UCLA Data Governance Task Force

FINAL REPORT AND RECOMMENDATIONS
MAY 2016

Data Governance Task Force, 2014–2016

Co-Chairs

Christine Borgman, Distinguished Professor and Presidential Chair in Information Studies
Kent Wada, Chief Privacy Officer and Director, Strategic IT Policy, Office of Information Technology

Members

L. Amy Blum Interim Vice Chancellor, Legal Affairs
Meg Buzzi, Opus Project Director, Office of the Vice Chancellor, Academic Personnel
Kristen Chamberland, Graduate Student Representative
Christina A. Christie, Professor of Education, Graduate School of Education & Information Studies
Mike Lee, Associate Director, Social Sciences Computing
Vickie M. Mays, Professor, Department of Psychology
Kristen McKinney, Director, Student Affairs Information and Research Office
Kelly Wahl, Director, Statistical Analysis, Office of Academic Planning and Budget
Neil S. Wenger, Professor of Medicine, General Internal Medicine and Health Services Research
(Unassigned), Undergraduate Student Representative

Policy Staff

Anna Joyce, Manager, Administrative Policies and Delegations, Administrative Policies and Compliance Office

Contents

I. Introduction

Data are critical to UCLA's ability to innovate, enhance, and execute our core missions of education, research, and service. We collect some kinds of data about the members of our community for purposes of basic operations, accounting, audit, and accreditation. Other kinds of data are necessary to improve user experience, whether with library services, instruction, research, travel, finances, or other aspects of the campus environment. We have mandatory obligations to report data on enrollments, time to degree, diversity, budgets, grants, library collections, and countless other functions to specific offices of the University of California, government agencies, accreditation bodies, and other entities. For these and many other reasons, the campus collects, produces, possesses, consumes, and shares tremendous amounts of data.

Our data are valuable assets, and as stewards of our data we must govern our data responsibly. Good data stewardship encompasses legal compliance, policies, and practices pertaining to data life cycle (records management and disposition), ownership, classification, privacy, incident response, disclosure (e.g., under California Public Records Act requests), curation, integrity, and management. Relentless cyber-attacks and a world of "information-radiant" systems challenge our ability to properly manage our data, and we are mindful that the trust chain begins with information security.

Cyber-attacks are not the only challenge to our ability to manage data responsibly. "Big data" analytics and the routine sharing of data with a diverse array of third parties have accelerated concerns for reconsidering our policy, funding, and technology frameworks for the appropriate use and protection of data. Data sources and stakeholders in our community change continually. The profile of our data stewards also is changing: undergraduate students are increasingly joining graduate students, faculty, and campus units in collecting, producing, and disseminating data. All of these stakeholders need to understand their responsibilities for data management. The campus needs to assess and address gaps faced by each of these constituencies.

The full breadth of institutional data stewardship is beyond the scope of our charge. Good data stewardship is, however, fundamental to campus operations and underlies the work of this Task Force. This report is a starting point, rather than an ending point, for the campus to comprehend the range of its data resources and effective approaches to managing them.

Report Focus

Based on our charge, we provide recommendations for an initial set of principles by which governance decisions can be considered, a UCLA data governance structure, and necessary surrounding processes. We focus on the governance uses of data about individuals within our

community, whether or not those data meet any existing legal or policy definition of personally identifiable information (PII).

Appropriate and ethical use of research data is in part embodied by the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979), the basis for Institional Review Boards (IRB) to protect human subjects in research. The IRB is a mechanism by which universities balance their core research mission with the requirement to do no harm. Yet a fundamental concern that led to the creation of this Task Force is the growing number of data-generating activities involving our students, faculty, and staff that fall outside of IRB purview. The scope of "human research" encompasses only projects intended to develop or contribute to "generalizable knowledge" (UCLA Office of the Human Research Protection Program, 2013). Sensitive uses of data often fall into a policy void if their conditions do not fall under the regulatory framework of the IRB. Universities need to develop policy frameworks to guide appropriate uses of data that fall outside the purview of the IRB. This report addresses that middle ground for data about our community.

