The current higher education landscape is replete with demands for improving accountability, increasing efficiency, and controlling costs. At the same time, information technologies make it easier to collect and analyze information to measure outcomes or to assist in decision making. Consequently, there is a higher demand for better information and also a never-before-available supply of data. This corresponding increase in both supply and demand creates the perfect storm for higher education to move into a new era of analytics. However, in a recent discussion session about the legal issues and campus policy dimensions for analytics, one participant joked: “This is where analytics efforts on campus come to die!” Although legal and policy dimensions do present considerable challenges, they should not be simply dismissed as obstacles with no solutions. Therefore, it is helpful to frame the policy context of analytics to include the policy drivers that are generating demand and the campus policy choices that should be guiding campus practice.
Policy Drivers
Internal and external forces are demanding greater accountability and transparency of college and university finances and operations. Policy-makers at the federal and state levels are expecting answers to questions: How are taxpayers’ dollars used? What is the impact of public investment in higher education? Are there ways to reduce costs and keep college affordable? Are institutions retaining students and ensuring that they complete college? There are also efforts to track individual students’ progress along the educational continuum—from K–12 to postsecondary education to employment—through the use of state longitudinal data systems. Increasing emphasis on learning outcomes requires that we identify measures of progress and develop better mechanisms to assess learning. Data helps to unlock the mystery about college costs, learning outcomes, institutional effectiveness, and other performance indicators.

Data helps to unlock the mystery about college costs, learning outcomes, institutional effectiveness, and other performance indicators.

In most cases, this data is collected, analyzed, and reported to satisfy regulatory requirements and is often viewed as a bureaucratic intrusion upon institutional autonomy. Yet there is also an increasing recognition of the inherent value of data-informed decisions and the power of information technologies to create dashboards, infographics, or other snapshot views to keep governing boards, campus administrators, and faculty informed in a variety of areas—which in turn can help them spot early signs of trouble where intervention may be necessary. A key component of an analytics program is first to identify the policy questions to be answered and then to engage in a risk-management exercise, including a cost/benefit analysis, to determine if analytics will provide the answers or feedback needed.

Campus Policy
Creating a culture of evidence on campus requires clear policies and processes with respect to the use of data. A key component of a solid analytics program is a system that is built on trust and normalized processes, transcending individual personalities or arbitrary decision making. A holistic approach to data management starts with effective governance, rational policies, and reliable procedures.

Data Governance. The most important step that any campus can take is to create a comprehensive data-governance structure to address all the types of data used in various situations. For example, although cloud computing or service outsourcing (with data stored or handled outside of the institution) creates unique challenges, most of the policy and process questions could be addressed through a data-governance body. Similarly, as research institutions move to apply privacy protections and security safeguards to research data in a manner similar to administrative data, they will find that including research data within the data-governance system will lead to more consistent and effective results.

Data Classification. One of the first tasks of a data-governance body is to inventory campus data sources and create a data-classification system that sorts data into categories that identify the necessary level of protection. This process may necessarily be different for a private versus a public institution, since state laws or regulations may require that certain information be available to the public. A typical classification scheme for a public college or university might include the categories 1. public, 2. non-public, and 3. sensitive or regulated. It is often difficult to define terms or draw lines (e.g., between “non-public” or “sensitive”), which is why some institutions have chosen more generic categories (e.g., levels 1, 2, 3). A data-classification or data-categorization exercise is a fundamental building block for an analytics program.

Roles and Responsibilities. Once the data has been classified, it is useful to clarify roles and responsibilities. This is often accomplished by assigning roles such as data stewards or data custodians. Data stewards are usually director-level employees who have the overall responsibility for...
Policy Dimensions of Analytics in Higher Education

specific data sets, including the enforcement of data standards and controls and the implementation of policies such as data access. Data custodians, in contrast, do not have the same policy authority but are often the departmental employees, IT staff, or users who have access to the data. As custodians, they have corresponding responsibilities to follow policy and procedures, but they do not typically have the power to authorize access.

Data Policies. The establishment of data policies and procedures takes on different forms at different types and sizes of institutions. In some cases, data policies may be dictated by legal or regulatory requirements that get translated into a formal compliance program. In other cases, data policies might establish the broad frameworks for data governance and classification and for the role of data stewards and custodians, as outlined above. Data access is perhaps one of the most important—and difficult—policy issues to clarify. Who should have access to the data and under what circumstances? Who is authorized to decide?

