college and may include IPEDS reporting, data for *US News & World Report*, and accreditation reporting. Business intelligence is included in this area.

- **Facilities:** Facilities is a broad area that can encompass many data domains. Included may be information related to physical spaces (buildings, rooms, green spaces, parking lots, etc.), building projects, work orders, staffing, inventory, physical assets, future building projects, and many other areas. Sensitivity ranges from public data to highly sensitive and confidential information. Because of this wide range, facilities data stewardship requires a clear understanding of institutional needs and what data may and may not be shared.
- **Human Resources:** Data in this area fall in many areas, including payroll, benefits, time management, personnel management, and possibly recruiting and training. Given the highly sensitive and confidential nature of the data in this domain, stewardship is essential. An understanding of laws, regulations, policies, data management best practices, and management of access to the data are all aspects of the stewardship role in this area.
- **IT:** IT tends to be the custodian of data created and processed by other areas. The data domains that IT typically stewards revolve around infrastructure services, systems, and processes that are owned or managed by IT itself, such as the service catalog, IT service management, backup/recovery, and disaster recovery.
- **Information Security:** Information security data stewardship tends to focus on policies, best practices, guidelines, logs, pen test results, incident response, and compromises. Some data, such as information regarding compromises and logs, are highly confidential and require close stewardship, while other areas, such as best practices and policies, should be public and shared with the community.
- **Legal/General Counsel:** This area is unique in that there is likely limited data to be stewarded; however, the ability to have legal counsel provide expertise on policies, practices, and other data-related activities is invaluable to the stewardship process.
- **Safety and Security:** Data stewardship in the institution's safety office is extremely important, in that life, health, and safety data are critical and confidential. The data in this area can be of many varieties—information may be related to police activities, such as arrest records, parking violations, and other law enforcement duties. Confidential data from federal and state agencies related to public safety, known offenders, warnings, and other security information may be housed or accessed in this area. Safety information related to inspections, hazardous material, required training, and similar items might all be maintained.

Research Data

Research data include administrative data associated with research, as well as the data generated by research.

- **Office of Sponsored Research:** This area focuses on data related to proposals, grants, contracts, and other federal, state, corporate, or privately funded research activities. The data maintained are typically for preaward proposals, awarded funds, postaward review, and accounting related to the research activities. Sensitive data are typically contained in this area, and, due to the fact that this area is regulated and audited on regular intervals, stewardship is critical to the institution and the research activities that are covered.

- **Institutional Review Board:** The data governed in this area relate to research protocols, primarily those administered for human subjects. Medical and administrative record data are the primary focus, and because these data are typically audited by federal agencies, strict governance is required.
- **Academic Research:** Although academic research is not the primary focus of this document, we recognize that data in this area require stewardship. Research in this area is defined as anything that has a principal investigator, including medical research, science research, digital humanities, and others. The focus here is on the actual data coming out of research.

Developing a Data Stewardship Program

To optimize services and operations, and to be competitive, universities must have an effective system to ensure authenticated and reliable data, enterprise-wide compliance, technical interoperability, and a common, informed mind-set that values data as an institutional asset for advancing the university's mission.

As technology advances, more people can easily create, store, and transmit growing volumes of data and digital collections. It is imperative to have a structure in place to manage and protect data assets, ensuring reliable and timely access to accurate data, within a framework that provides built-in privacy and security safeguards and data-management and sharing capabilities that meet federal mandates.

Some data stewardship programs will focus on data policy. Others may focus on data definition and workflow, and still others on risk management. Regardless of the focus, most data stewardship programs will share four essential first steps:

1. **Assign Responsibility:** Confer the designation of data steward on managers of operational areas that are known to collect, create, or manage data that are vital to university operations or that pose a risk if compromised.
2. **Define the Data Steward:** Clarify in written policy or procedure exactly what the data steward designation entails. Regardless of the particular function or operations the manager oversees, consistent stewardship responsibilities, activities, and duties of care apply to vital data.
3. **Identify the Data Stewards:** Publish a roster of data stewards to clearly establish the point person to be consulted by functional or operational area.
4. **Specify Stewardship Coordination:** Develop a formal specification of the ways that data stewards are expected to collaborate and cooperate with each other for the betterment of the university's data.

Beyond these first steps, the institution should have a clear understanding of the key components of the stewardship program.

