COLLEGE OF CHARLESTON

OFFICE OF THE CIO

Data Principles

This set of principles address data quality/integrity, security, and accessibility.

Note: The framework of these principles (Statement, Rationale, and Implications) and some of the phrasing was taken from the TOGAF 9.2.
(The TOGAF standard is an architecture framework. It provides the methods and tools for assisting in the acceptance, production, use, and maintenance of an Enterprise Architecture. It is based on an iterative process model supported by best practices and a re-usable set of existing architecture assets.)

Guiding Principles: 2
Supporting Resources: 2
Data Management Standard: 2
Data Classification Standard 2
Definitions: 2
Data Governance Council 2
C.I.A Triad 3
Confidentiality: 3
Integrity: 3
Availability: 3
Institutional Data 3
Information Domain 3
Principles 4
Principle 1: Data are Assets 4
Principle 2: Data are shared 4
Principle 3: Data are Accessible 5
Principle 4: Data follow CofC Data Management standards 6
Principle 5: Data have a Common Vocabulary and Data Definitions 7
Principle 6: Data follow CofC Data Classification standards 7
Principle 7: Data will be Analyzable 8
Principle 8: Data are the enterprise memory of service delivery 9
Principle 9: Data update should be the natural result of automated and/or self-service delivery 9
Principle 10: Compliance with Law 10
Principle 11: Training 10

Guiding Principles:

- Information is one of the College of Charleston's most valuable resources and as such requires responsible management and use by all members of the College community. The College seeks to enable data sharing in a way that allows appropriate use to drive decision-making while at the same time protecting the security and privacy of information.
- P.A.I.R.
    - Personalized Experience
    - Automation
    - Integration
    - Reporting/Results

Supporting Resources:

Data Management Standard:

A data management structure is required at the College to ensure proper handling of **institutional data**. See Data Management Standard.

Data Classification Standard

All **institutional data** will be classified under this standard and used with appropriate and relevant levels of access and with sufficient assurance of its security, privacy, and integrity in compliance with applicable laws, rules, and regulations. This standard describes the classification schema for **institutional data** and is based on the State of South Carolina's Data Classification Schema to categorize institutional data so the College knows who can access the data, where the data is located and prevent the unauthorized use or disclosure of information.

Privacy Policy

The College is committed to maintaining the privacy, integrity, security, and availability of confidential information created, received, maintained and/or stored by it, regardless of form. This Policy explains the obligations of all members of the College community to protect non- public information and records from unauthorized use or disclosure. It is designed to address applicable federal and state law governing privacy and confidentiality of information, as well as any applicable international privacy regulations. This Policy speaks to principles such as need-to-know collection of and access to information, record retention, acceptable use, information access safeguards and the South Carolina Freedom of Information Act and exemptions.

Definitions:

Data Governance Council

Made up of key campus stakeholders that includes all data stewards and many data managers. Responsible for coordination of enterprise data management activities including:

- Establishment of standard definitions and documentation of data relationships for administrative data, ultimately leading to an enterprise data model.
    - Update and maintain currency of the Data Dictionary
- Establishment and implementation of data standards, including those ensuring security of sensitive data

- Implementation of consistent processes for collecting, matching, aggregating, quality-assuring, securing, and distributing data throughout the organization to ensure consistency and control in the ongoing maintenance and use of the data.
- Resolves issues related to:
    - process flow
    - data integrity assurance
    - data coordination across divisional lines
    - data definitions
    - record retention.

Recommend institutional descriptive and diagnostic dashboards to support decision making

C.I.A Triad

Confidentiality:

Data confidentiality is about protecting data against unintentional, unlawful, or unauthorized access, disclosure, or theft and means access to data are restricted to authorized users. Protecting confidentiality is dependent on being able to define and enforce certain access levels for information. In some cases, doing this involves separating information into various collections that are organized by who needs access to the information and how sensitive that information actually is - i.e., the amount of damage suffered if confidentiality was breached.