Privacy is a fundamental consideration in appropriate uses of data, thus we examine relevant privacy models. Privacy principles traditionally have been based on the Code of Fair Information Practices (FIPs) established in the 1970s and subsequently revised (Gellman, 2014; The Organisation for Economic Co-operation and Development, 2013). Core to FIPs—and laws that are based on FIPs, such as Family Educational Rights and Privacy Act (FERPA) and Health Insurance Portability and Accountability Act (HIPAA)—are the principles of *notice*, that data collection systems must be known to the subjects of that data collection and not be secret, and *consent*, that data collected for one purpose should not be used for another purpose without the express permission of the person who is the data subject. These privacy policies emerged from a much earlier area of technology practice and policy. In today's era of "big data" analytics, social media, and ubiquitous online transactions, small amounts of data are being collected by many systems, on a continual basis. These small bits of data become much more valuable as they become integrated across systems and services. Individuals may have notice of one party that is collecting data about them, but not be aware of secondary, tertiary, and other parties who later gain access to those data. Similarly, they may have given consent for a specific use, but not be aware of the value of those small bits of data when combined with many other small bits of data. Even when data are de-identified, individuals can be re-identified through the aggregation of data. Opting out of data collection is rarely an option in university environments, because access to basic university and public services depends upon myriad uses of data for teaching, learning, library services, travel, personnel, payroll, recreation, and well beyond.

Recommendations

The governance mechanisms recommended in this report are intended (1) to resolve legitimate institutional disagreements about appropriate data use when not fully resolvable by policies or laws, and (2) to provide a path forward for fuller engagement of the campus community in data governance issues. Our goal is to support innovation in education, research, and service with a consistent set of expectations about the appropriate uses of data about our community.

II. Framing the Challenges for UCLA

Our data are valuable institutional assets. Their value increases, both to the University and to external third parties, as they accumulate and can be reused and remixed in new ways. "Big data" and predictive analytics define a new generation of opportunities across the institution, whether in student success, research, precision medicine, or administrative effectiveness.

Risks of breach, misuse, and misinterpretation of information about our community also increase. Data that may not appear to be sensitive at the time of collection, such as student traffic to a course website, may become extremely rich when combined with other data such as a student's grades, medical records, library usage, food purchases, and social media habits. Similarly, information that is nominally public, such as a faculty member's bibliography of publications, can become extremely sensitive when combined with proprietary analytics used to rank individuals, departments, universities, and countries. As data, metadata, algorithms, and analytics are shared within and between universities and with third parties, the complexity of data governance increases.

Thus, for the purposes of this report, data governance concerns all of these cases:

- *Uses of University data internal to the University* (e.g., course management systems, faculty personnel dossiers, institutional analytics)
- *Uses of University data in collaboration with external partners* (e.g., student, faculty, or research data shared with partners at other universities or institutions)
- *Releases of University data to suppliers* (e.g., Internet-based service providers that are given access to data about our community to provide; email, cloud storage, or web hosting services; or that have such access as a byproduct of providing a different primary service; for example, a course management or cybersecurity system)

To make the data governance problem more tractable, the Task Force focused on the following three UCLA scenarios. These present-day scenarios are intended to prompt discussion around the need for campus mechanisms to govern appropriate uses of data rather than to be evaluated in depth or to solve specific questions that have emerged from these scenarios.

Scenario 1: *Internal capture and uses of data about members of the UCLA community*. The development of the Opus Faculty Information System improves efficiency, transparency, and accuracy of the academic review process and provides robust reporting features. Data will be readily available to individual faculty; to department chairs, deans, and review committees; and to the public through California Public Records Act requests (these data have always been subject to the Public Records Act, but not always accessible as a practical matter when on paper). Such increased access raises both opportunities and concerns about how the campus captures and uses information about UCLA faculty.

Scenario 2: Capture and analytics of student behavior data for student success. Online learning allows capture of student activity and behavior at very granular levels. These data—often nontraditional metadata such as what time of day or night assigned readings were accessed—can be used by instructors in conjunction with more traditional indicators to evaluate students and guide them, in a highly personalized manner, to patterns of success. Such monitoring is new territory for the ethics, privacy, and autonomy of students. If data are shared with third parties, such as a course management systems vendor, further concerns over ownership and voice in aggregated data products arise. Similarly if data are shared with advisors, parents, prospective employers, graduate schools, or other parties, a range of ethical and legal issues arises.

Scenario 3: Sharing data about members of the UCLA community with external organizations. Cloud services rapidly have become integral to the services provided by universities, governments, and companies alike. Such outsourcing inherently exposes data about our community to third parties. Even vendors of purchased proprietary systems are gaining access to UCLA data that they can combine with data from other institutions, and potentially sell their insights. Similarly, scholarly data products are attractive as they provide sophisticated, aggregated patterns and trends that show performance within campuses and across institutions. This is new territory for ethics, privacy, autonomy, and control of data about our community. Among the concerns of sharing data with third parties are how to maintain trust when the entities involved may have different goals, and therefore different principles and values that drive their decisions; and how to protect metadata, algorithms, and analytics.