Data Stewardship Program Components

In order to fully understand how to develop an effective data stewardship program, it is helpful to break down the components of such a program into the four Ps: purpose, people, policy, and practice.

Purpose

One critical advantage for any institution embarking on a path toward comprehensive data governance is the opportunity it provides to appraise the existing tangible components that constitute the institution's data assets—identifying what types of assets they are, where they exist, and what risk the institution incurs is if those assets are lost or compromised at any phase of their life cycle. As stated in the Confidential Data Handling Blueprint,[8] an initial institution-wide risk management program enables the institution to gather information about the types of data and conduct a risk assessment of those assets. Understanding the types of data that an institution creates, obtains, and stores is the first step in helping determine how those data should be handled and managed across their lifetime.

Rather than approach this topic only from the perspective of managing threats, it is important to remind the campus community of the value of all data held by the university and how effective management of the data can realize potential benefits from academic and research activities. While an effective data stewardship program helps prevent the misuse of data by those without appropriate access rights, it also helps ensure that the aggregation and analysis of high-quality and officially recognized data elements produces publication-ready data.[9]

Data stewardship programs can serve several key objectives:

- Ensure that institutional data are reliable, consistent, and of high quality
- Ensure that institutional data are accessible for appropriate purposes, people, and systems
- Ensure that institutional management practices comply with federal and state legislation (e.g., HIPPA, FERPA) as well as with industry regulations (e.g., PCI), especially regarding privacy and security
- Ensure that institutional data management practices comply with standards and best practices in higher education (e.g., NCAA, IPEDS,[10] accreditation agencies, conventional uses)

A sound stewardship structure increases the institution's ability to efficiently deliver data and related reports that are reliable, consistent, and of high quality. Policies and procedures ensure that users have access to the specific types of data needed based on their individual responsibilities, while not exposing additional data that are not needed for a given analysis or reporting function. Guidelines and ongoing training provide all users with clear instructions governed by roles and responsibilities of all those who have appropriate access. Process and policy components resulting from oversight of a data governance program ensure workflow efficiency and data security are standard when interfacing with other established systems across campus. A thorough understanding of data stewardship at the institutional level also ensures that requirements by additional external entities such as state systems and federal funding agencies are being met.

People

All higher education institutions have managers charged with the oversight of key operational areas under administration, finance, human relations, student services, and many other organizational units. Almost without exception, position descriptions and job duties for such managers include responsibility for managing human resources, but only in recent years have some of those same position descriptions also addressed similar responsibility for managing data and information, despite the fact that these various organizational areas almost assuredly collect, produce, or manage data. At many institutions, data

responsibility and duty of care may only be implicit; some managers may take active responsibility for data because doing so naturally aligns with their other job duties. To ensure that data are being properly cared for, however, in alignment with and for the furtherance of institutional goals, a formal stewardship approach—ideally, through an institution-wide program—is necessary.

A data stewardship program may also consist of the roles shown in figure 1 to support the data steward (note that not all of these may not exist, depending on the size and complexity of the institution).

