

3357:13-19-30 Data Integrity/Data Request For Institutional Research Policy

(A) Policy. It is the policy of North Central State College that openness leads to a better informed citizenry, which leads to better government and better public policy. It is therefore, the policy of North Central State College to strictly adhere to the state's Public Records Act as it may pertain to data and other information relative to the Institutional Research function of North Central State College. Such data and other relevant information shall be made available on the Institutional Research webpage. All requests for data and/or relevant data not currently available on the Institutional Research repository will be reviewed and evaluated for production and release as required under the Public Records Act and the Ohio Revised Code. For detailed procedures on requesting data from the Institutional Research Office, refer to procedure [3357:13-19-101](#) Data Request Procedures for Institutional Research.

(B) Purpose

(1) The purpose of NC State's Data Integrity Plan is to promote shared accountability among key college constituents that are responsible for the use, production, data entry, and analysis of university data. This plan is designed to recognize that there is no single accountable person or office, but that multiple administrative offices share this responsibility with the academic units. This plan lays the framework for the development and on-going review of systematized processes that verify the accuracy of and allow for corrections to official college data. The plan is administered by jointly by the offices of Institutional Research and Information Technology, who collectively form the "Data Integrity Group." However, it relies upon a collaborative team representing key administrative offices and representatives.

(2) The Data Integrity & Request Policy and appendices include:

- (a) Policies and procedures for a data integrity group and data users group
- (b) Policies, procedures and forms for managing ad hoc data requests
- (c) Links to an updated rolling project log to track/manage all data projects and hours allocated
- (d) Links to an online data repository for posting a variety of categorized reports (in process)

(C) Steps involved

(1) Definition of data integrity

- (a) Accurate – the data are free from errors;
- (b) Complete – all values are present;

- (c) Consistent – the data satisfy a set of constraints and are maintained in a **consistent** fashion;
- (d) Timely – the data are available when required; and
- (e) Flexible – the data definitions are understood so that data can be analyzed in a number of ways.

(2) Business practices related to the collection and recording of data

- (a) How and when are data collected?
- (b) Where and how in Colleague are the data recorded?
- (c) Are data recorded in more than one system? If so, how reconciled?
- (d) What quality checks ensure accurate data capture?
- (e) Who is the “custodian” of each data element?

The data custodian, or data supplier, is responsible for collecting and supplying the data. The custodian is associated with the organizational unit vested with the operational responsibility for a specific set or organizational activities. The Registrar, for example, is often the custodian of student data. ¹

- (f) Who is the “steward” and actually creates the data? (Who collects and enters the data?)

The custodians must assign data stewards – operational staff – who are responsible for data administration within a specific set of elements and codes. The data stewards assure proper collection, storage and editing of operational data for completeness and accuracy.

- (g) What types of reporting are required?

The data broker obtains data from the various data sources and transforms them into information reports. This requires the ability to interpret the data and determine the type of analysis needed.

- (h) Who uses the data?

The manager takes the information and applies it to the situation. The manager needs to work with the broker to enhance the reliability of data, and then work with the broker to enhance the internal validity (interpretability) of the data.

¹ The custodian/steward, manager and broker definitions come from *People, Processes and Managing Data* by Gerald W. McLaughlin and Richard D. Howard.

(i) Results of this business review should include:

- (i) A comprehensive understanding of NC State's business practices that impact data quality;
- (ii) Identification of the individuals responsible for data collection and quality control;
- (iii) Identification of all data users and their requirements;
- (iv) Preliminary metrics to use in a quality assessment; and
- (v) Initial identification of problem business practices.

(j) In the immediate term, these questions should be answered in the contexts of external reporting: HEI, IPEDS, JBL, etc. In the longer-term, the college should attempt to apply these questions to all data collection.

(3) Assess existing data quality and completeness (on external report or sampled basis)

(a) Establish metrics to assess the validity and the data extracted from the Colleague. The metrics may include:

- (i) Acceptable date ranges for elements such as birth date, registration date, and achievement date;
- (ii) Acceptable syntax and values for elements such as course codes and program codes;
- (iii) Estimates of counts of total student registrations and student records; and
- (iv) Business rules with which data must comply. For example, a course section start date must come before the course section end date.