Integrity:

Integrity means an assurance that data are reliable for their intended use, accurate, complete and up-to-date. This is designed to protect data from deletion or modification from any unauthorized party, and it ensures that when an authorized person makes a change that should not have been made the damage can be reversed. This also provides an opportunity for authorized users to review information and ensure that data is accurate, complete and up-to-date.

Availability:

Availability means data are accessible by authorized users as needed to perform their authorized functions. Authentication, authorization mechanisms, access channels and systems work together to protect and ensure it's available when it is needed.

High availability (HA) systems are the computing resources that have architectures specifically designed to improve availability. Based on the specific HA system design, this may target hardware failures, upgrades, or power outages to help improve availability, or it may manage several network connections to route around various network outages.

Institutional Data

By default, all institutional data follows the data classification policy. College employees will have access to these data for use in conducting College business. Access permission to view or query institutional data will be based on legitimate need and role of the data user.

Information Domain

A three-part concept for information sharing, independent of, and across information systems and security domains that 1) identifies information sharing participants as individual members, 2) contains shared information objects, and 3) provides a security policy that identifies the roles and privileges of the members and the protections required for the information objects.

Principles

Principle 1: Data are Assets

Statement:

- Like personnel and facilities, data are a highly valuable asset to the College that must be managed accordingly.

Rationale:

- Data are valuable college resources and has real, measurable value. Accurate, timely data are the foundation of our decision-making. We must carefully manage data to ensure that we know where it is, can rely upon its accuracy, and can obtain it when and where we need it and know that it is protected from unauthorized use or disclosure.

Implications:

- The first of three data principles, which are closely related: data are an asset; data are shared; and data are accessible. There is an education task to ensure that all stakeholders within the College understand the relationship between value of data, sharing of data, and accessibility to data. This should include an opportunity to provide notice and transparency to users about the purpose for collection and ensure responsible use of information.
- Data stewards must have the authority and means to manage the data for which they are accountable.
- We must make the cultural transition from "data ownership" thinking to "data stewardship" thinking.
- The role of the data steward is critical because obsolete, incorrect, or inconsistent data could be passed to college personnel and students, and adversely affect decisions across the enterprise.
- Part of the role of the data steward, who manages data, is to ensure data quality/integrity. Procedures must be developed, documented, and used to prevent and correct errors in the information and to improve those processes that produce flawed information.
- Data quality will need to be measured and steps taken to improve data quality – it is probable that policy and procedures will need to be developed for this as well.
- The Data Governance Council, with comprehensive enterprise-wide representation, should decide on process changes suggested by the steward.
- Since data are assets of value to the entire college, data stewards, who are accountable for properly managing the data, must be assigned at the enterprise level.

Principle 2: Data are shared

Statement:

- Users have access to data necessary to perform their duties; therefore, data are shared across the College functions and organizations on a need-to-know basis.

Rationale:

- Timely access to accurate data is essential to improve the quality and efficiency of college decision-making. It is less costly to maintain timely, accurate data in a single application, and then share it, than it is to maintain duplicate data in multiple applications. The College holds a wealth of data, stored in over 150 systems. The speed of data collection, creation, transfer, and assimilation is driven by the ability of the organization to efficiently share/integrate these systems across the College.
- Shared data will result in improved decisions since we will rely on fewer (ultimately one) sources of more accurate and timely managed data for all decision-making. Electronically shared data will result in increased efficiency when without re-keying, to create new entities.

Implications:

