Information Use and Standards Policy for the Bentley Data Warehouse[1]
February 5, 2004

I. Introduction: Administrative data, owned by Bentley and maintained by various departmental stewards, is a valuable institutional resource. While these data may reside in different database management systems and on different machines, in aggregate they may be thought to form one logical resource.

A "data warehouse" is a data repository designed to support integrated, cross-functional analysis, institutional research, reporting and executive decision-making. It contains sharable historic data from these multiple operational systems-of-record, as well as transactional data derived from the operational data and deemed to be useful management information.

This policy establishes uniform data management standards and identifies the shared responsibilities for assuring that the data warehouse provides security, protects privacy and has integrity while it efficiently and effectively serves the needs of Bentley College. This policy applies to those data that are critical to the administration of the institution regardless of whether the data are used or maintained by administrative or academic units.

II. Data Security, Privacy, and Access Philosophy: The overarching goal of the data policy is to strike a balance between data access and data security and privacy. The value of data as an institutional resource is increased through its widespread and appropriate use; however, its value is diminished through misinterpretation, misuse, or abuse. Of the two concerns, data security and privacy is the more critical and delicate. Access can be expanded as needed, but privacy, once violated, can seldom be repaired and security, once violated, can compromise the financial integrity, reputation, functionality, and stability of the institution.

The data warehouse exists to support institutional self-knowledge and planning. Operational and planning information should be readily available within the institution, but on a need-to-know basis. Permission to view or query data contained in the data warehouse should be granted only for legitimate institutional purposes.

III. Data Access Standards:

- ***Access to the data warehouse*** is limited to employees of Bentley. Access will not be granted to outside contractors or consultants, or to student-employees. Update access to the warehouse is restricted to the Information Technology professionals administering it. Normally, access will not be granted to support individual faculty or student research projects. Access is granted under this policy to institutional researchers, technical professionals and officers of the institution. Access to others will be approved by the Committee on Information Use and Standards upon the recommendation of the applicable data steward. Permission to view and query data contained in the data warehouse should be granted only for legitimate departmental or institutional purposes. The breadth and depth of access will be determined by the role of the individual and may be contingent upon training on applicable data policies and responsibilities. Security and privacy will be prioritized over access, except as required by business need.

  The warehouse will be used by three levels of users. Record-level data that are individually-identifiable will reside in the warehouse. Indeed, such data must reside there to support operational reporting. In some cases, staff doing operational reporting have the most extensive and frequent access to individually-identifiable data and the least training and

experience with responsible data use. The inauguration of the DW is an opportunity to review and revise the data access privileges of Level 1 and Level 2 users, conceivably creating enhanced protection for individually-identifiable data.

- ***Data Types Residing in the Warehouse:*** Two kinds of data reside in the warehouse that relate to people: public domain data and confidential individually-identifiable data.

  Public domain data includes:
  - Directory information as defined by the institution, respecting students rights under the Family Education Rights and Privacy Act to withhold release of directory information;
  - Staff directory information, unless restricted by the individual;
  - and certain aggregated data, at determined by the institution which have been reported to the federal government, accrediting agencies, or other sources.

  Confidential individually-identifiable data includes all sorts of record level data that are associated with the person whom it describes. The relationship between individual and attribute is intact and visible.

  This table summarizes distinctions among the user levels in terms of purpose, competence and data-type access:

	L1: Operational	L2: Unit Specific Analyst	L3: Enterprise Analyst or Information Consumer
Purpose	These users run operational reports to support routine office functions.	These users make more complex use of unit-specific data and may engage in data analysis, interpretation, planning and evaluation for the office.	These users are senior officers, researchers and planners and others so assigned who perform complex analyzes, often cross-unit in scope and involving data under the stewardship of more than one division.
Competence	Know data. Trained to execute (and perhaps create) simple reports using a report writing product targeted at end users.	Know data, techniques and approaches specific to the functional area. Trained to write reports with end user-oriented report writing tool. Competent in basic data manipulation, file handling, and descriptive statistics. Know the practical and statistical limits of the data.	Info Consumers: Know the questions to which they want answers, and are intelligent consumers of quantitative information. Enterprise Analyst: Know data. Professional caliber in terms of research design, data collection techniques, data manipulation and analysis, descriptive and inferential statistics, graphical representation and presentation skills. Know the practical and statistical limits of the data.
Access Scope	Office-specific in scope, minimum needed to perform job.	Office-specific or Division-specific in scope, minimum needed to perform job.	Enterprise-wide, with special obligations to engage in data disguising practices to reduce view

			access to individually-identifiable data and to protect them from disclosure.
Access to individually-identifiable data	Yes. Record level data are necessary to verify correct data entry, generate administrative control lists and the like.	Yes, consistent with need. But sometimes unnecessary. The use of individually-identifiable data should be avoided unless necessary to complete the assignment.	Yes, but generally unnecessary. The use of individually-identifiable data should be avoided unless necessary to complete the assignment.