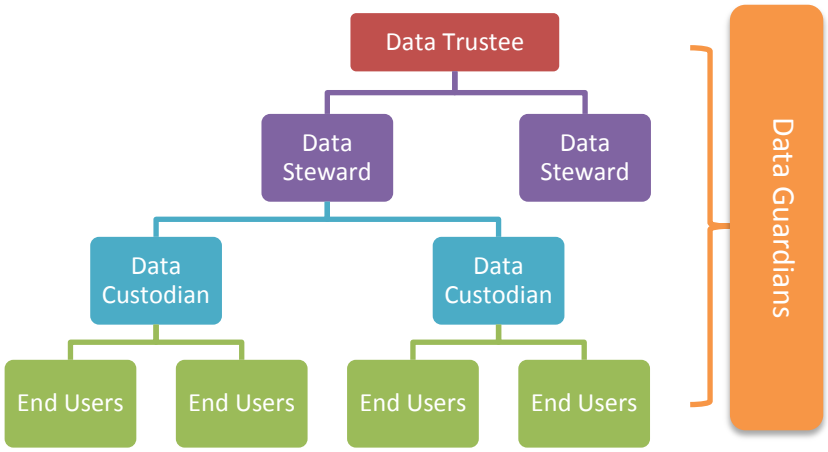

Figure 1. Data stewardship and support roles

- **Data Guardians:** Data guardians are supportive, but parallel, to the rest of the roles outlined here. They are typically made of technology and personnel responsible for monitoring activities affecting data at rest and in transit and for implementing access and security controls required by data stewards based on data classification. Example guardians include personnel in networking, data center operations, systems administration, and information security.
- **Data Trustees:** Trustees are the highest-ranking individuals accountable for what happens with and to data at your institution. They might be members of the highest oversight organization to which issues are escalated, and they likely have strategic-planning and policy-setting authority. Data trustees provide a broad, university-wide view of data, approve policies, resolve questions of procedure, and ensure that data plans are consistent with and in support of university strategic plans. Data stewards typically fall under trustees but above custodians. Example trustees include vice presidents and vice provosts.
- **Data Custodians:** These individuals are responsible for ensuring that policies are followed within specific user areas (e.g., colleges, departments, and administrative offices) and that local procedures are consistent with university policies and procedures. Example custodians include departmental HR managers and academic advisors.
- **End Users:** End users include the people, organizational units, and information systems that are granted access to data for specific uses. Among many examples, end users may include staff tasked with entering data and managing the records of students, employees, financial transactions, etc.; persons granted access to data for analysis and reporting; or downstream information systems such as parking or residence hall management programs that ingest and/or transform data for a specific

purpose. End users should clearly understand their responsibilities with data entrusted to them. Example end users include HR administrators and registration processors.

In addition to the important relationships they have with data stewards, these designees often have additional relationships with individuals and offices that may not themselves be data stewards. These vital relationships include:

- **Institutional Research Office:** Responsible for official reporting of data from multiple operational and functional areas
- **Institutional Archivist:** Responsible for the official schedule of records retention, archive, and destruction, regardless of format or medium (electronic data, images, film, paper documents, etc.)
- **Institutional Audit and Risk Management Personnel:** Responsible for ensuring policies and procedures are followed
- **Data Governance Personnel:** Responsible for ensuring coordination among data stewards and others with duties of care for data (e.g., chief data officer)
- **Information Security Office:** Responsible for the implementation and execution of systems and services to protect vital data from unauthorized access and use
- **Data Center/Information Technology Unit:** Responsible for the technical administration, support, and maintenance of information systems that collect, create, manage, or distribute data, as well as hardware, databases, and networking capabilities (including data warehouse)

Each of these relationships could be classified under one or more roles (guardian, trustee, steward, custodian, or end user).

Policy and Practices

Data stewardship is concerned with the exercise of authority, control, and shared decision making over the management of data assets. Policies and best practices are necessary to ensure consistency in the activities of the stewards.

Data stewards often work to develop procedures that define how policy is implemented, e.g., what access can be granted to specific data based on that data's classification. Other common policies may focus on data quality, performance metrics, data usage, or data retention. Stewards may develop the procedures, but typically data custodians will carry out the responsibility of following those procedures.

Goals

Data stewardship policies must be tailored to each institution's needs. Common goals include:

- Improve data quality
- Improve data consistency
- Reduce operational friction by improving the organization's ability to share data between units (reduce costs, improve effectiveness)
- Support the needs of data stakeholders

- Build and implement transparent, standard, repeatable processes
- Comply with internal and external regulations and standards for data usage
- Manage change in the environment over time

Representation

With a set of goals in hand, organizations can then begin to develop specific policies. Enterprise data policy development requires broad representation from across the institution, including functional groups (units that serve core business functions and manage data that are used across the institution), risk management/general counsel/audit, IT, and research.

Scope

Examples of policy scoping issues include:

- **Data:** What data are covered by the policy and for what purposes?
- **Data Classification:** While an organization's needs may extend beyond these classifications, most organizations need at least three:
  - Public: Free to all.
  - Protected: Unauthorized release/modification/etc. This classification is typically for data that have ethical and/or privacy considerations.
  - Confidential: Unauthorized release/modification/etc. This classification is typically for data with specific legal ramifications.
- **Roles:** Define roles and responsibilities with regard to data management within the organization.
- **Access Rights:** Generic access rights include administrative, read, write, and delete. These will need to be determined and will vary by functional area and role and will be constrained/directed by underlying technologies.
- **Retention Periods:** These periods will need to be determined and will vary.