- The second of three data principles, which are closely related: data are an asset; data are shared; and data are accessible. There is an education task to ensure that all stakeholders within the College understand the relationship between value of data, sharing of data, and accessibility to data. This should include an opportunity to provide notice and transparency to users about the purpose for collection and ensure responsible use of information.
- To enable data sharing we must develop, document, and abide by a common set of policies, procedures, and standards governing data management and access for both the short and the long term.
- Effective data sharing and systems integration promotes improved processes through automation and triggered actions.
- For the short term, to preserve our investment in legacy systems, we must invest in software capable of migrating legacy system data into a shared data environment.
- Using the Data Cookbook, we will develop standard data models, data elements, and other metadata that defines this shared environment. The Data Cookbook is a central repository system of data definitions and a data governance tool. For the long term, as legacy systems are replaced, we must adopt and enforce common data access policies and guidelines for new application developers to ensure that data in new applications remain available to the shared environment and that data in the shared environment can continue to be used by the new applications.
- For both the short term and the long term we must adopt common methods and tools for creating, maintaining, and accessing the data shared across the College.
- This principle of data sharing will continually come into conflict with the principle of data security —under no circumstances will the data sharing principle cause confidential data to be compromised.
- Data made available for sharing will have to be relied upon by all users to execute their respective tasks. This will ensure that only the most accurate and timely data are relied upon for decision-making. Shared data will become the enterprise-wide, single source of truth.

Principle 3: Data are Accessible

Statement:

- Data are accessible for users to perform their functions.

Rationale:

- Wide access to data leads to efficiency and effectiveness in decision-making and affords timely response to information requests and service delivery. Using information must be considered from an institutional perspective to allow access by a wide variety of users. Staff time is saved, and consistency of data is improved.

Implications:

- The third of three data principles, which are closely related: data are an asset; data are shared; and data are accessible. There is an education task to ensure that all stakeholders within the College understand the relationship between value of data, sharing of data, and accessibility to data. This should include an opportunity to provide notice and transparency to users about the purpose for collection and ensure responsible use of information.
- Accessibility involves the ease with which users obtain information.
- The way information is accessed and displayed must be sufficiently adaptable to meet a wide range of enterprise users and their corresponding methods of access.
- Access to data does not constitute understanding of the data — personnel should take caution not to misinterpret information.

- Data managers, designated by the data stewards, are generally operational managers within a functional area overseeing the data for a particular subject area. Data managers have day-to-day responsibility for managing administrative processes and establishing business rules for transactional systems. They have operational responsibility for the data management activities related to the collection, maintenance, protection, and dissemination of data in their functional areas.
- Data managers may authorize operational tasks to be performed by data users outside the units that report to the data manager. The data managers are accountable for the data subsets they manage, whether the data are collected or maintained directly by the data manager (or their staff), by data users in other units or by external sources.
- Data users are CofC employees who have been granted authorization by the data managers to access **institutional data**. Authorization is granted for a specific level of access, as defined by the data management policies, solely for the conduct of institutional business. Responsibilities include:
  - Following the policies and procedures established by the data stewards for the responsible use of the College data.
  - Using **institutional data** only as required to conduct College business.
  - Ensuring the privacy of data by viewing and storing data, and the information derived from data, under secure conditions.
  - Ensuring the accuracy and timeliness of the data they enter or update.
  - Collecting, preparing, entering, or maintaining data for the authorized unit(s), if authorized by the data manager.
- Access to data does not necessarily grant the user access rights to modify or disclose the data. This will require an education process and a change in the organizational culture, which currently supports a belief in "ownership" of data by functional units.

Principle 4: Data follow CofC Data Management standards

Note: this principle is an extension of TOGAF 9.2 framework, which deals only with the Data Trustee. The College has established a data management standard that defines the Trustee, the Steward, the Manager, and the User. This is how the base principles of the framework may be extended or adapted to suit the requirements of the College.

Statement:

- Administrative data is owned by the College of Charleston. Each data element has a data trustee accountable for data quality. The data trustee delegates authority to a data steward who delegates operational responsibilities to data managers.

Rationale:

- One of the benefits of an architected environment is the ability to share data (e.g., text, video, sound, etc.) across the College. As the degree of data sharing grows and stakeholders rely upon common information, it becomes essential that only the data trustee and steward make decisions about the content of data. Since data can lose its integrity when it is entered multiple times, the data trustee will have sole responsibility to ensure data entry processes minimize redundant human effort and data storage resources.

  Note: A trustee is different than a steward – a trustee is responsible for accuracy and currency of the data, while the responsibilities of a steward may be broader and include data standardization and definition tasks.