IV. Data Use Standards:

- ***Culture of Responsible Data Use:*** These standards aim to maximize the use of institutional information, while protecting individually identifiable information from disclosure. The approach includes reasonable data protection techniques and sanctions for irresponsible conduct, but depends more substantially on education, elevated consciousness, mutual trust and shared responsibility about responsible data use. It also seeks to develop an appreciation of the costs and risks of ethical lapses, neglect and inadvertent outcomes.

- ***Selecting/Accepting Projects:*** At the design stage, persons initiating analytic projects will thoroughly explore the degree of invasion of privacy and the risks of breach of confidentiality that are involved, will weigh them against potential benefits, and will make a recommendation to the person(s) commissioning the project as to whether the project should be executed and under what conditions.

  When appropriate, users will adopt a written description of special precautions beyond the regular guidelines described here necessary during an assignment to ensure the protection of aspects of privacy and confidentiality that may be at specific risk.

- ***Data Editing Techniques:*** Level 2 and 3 analysts will use, as appropriate, data editing techniques to eliminate unnecessary visual access by the researcher to individually-identifiable data and to protect against the release of such data, either directly or by deduction.

    - Data disguising/linking: Relationship between individual and attribute is intact but invisible to the researcher, allowing him/her to work with data without knowing to whom it refers, and enabling, as appropriate, linkage to other data sets to enable longitudinal studies. Use random identifier, not name or social security number.
    - Coding or Coarsening: This is a disclosure limitation technique that protects individual-identifiable data by reducing the level of detail used to report some variables. Examples of this technique include: recoding continuous variables into intervals; recoding categorical data into broader intervals; and top or bottom coding the ends of continuous distributions.
    - Rendering data anonymous: In the data file used for analysis, names, addresses, social security numbers and other positive identifiers are stripped.
    - Cell suppression: A data item in a table which could lead to disclosure may simply be suppressed, i.e., the cell value is omitted and replaced by an asterisk or other symbol which indicates that the number is being omitted to maintain confidentiality

for the subjects. However, care must be taken to assure that the omitted value may not then be deduced by subtraction, which requires that another cell value in the same row and another in the same column also be suppressed, assuming it is desired that no changes be made in the row and column totals.

- Attention to Sample Size and Cell Frequency: When sample sizes are small cells with 1 or 2 cases may occur. In lieu of cell suppression, the researcher might simply change the intervals to combine cells with small counts, thus protecting the identity of the research subject from the possibility of being revealed by deduction.

- ***Secure Storage and Transmission:*** Whether spoken, in hardcopy or electronic form, users shall organize, distribute, print, store, maintain, analyze, and/or transfer data, under their control in such a manner as to reasonably prevent loss, unauthorized access or divulgence of confidential information. Data files containing individually-identifiable information and/or supporting research findings shall be stored and archived securely.

- ***Data Destruction/Archiving:*** If materials containing individually-identifiable data are to be destroyed, the method of destruction shall be appropriate. Such materials shall not go into normal trash or recycling bins. Destruction should be by shredding or other protective disposal technique. Electronic records are subject to comparable controls. Unless stored and archived securely as necessary to support research findings, data files should be destroyed promptly after serving their purpose. Special care shall apply to the control, management and destruction of various export formats offered by standard query tools including but not limited to spreadsheet, comma-delimited, pdf and html. Level 3 users shall apply all reasonable means to prevent irrevocable loss of data and documentation during its immediate useful life, and being aware of the role of data as institutional historic resource, shall act as an advocate for its documentation and systematic permanent archiving.

- ***Release of Individually-Identifiable Data:*** Level 1 users shall not allow individually-identifiable data to be released in any form outside the office without the explicit permission of the appropriate data steward. Level 2 and 3 users shall not allow individually-identifiable data to appear in reports, spreadsheets, email messages or other media that will be made public to the campus community or beyond it.

- ***Release of Institutionally-Identified Information:*** Except as allowed below, institutionally-identified information derived from warehouse data shall not be available to the general public through any medium. This restriction applies to release to the media (including Bentley student print and broadcast media), corporations, associations, agencies or commissions; such data releases may only be made by the Office of Public Affairs or other authorized office. The Office of Institutional Research and Planning responds to many external mandates and requests for statistical information about Bentley. It may also release as a courtesy to scholars and researchers at other institutions data commonly in the public domain, and may share other data as permitted by formal data sharing agreements (e.g., Higher Education Data Sharing Consortium). Institutional researchers and other Level 3 analysts are constrained from revealing institutionally-identified data in scholarly and professional publications, except as allowed by such agreements or by explicit permission of his/her vice president.