Metrics

As data stewardship efforts mature within an organization, metrics become valuable for measuring success and directing future developments. Data metrics may be developed to measure:

- Accuracy
- Compliance
- Completeness
- Consistency
- Utility
- Efficiency
- User satisfaction

Program Paths and Stages

It seems a daunting task, but many paths can be taken toward a data stewardship program. Administrative units often have business practices and policies on how to handle sensitive data—from auditing to accessing, manipulating, and reporting. A strongly recommended first step is to review the current cultures, practices, and policies. This will help establish the roadmap showing where your institution has already made progress, where it lacks depth, and what you need to do to establish a complete program.

Based on culture, available resources, and the maturity of your institution's data management, one or more of the following routes may provide an appropriate way to build out a data stewardship program that is sustainable and effective. All three capitalize on resources and efforts that require many of the same conversations and decisions.

- **Align the Program with Compliance and Risk Objectives:** If your institution is focusing on securing data (encryption of devices or data at rest, in transit, for reporting, etc.), then partnering with the committee or department with this charge is an excellent option.
- **Align the Program with a Product Rollout:** If your institution is implementing a new reporting tool, a new payroll system, or some type of software that involves data integration and access, you can use it as microcosm to model a stewardship program. Include it as a deliverable in the implementation and then raise awareness as the product is deployed.
- **Align with Other Campus Initiatives:** Initiatives such as those involving document life cycles or contracts management typically involve building approval workflows and identifying owners of certain steps or processes and can tie in nicely with data stewardship goals.

Regardless of your institutional environment, building a campus-wide data stewardship program is an ongoing process with recognizable stages that are evolutionary in nature and help indicate the maturity of a data stewardship program. Some institutions may find that they need to progress through all the stages, while others may move more quickly and skip stages.

Stage 1: Locate Siloed and/or Independent Data Stewards

In this early stage, an institution will typically start by identifying individual data stewards for business units that approve access to data. Typically this includes roles such as the registrar, comptroller, and vice president of HR. Policies and procedures may or may not be formalized. Stewards in this stage may or may not meet and collaborate regularly.

Stage 2: Set Up Initial Collaboration and Cooperation

During this stage a data steward task force may be established that comprises the individual data stewards plus others who are involved with data and its protection (e.g., IT, library, legal, risk/compliance office, institutional research). The task force should meet regularly, perhaps biweekly or monthly, and have access to a tool or procedure to document findings collaboratively. Suggested goals are:

- Create a document that lists existing policies and procedures, identifies and classifies risks, and identifies business units or groups with data responsibility and little or no stewardship activity.
- Identify paths to raise awareness.

- Organize subgroups to focus on needed changes and set the objectives to accomplish those changes. Each subgroup should include a task force member and persons whose interests coincide with the needed changes. The duration and frequency of the subgroup meetings may depend on the group's assignment, but monthly updates should be provided to the task force.
- Develop a proposal for common data policies, practices, and definitions, as well as task force objectives and general structure. A comprehensive vision can then be presented with metrics and observed impacts to senior management along with a recommendation to establish a permanent governing body.

Stage 3: Obtain Leadership and Official Recognition

This stage marks the initial formation of a data stewardship program. It begins by gaining institutional buy-in of the importance of having a stable governing body with the authority to approve university data policies and, when needed, review local data policies. Breadth of participation in the subgroups in stage two will make this easier.

This stage can take on many forms. A chief data officer might be designated to manage the program, or a permanent data stewardship committee might be created to take on that task. The critical outcome is that policy and procedures exist for defining and approving data access, definitions, classification, preservation, auditing, encrypting, etc.

At this stage, the core infrastructure (data stewardship group and ground rules for how they interact, policies, and an institutional approach) is established.

Stage 4: Expand Stewardship Maturity and Sustainability

This stage is reached once a formal stewardship program has been up and running and can point to some clear accomplishments. At this stage, the scope of the data stewardship program expands to include aspects that may have been overlooked or put on hold earlier, such as institutional analytics or ancillary systems such as gym membership or event management. In addition, a constant environmental scan will likely be necessary at this stage to ensure practices and policies scale with new technologies and demands.

Conclusion

Data are an important asset, just like cash and other physical assets. To maximize the benefit and minimize the risk of data assets, many universities are establishing data governance programs, central to which is the role of the data steward. Enabling successful data stewardship is the key to an effective data governance program and ultimately to the effective use of institutional data assets. If your data stewards are successful, your data governance program will be successful.