Implications:

- Real trusteeship dissolves the data "ownership" issues and allows the data to be available to meet all users' needs. This implies that a cultural change from data "ownership" to data "trusteeship" may be required.
- The data trustee will be responsible for meeting quality requirements levied upon the data for which the trustee is accountable.
- It is essential that the trustee can provide user confidence in the data based upon attributes such as "data source."
- It is essential to identify the true source (i.e., system of record or authoritative source) of the data in order that the data authority can be assigned this trustee responsibility.
- Information should be captured electronically once and immediately validated as close to the source as possible. Quality control measures, managed by the Data Governance Council, must be implemented to ensure the integrity of the data.
- As a result of sharing data across the College, the trustee is accountable and responsible for the accuracy and currency of their designated data element(s) and, subsequently, must then recognize the importance of this trusteeship responsibility.

Principle 5: Data have a Common Vocabulary and Data Definitions

Statement:

- Data are defined consistently throughout the College, and the definitions are understandable and available to all users.

Rationale:

- The data that will be used in the development of applications must have a common definition throughout the College to enable sharing of data. A common vocabulary will facilitate communications and enable dialog to be effective. In addition, it is required to interface systems and exchange data.

Implications:

- We are lulled into thinking that this issue is adequately addressed because there are people with "data administration" job titles and forums with charters implying responsibility. Significant energy and resources must be committed to this task. It is key to the success of efforts to improve the information environment. This is separate from, but related to, the issue of data element definition which is addressed by a broad community – this is more like a common vocabulary and definition.
- The College must establish the initial common vocabulary for the business; the definitions will be used uniformly throughout the College.
- Whenever a new data definition is required, the definition effort will be co-ordinated and reconciled with the corporate "glossary" of data descriptions. The College data administrator will provide this coordination through the Data Governance Council.
- Ambiguities resulting from multiple parochial definitions of data must give way to accepted higher education definitions and understanding (i.e., IPEDS).
- Multiple data standardization initiatives and procedures need to be coordinated.
- Functional data administration responsibilities must be assigned to data managers and data users.

Principle 6: Data follow CofC Data Classification standards

Note: this principle is an extension of TOGAF 9.2 framework, which deals only with the Data Security. The College has established a data classification that defines not only Security, but all data classifications. This is how the base principles of the framework may be extended or adapted to suit the requirements of the College.

Statement:

- Data are protected from unauthorized use and disclosure. In addition to the traditional aspects of confidential classification (i.e., data governed by federal, state, or other regulatory agencies), data security and privacy protects pre-publicized, internal use data — source selection-sensitive and proprietary information.

Rationale:

- Existing laws and regulations require the safeguarding of data using industry standard security and privacy frameworks, while permitting limited access as appropriate under the South Carolina Freedom of Information Act. Pre-publicized (work-in-progress, not yet authorized for release) information must be protected to avoid unwarranted speculation, misinterpretation, and inappropriate use.

Implications:

- Aggregation of data, both classified and not, can generate a dataset that requires review and de-classification procedures to maintain appropriate control. Data stewards and/or data managers must determine whether the aggregation results in an increased classification level. Appropriate policy and procedures, established by the Data Governance Council, will be needed to handle this review and de-classification. Access to information based on a need-to-know policy will force regular reviews of the body of information.
- To adequately provide access to public information while maintaining secure information, security needs must be identified and developed at the data level, not the application level.
- Data security safeguards can be put in place to restrict access to "view only", or "never see." Sensitivity labeling for access to pre-decisional, decisional, classified, sensitive, or proprietary information must be determined.
- Security must be designed into data elements from the beginning; it cannot be added later. Systems, data, and technologies must be protected from unauthorized access and manipulation. All information must be safeguarded against inadvertent or unauthorized alteration, sabotage, disaster, or inappropriate disclosure.
- New policies are needed on managing duration of protection for pre-decisional information and other works-in-progress, in consideration of content freshness.

Principle 7: Data will be Analyzable

Note: this principle is an extension of TOGAF 9.2 framework. This is how the base principles of the framework may be extended or adapted to suit the requirements of the College.