The table below summarizes the distinctions among the user levels in terms of data use, protection and disclosure responsibilities and/or prohibitions. The professional conduct of Level 3 users with respect to privacy and confidentiality is often outlined in codes of ethics associated

with their professional associations; so while this local codification may be new to Bentley, the behavior it mandates is likely consonant with current practice.

	L1: Operational	L2: Unit Specific Analyst	L3: Enterprise Analyst
Data Storage & Destruction	Reports, data files and other hardcopy and electronic data sources are subject to secure storage and disposal.	Reports, data files and other hardcopy and electronic data sources are subject to secure storage and disposal.	Such materials are subject to secure storage and disposal. Some products of complex analyses become valuable institutional data sources in their own right and shall be documented and permanently archived.
Circulation within functional office	Such reports are for office use only. Access to them is limited to persons who need the information in order to perform their jobs.	Reports with individually-identifiable data are limited to persons who need the information to perform their jobs. Circulation of reports and analyzes with aggregated data at the discretion of the data steward.	Reports and other final products of such work shall contain no individually- identifiable material. Access to materials upon which analyzes and reports are based containing individually-identifiable data is limited to the researcher/analyst. Reports containing data in the public domain may be shared freely.
Authorization to share with other offices at the institution.	No, except with respect to data in the public domain or with permission of data steward.	No, except with respect to data in the public domain or with permission of data steward.	Reports and other accounts of such work shall contain no personally- identifiable material. Circulation of special reports and analyzes with aggregated data determined by client(s) who commissioned them. Some broadly based statistical reports of institution-wide interest, such as may be produced by the Office of Institutional Research and Planning or MIS, are considered to have been commissioned by the institution and owned collectively; their circulation is determined by the Cabinet. Reports containing only data in the public domain may be shared freely
Authorization to share outside the institution.	No.	No, except with respect to data in the public domain or with the permission of the data steward.	Yes, as outlined in #14 below.

Compliance with Standards: By accessing and using the Data Warehouse, you will be deemed to have agreed to all the Standards of Conduct contained herein. Any user found in violation of these standards will be penalized by loss of access privileges to the reporting environment and may be subject to more severe sanctions, consistent with existing employee disciplinary policies and procedures.

V. Standards of Conduct:

1. I understand that my access to the data warehouse is limited to a "need to know" in order to perform my job.
2. Unless otherwise specifically allowed in these standards, I understand that I may not reproduce, republish, distribute, sell, trade, or share data. I understand as well that I may not modify or alter any content or data residing in the Warehouse.
3. I agree to exercise my responsibilities in the use of confidential data in such a way as to bring no harm to Bentley, its students, faculty, or staff.
4. I understand that embedded in my log-in credentials (username and password) are my personal access privileges. I will only gain access to the DW using my personal username and password. I will not share my password with others.
5. I agree that I will not undertake data manipulation, analysis or reporting unless suitably prepared by training and/or experience, or unless working under the guidance of someone so qualified. In general, serious inquiries with important ramifications for the institution should be conducted by the Office of Planning and Institutional Research, or similarly trained professionals.
6. I agree to conduct all tasks in accordance with accepted technical standards, and to use statistical methodologies suitable to the data and to obtaining valid results.
7. If applicable, I agree to report the limits of statistical inference of the study and possible sources of error.
8. I agree to conduct my work with objectivity, approaching my assignments with an unbiased attitude and striving to gather evidence fairly and accurately.
9. I will be particularly sensitive to avoid personal conflicts of interest when performing information services, disclosing any conflicts of interest, financial and otherwise, and resolving them. This may sometimes require divestiture of the conflicting interest or recusal from the project.
10. Bentley complies with federal, state and local laws governing privacy and confidentiality. I understand that my work may be subject to such regulation. Accordingly, I will take steps to familiarize myself with whatever obligations follow from such compliance and strive to conduct my work in accordance with them. Specifically, if using human subjects in research, I will contact the Office of Sponsored Programs to determine whether the project is subject to review by Bentley's Institutional Review Board.
11. I agree to employ, as appropriate, disclosure avoidance techniques such as those illustrated in this document.
12. I agree to provide secure storage and appropriate disposal of confidential materials in my custody.
13. I agree that I will not make individually-identifiable data available in reports, spreadsheets, email messages or other media that will be made public within the campus community or beyond it, except as explicitly allowed in this document.
14. I agree that I will not make institutionally-identifiable data available to the general public through any medium. This restriction applies to members of the media (including Bentley student print and broadcast media), corporations, associations, agencies or commissions; such data releases may only be made by the Office of Public Affairs. The Office of Institutional

Research and Planning may release as a courtesy to scholars and researchers at other institutions data commonly in the public domain, may respond to external mandates requests for statistical information about Bentley and may share other data as permitted by formal data sharing agreements (e.g., Higher Education Data Sharing Consortium).