Data Privacy and Fair Information Practices. The United States does not currently have a comprehensive national privacy law. As a result, federal and state laws and regulations have taken a piecemeal approach to regulating data according to the type of data collected. For example, the Family Educational Rights and Privacy Act (FERPA) governs the use of students’ education records. The Health Insurance Portability and Accountability Act (HIPAA) is concerned with the use of personal health information. The Gramm-Leach-Bliley Act covers financial information. Employee information tends to be addressed by state law or through institutional policy or employment contracts. Since there are generally very few absolute mandates for data privacy, it comes down to choices about how to balance individual concerns for privacy with institutional needs. A common ethical framework for the development of campus privacy policies and practices can be found in the Fair Information Practice Principles. The Federal Trade Commission summarizes the principles as follows:

1. Notice/Awareness: Individuals should be given notice of an entity’s information practices before any personal information is collected from them.
2. Choice/Consent: Individuals should be given options as to how any personal information collected from them may be used, and secondary uses should be permissible only after consent.
3. Access/Participation: Individuals should have access to data about them and should have the ability to insist on that data’s accuracy and completeness.
4. Integrity/Security: Data must be accurate and secure, with controls to protect against unauthorized access, destruction, use, or disclosure.
5. Enforcement/Redress: There must be mechanisms to ensure compliance and appropriate means of recourse by individuals.

De-Identification of Data. Institutions may have a number of reasons for using de-identified data for business, academic, or operational functions. For instance, data can be made available for institutional use, without identifying the underlying data subjects, for research purposes, institutional effectiveness studies, performance and operational studies, IT security and operational reviews, and public health purposes. Other uses of de-identified data may require the ability to retain unique identifiers for individuals in the data set, without identifying the actual identity of the individuals. For example, a researcher may need to know that certain actions were all taken by the same individual, in order to form conclusions about how individuals use the data or service. A website designer may want to determine how long indi-
individuals stay on the site or how individuals traverse the site in order to find the information sought. Systems development, test, and training environments may require the use of data that simulates real production data while not actually consisting of real data elements (e.g., Social Security numbers). In such cases, de-identification processes are complicated by the need to replace unique identifiers, such as Social Security numbers or IP numbers, with alternative unique identifiers that cannot be used to identify the individual. Yet though de-identifying data is a useful step toward protecting privacy, the de-identified data can still carry a number of privacy risks.3

Information Security. Although closely coupled with many of the policy choices identified above, information security concerns not the restrictions put on data but, rather, the availability of the data and the resiliency of its underlying information systems. Business continuity planning is thus a critical aspect of any information security program.4

Conclusion
For too long, IT professionals in higher education have been fixated on the “T” (technology) without an appreciation of the importance of the “I” (information). Fundamentally, analytics centers on the effective use of data and information. Therefore, the starting point for the examination of legal issues and campus policies must be the collection, storage, use, and disclosure of data—which necessarily align closely with matters of data privacy and information security. As enablers of the effective use of analytics, information security and technology systems and networks should not be viewed as barriers. With the increasing demand for more and better analytics, the IT community needs to work with other campus stakeholders to ensure that appropriate data governance, data classification, data roles and responsibilities, and other data policies and procedures are in place.

Notes
1. For more information, see the EDUCAUSE Data Governance resource page: <http://www.educause.edu/Resources/Browse/Data%20Governance/34511>.
2. For more information, see the EDUCAUSE/Internet2 Higher Education Information Security Council’s Data Classification Toolkit: <https://wiki.internet2.edu/confluence/display/itsg2/Data+Classification+Toolkit>.
5. For more information, see the EDUCAUSE/Internet2 Higher Education Information Security Council’s Guidelines for Data De-Identification or Anonymization: <https://wiki.internet2.edu/confluence/display/itsg2/Guidelines+for+Data+De-Identification+or+Anonymization>.

© 2012 Rodney J. Petersen. The text of this article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Rodney J. Petersen (rpetersen@educause.edu) is Senior Government Relations Officer and Managing Director, Washington Office, for EDUCAUSE.