(b) To put it another way, many metrics check for standardization of the data amongst four major types of errors:

- (i) There needs to be data present for the required attributes of the entity;
- (ii) The data that are present must be in the allowable range or as an allowable category;
- (iii) The data that are present must be in a category that is consistent with the data that represent related elements. For example, financial aid information should not be available for high school dual credit students; and

- (iv) The data that are present should have a reasonable likelihood of occurring. For example. Students who are paying out of state fees should in general have out of state addresses.
- (c) To the extent possible, these checks and edits should be completed before the data is “frozen” in a data warehouse. The DIG group should collaborate with some of the major users and custodians on edit and validation routines. One possibility is to prepare “data management reports” to send out to data custodians prior to benchmark, indicating any key fields with missing or inconsistent data so they can be fixed prior to creation of the frozen census file.
- (d) Other tips, especially for immediate-term checking of external reports, include:
 - (i) Guarantee there are no technical error messages or warnings generated by the computer edit program;
 - (ii) Ensure the cohort size matches corresponding data in the official frozen census file;
 - (iii) Decide whether results make sense;
 - (iv) Have someone else check and edit the report before it goes out. If the data is initially prepared by the IT department (such as HEI and JBL), the Institutional Research Department should check every file given their contextual knowledge of the college. If the IR coordinator is preparing the report (IPEDS), the IR director should check that report likewise; and
 - (v) When applicable, data custodians and managers should also review relevant reports (e.g., controller review IPEDS finance).
- (e) For every quality problem revealed, the assessment will identify the source of the error, the magnitude of the problem (one record or thousands?), corrective action and a strategy to prevent this error from occurring again. However, it must be realistic in considering things like simple keying errors.

(4) Data Cleansing

- (a) Data cleansing goes to the source and corrects the errors and other problems identified during the quality assessment. As noted before, the preference is to do as much data cleansing as possible in the live Colleague system so that all future users of data will have accurate information. If data cleansing takes place after extraction of data from Colleague, cleansing will have to be repeated every time the data is re-extracted. Typical cleansing includes:
 - (i) Correcting of data entry errors;
 - (ii) Removing or correcting nonsensical dates;

- (iii) Deleting garbage records that don't contain valid data;
 - (iv) Combining and/or deleting duplicate records.
- (b) Data cleansing IS time consuming, and it is often necessary to perform extensive data cleansing as part of implementing a data quality improvement process. But the ultimate goal is to reduce the amount of data cleansing required over time. Changing business practices, combined with regular quality reviews, should help limit the amount of routine cleansing required.
- (c) For the immediate-term, the college will have to prioritize necessary data cleansing as it relates to external reporting elements. A good example would be the reporting of internships/co-ops/clinical on the HEI system, which appear to be under-reported in comparison to other colleges. This will likely entail a review of codes assigned by the Academic Division and reconciliation to match HEI criteria.

(5) Changing Business Practices

- (a) Implement changes to business practices that will improve data quality and adopt an ongoing process to regularly review and improve data quality. Business practices can range from simply reviewing data collection and entry practices with responsible staff to a complete re-engineering of business processes. Typical business practice changes could include:
- (i) Educating data entry staff on the importance of accurate data entry;
 - (ii) Updating code sets;
 - (iii) Centralizing the creation of new codes;
 - (iv) Ensuring that data is entered in the correct location in Colleague;
 - (v) Eliminating redundant fields; and
 - (vi) Implementing validation routines within Colleague.
- (b) For the immediate-term, this would entail addressing issues at the source level that are the true root causes that impair data integrity in external reporting.
- (c) In the long-term, the College should adopt an ongoing process for managing standardized data by incorporating the Shewhart Cycle for the support of data management. This essentially entails all the steps described above:
- (i) PLAN – Data Management Structure

- (a) Identify and establish an official source for critical entities with a list of standard values for key attributes (beginning of centralized data repository)
 - (b) Assign data custodial responsibility and accountability
 - (c) Establish and implement policies that balance accessibility with security
- (ii) DO – Data Standardization
 - (a) Standardize and distribute data descriptions, definitions, and documentation
 - (b) Apply consistent definitions over time (historical data)
 - (c) Cross-reference all occurrences of a data element across the organization
 - (d) The most important standard is the standard of consistency – consistency of data naming, data attributes, data design and data use.
- (iii) CHECK – Processes and Procedures
 - (a) Implement systematic edits and validation to ensure completeness and accuracy
 - (b) Establish audits for accuracy and measures of accountability
 - (c) Develop a process for reporting the results of the data edits, audits and checks
- (iv) ACT – Implement and Monitor Data Access and Data Use
 - (a) Create query capability to identify data sources and data modification procedures
 - (b) Ensure the retention of historical data as well as ready access to timely and historical data by trained users
 - (c) Survey users to measure the extent that data usage is clear and meaningful
 - (d) Again, this requires that the operational offices supply reliable data; the central data management function integrates and refines the data into a usable form and produces internally consistent information; the customers have access and training such that they can generalize the information they receive to their needs and situation; and the three groups communicate, coordinate and cooperate.

Effective: April 7, 2014

Expires: April 1, 2019

Review Dates: 4/7/19