15. I will not include institutionally-identifiable data derived from the data warehouse in scholarly publications, in other published sources, or in presentations delivered at professional meetings off campus, without the permission of my vice president.
16. I agree to promptly and publicly correct any errors discovered after the report is released.

VI. **Data Integrity, Validation and Correction Policy**: Data integrity will be maintained within the source systems that feed the warehouse. Data Stewards are responsible for assuring that the applications that capture and update data incorporate edit and validation checks to protect the integrity of the data.

Data Experts are responsible for correcting data problems and inaccuracies. Data Users are responsible for supplying as much detailed information as possible about the nature of the erroneous data.

Upon written identification and notification of erroneous data to the Data Stewards by data warehouse technical professionals, corrective measures should be taken as soon as possible to:

- Evaluate and appropriately correct the data at the source. Errors will be corrected in the live data whenever possible, and are the responsibility of Data Stewards and Data Experts. (Erroneous data will not be corrected in the historical tables maintained in the data warehouse.) Corrective measures may include, but are not limited to**,** modification (by authors) of reports accessing erroneous data in the data warehouse, updates by data warehouse technical professionals to correct erroneous data in data mart tables abstracted from the warehouse and used for reporting, addendums or notations by authors to published institutional reports**,** and process changes enacted by Data Stewards to prevent on-going occurrences of erroneous data entry in the source systems.
- Report corrective action in writing to the Committee on Information Use and Standards for review.
- Notify Divisional Key Users and Data Experts who have received or accessed erroneous data.

VII. Data Management Roles and Responsibilities:

President and Cabinet- The President determines overall planning and policy-making responsibilities for institutional data and makes such delegations as deemed appropriate. The Cabinet members, as individuals, are responsible for overseeing compliance with data management policies and procedures for units within their division and for the assignment of data management accountability with appropriate data steward(s). College-wide data policies and accountabilities are reviewed by the cabinet as a whole and approved by the CIO.

Chief Information Officer (CIO) - The Vice President for IT & Vice Provost (CIO) is responsible for overseeing the management of institutional information resources and security. The CIO has signature approval authority for all data policies.

Data Stewards- These are Directors or above (e.g., the Controller, Registrar, Dean, Executive Director) who oversee the capture, integrity, maintenance and dissemination of data for a particular operation according to the standard data policies and procedures. Data stewards and data experts provide data descriptions and definitions so data warehouse users know what

shareable data are available, what the data mean, and how to access and process the data. It is the responsibility of Data Stewards and Data Experts to notify data warehouse technical professionals whenever changes occur in data descriptions, definitions, and meanings. They also share responsibility for data security and privacy with Director level and above managers within the Information Technology division.

Data Experts- These are technical staff and operational managers (e.g., Director of Enrollment Systems, Director of Financial Systems, Associate Registrar for Technical Services, Administrative Systems Manager, Director of Advancement Systems) within a functional area with day-to-day responsibilities for managing business processes and establishing business rules for the production transaction systems. Such individuals may also serve in the roles of Divisional Key User and Departmental Key User. Data stewards and data experts provide data descriptions and definitions so data warehouse users know what shareable data are available, what the data mean, and how to access and process the data.. It is the responsibility of Data Stewards and Data Experts to notify data warehouse technical professionals whenever changes occur in data descriptions, definitions, and meanings.

Data Warehouse Technical Professionals: IT technical professionals build, extend and support the data warehouse and implements access to it. They also share responsibility for data security and privacy with data stewards and data experts. These technical professionals store metadata (data about data) in a suitable way, so that users and analysts can easily access it and appropriately use and interpret reported data.

Data Users- Data Users are individuals who access warehouse data in order to perform their assigned duties or to fulfill their role in the community. Data users are responsible for protecting their access privileges and for proper use and protection of the data they access, and are held accountable for their own use of data.

Committee on Information Use and Standards: This group, typically composed of Data Stewards and Data Experts, regulates access to the data warehouse, reviews the operational effectiveness of data management policies, and makes recommendation to the CIO and Cabinet for improvement or change. It is chaired by the Dean for Information Resources.

Recommended by the Committee on Information Use and Standards:

_____ _____

Barbara H. Palmer, Chair Date
Dean for Information Resources

Approved, after Cabinet review and concurrence:

_____ _____

Traci Logan Date
Vice President for Information Technology

[1] Bentley College borrowed extensively from the data policies of Renesselaer Polytechnic Institute and Virginia Polytechnic Institute and State University, and acknowledges this debt with gratitude. We also acknowledge a debt to the following sources: EDUCAUSE Core Data Service Appropriate Use Policy; Association for Institutional Research Code of Ethics; and the American Statistical Association Ethical Guidelines for Statistical Practice.