Establishing a data governance program consists of three tasks: defining data steward responsibilities, dividing institutional data into segments, and assigning the segments to individual data stewards. The purpose of everything beyond this is to help the data stewards succeed in their stewardship. Oversight councils help data stewards resolve shared data domain issues. Principles, policies, and procedures institutionalize stewardship decisions. Data government offices facilitate stewards and oversight councils.

Additional Resources

For more information on data stewardship and data governance:

- Albrecht, Robert, and Judith A. Pirani. *Revitalizing Data Stewardship through Risk Reduction: Managing Sensitive Data at the University of Virginia*. ECAR case study, December 1, 2009.
- Blair, Douglas, et al. *The Compelling Case for Data Governance*. ECAR working group paper, March 17, 2015.
- Chapple, Michael J. "Speaking the Same Language: Building a Data Governance Program for Institutional Impact." *EDUCAUSE Review*, December 6, 2013.
- Kelly, Michael C. "The Chief Data Officer in Higher Education." *EDUCAUSE Review*, June 8, 2015.
- Ladley, John. *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program* (Waltham, MA: Elsevier, 2012).
- Plotkin, David. "Common Data Stewardship Issues You'll Need to Deal With, Part 1." *Dataversity*, October 25, 2011.
- ———. "Common Data Stewardship Issues Part 2." *Dataversity*, November 8, 2011.
- ———. *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance* (Waltham, MA: Elsevier, 2014).
- DAMA (Data Management Association) International
- DAMA Body of Knowledge
- The Data Administration Newsletter
- "Data Classification Toolkit." *HEISC Information Security Guide*, July 2015.
- Data Governance Institute
- Data Governance Professionals Organization
- EDUCAUSE Library:
  - Data Administration and Management
  - Data Governance
- "Top Information Security Concerns for Campus Executives & Data Stewards." *HEISC Information Security Guide*, September 2010.

Authors

Special thanks go to the following ECAR-DATA Working Group members who authored this report.

Nickolas Backscheider
Associate Executive Director, OIT
Auburn University

Douglas Blair
IT Consulting Engineer
Carnegie Mellon University

Kelly Briner
Director, Data Governance
Arizona State University

Vaibhav Dani
Assistant Director, Database Architecture
Old Dominion University

Michael Fary (Co-Chair)
Enterprise Data Architect
University of Chicago

Jason Fishbain
Chief Data Officer
University of Wisconsin–Madison

Sherri Flaks
Director, Enterprise Data & Analytics
The Johns Hopkins University

Michael C. Kelly (Co-Chair)
Chief Data Officer
University of South Carolina

Kathryn Matuch
AVP, Core Enterprise Systems, IRT
Drexel University

Kim Owen
Advanced Applications Outreach
North Dakota State University

Citation for This Work

Backscheider, Nickolas, et al. *Establishing Data Stewardship Models*. ECAR working group paper. Louisville, CO: ECAR, December 18, 2015.

For more information about this paper and ECAR working groups, contact Karen A. Wetzel, Program Manager, EDUCAUSE.

Notes

1. As was detailed in Douglas Blair et al., *The Compelling Case for Data Governance*, ECAR working group paper, March 17, 2015, effective data governance is the key to effectively managing an institution's data.
2. A business glossary is "a software application used to communicate and govern the organization's business concepts and terminology along with the associated definitions and relationships between those terms." From Lowell Fryman, "What Is a Business Glossary?" BeyeNETWORK, September 13, 2012.
3. Clifford A. Lynch, *Big Data in the Campus Landscape: Curation*, ECAR working group paper, November 20, 2015.
4. To learn more, see "Data Classification Toolkit," *HEISC Information Security Guide*, July 2015.
5. See PCI Security Standards Council.
6. To learn more about learning analytics data, see ECAR-ANALYTICS Working Group, *The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success*, ECAR working group paper, October 7, 2015.
7. For more about library data and patron privacy, see "Consensus Framework to Support Patron Privacy in Digital Library and Information Systems," National Information Standards Organization.
8. See "Confidential Data Handling Blueprint," *HEISC Information Security Guide*, September 2009.
9. Blair et al., *The Compelling Case for Data Governance*.
10. See IPEDS.