Statement:

- Data assets provide invaluable information to the enterprise for research and business intelligence decision-making when gathered, stored, and accessed correctly.

Rationale:

- Making sound business decisions at all levels across the enterprise is aided significantly by having access to timely accurate data from all systems, presented in the most appropriate way for the consumer. Holding this data for all systems, services, functions, and processes will enable an enterprise-wide view on our estate, capabilities, and the best utilisation of our resources throughout. Systems included go beyond our traditional monitoring points of student, staff, and financial data and will include areas such as, but not limited to, physical space usage, telephony usage, email usage, network statistics, all other areas of IT and beyond.

Implications:

- Data will be gathered from all capable systems and stored in a centralized data repository.

- All new systems will be procured with the capability to store logs and other pertinent information within the repository.
- Suitable analytical software will be used to extract required data from the repository and present to the consumer in the most appropriate way.
- This data and the subsequent use of it will be subject to information security, privacy and information management policies and classifications as set out by the College university.
- Data will meet data quality standards of accuracy, validity, reliability, timeliness, relevance, completeness, and compliance with College policy and legal regulations and statutory obligations.

Principle 8: Data are the enterprise memory of service delivery

Note: this principle is an extension of TOGAF 9.2 framework. This is how the base principles of the framework may be extended or adapted to suit the requirements of the College.

Statement:

- Data does not exist independently of business processes, which only exist to serve the delivery of services.

Rationale

- Data represent what occurred in the context of service delivery.
- The service delivered at the time is the business process at the time.
- Data only have meaning because of a business process executing to deliver a service.
- Service delivery is continuous, is not isolated, and builds on each other. (i.e., recruitment, admissions, matriculation, graduation, and alumni services are all connected).

Implications

- The semantic meaning of the data is based on the service delivered at the time
- Data and service delivery are connected through process execution
- Processes create data. Data aggregation and analysis becomes information, information, when combined with education and experience, generates knowledge allowing us to draw conclusions.
- Data and information domain leads to knowledge management where we can consistently analyse and apply learning to create new frameworks for managing our operation.

Principle 9: Data update should be the natural result of automated and/or self-service delivery

Note: this principle is an extension of TOGAF 9.2 framework. This is how the base principles of the framework may be extended or adapted to suit the requirements of the College.

Statement:

- Self-service provides the foundation for effective automation.

Rationale

- The more data are handled, the greater opportunity for data errors.
- Automation can only be achieved from good business processes.
- Systems and data must be fully integrated to ensure a single source of truth.

Implications

- Ensure business processes support self-service
- Create auditing methodologies to ensure data integrity
- Use technology to ensure data consistency and accuracy (i.e., clean address, field masking, etc.)
- Use technology to notify and confirm data modifications to users' profile.
- Use technology to allow users to make choices about and consent to collection and use of personal information.

- Use technology to allow users to review, correct or update personal information.

Principle 10: Compliance with Law

Statement:

- College information management processes comply with all relevant laws, policies, and regulations.

Rationale:

- College policy is to abide by laws, policies, and regulations. This will not preclude business process improvements that lead to changes in policies and regulations.

Implications:

- The College must be mindful to comply with laws, regulations, and external policies regarding the collection, retention, and management of data.
- Education and access to the rules.
- Efficiency, need, and common sense are not the only drivers. Changes in the law and changes in regulations may drive changes in our processes or applications.

Principle 11: Training

Statement:

- Periodic and regular training is necessary to ensure that competencies are maintained.

Rationale:

- Technical literacy is key to ensuring that efficient processes are maintain.
- Good and efficient processes generate reliable data.
- Periodic training to address new features and functionality is needed to ensure that the system is being used properly.
- Competencies are high during early years of implementation, over time, competencies drop
    - Attrition
    - Lack of sustained training program
    - New functions/features

Implications:

- Appropriate training and orientation are key to onboarding all new employees.
- Continuing education/training is key to maintaining competencies
- The College must be mindful that functional and feature updates are a part of the normal technology evolution.