These three scenarios differ in domain (e.g., student, faculty, and other members of the community), in portability (e.g., degree of internal vs. external control of data), and in degree of protection by law (e.g., student records, faculty personnel records, patient records). However, their commonalities are greater than their differences. Our concerns in each of these scenarios address the power of data aggregation, the re-identification of individuals, risks of misinterpretation, perceived value to stakeholders, competing values and business models, expectations of confidentiality, expectations of reuse, the availability of multiple data streams, quality control, verification, interoperability, legal constraints and gaps, and ethics. These properties characterize the need for governance mechanisms focused on appropriate use of data rather than on particular data elements or owners, thus assuring that University principles are applied to data governance.

III. Meeting the Challenges and the Charge

UCLA has long been a leader in joint Academic Senate and Administration governance of information technology, establishing models and principles that are adopted throughout the University of California and elsewhere in higher education. Our inquiries suggest that UCLA is the first major research university to address the data governance issues in the scope of this charge. UCLA established the Information Technology Planning Board in 2000 and the Privacy and Data Protection Board in 2005. Members of these boards played essential roles in the UC President's Privacy and Information Security Initiative (2014).

In governing information technology over the last 15 years, UCLA faculty, staff, and students have addressed challenges such as design, deployment, budget, privacy, information security, and intellectual property. Safeguards against surveillance and inappropriate monitoring of behavior are pervasive concerns, which often overlap with questions of appropriate uses of data about people. University responsibilities for managing intellectual property, security, and integrity of our technical infrastructure and of contents are essential but outside the scope of this report. Other campus entities are deeply engaged in cybersecurity and intellectual property issues for UCLA.

In considering our charge, the Task Force drew heavily from the UC Privacy and Information Security Initiative's work throughout, beginning with its operating principles (2014, p. 8):

- We must maximally enable the mission of the University by supporting the values of academic and intellectual freedom.
- We must be good stewards of the information entrusted to the University.
- We must ensure that the University has access to information resources for legitimate business purposes.
- We must have a University community with clear expectations of privacy and information security*—both privileges and obligations of individuals and of the institution.
- We must make decisions within an institutional context.
- We must acknowledge the distributed nature of information stewardship at UC, where responsibility for privacy and information security resides at every level.

This report is organized by the three charges to the Task Force—principles, governance structure, and governance operations—preceded by a discussion of the scope of data to which these principles and governance mechanisms apply. We conclude with recommendations to the UCLA Academic Senate and campus administration. While this report is focused on the UCLA context, in keeping with our charge, we recognize that our findings have a wider audience in the University of California and in higher education generally.

* "and information security" has been added to the original text.

1. Scope of Data

Determining the scope of data was our most difficult task. The charge letter refers to "data about the members and activities of our campus community," and states "Good data governance gives assurance to the UCLA community that the campus is acting with credibility and trust with respect to data about its faculty, students, and staff." We struggled with questions of how to define "campus community," and how to bound the categories of "faculty, students, and staff." These are notoriously difficult groupings that vary by context throughout the University. Students may also be staff, if they are employed by UCLA in any capacity, whether in dining halls or research labs. Similarly, alumni, volunteers, donors, and patients are considered members of the UCLA community in certain contexts, but concerns for data governance differ widely. All persons affiliated with UCLA in any of these ways have reasonable expectations that their privacy will be respected and their information be held securely. However, not all of these cases necessarily fall within the scope of the charge.

We also struggled with distinctions between data about individuals collected by the University, collected by other parties, held by the University, used by the University, or used by other parties, regardless of origins. Partitioning data by the purposes for which they were collected was of limited value, as many of our concerns focus on the reuse of data for other purposes, regardless of their origins. We also considered legal and scholarly definitions of data (Borgman, 2015).

A further challenge was to distinguish between access to data and uses of data. In terms of information retrieval, it is helpful to distinguish between discovery, location, retrieval, and use. One must first discover the existence of a database or other store of data. Discovery may require a directory or index of campus data resources. No such directory currently exists, and to create and maintain one is a complex and expensive process. If individuals inside or outside the University can determine easily where information about individuals exists, that directory itself becomes a target for cyberattack, for legal discovery, and for all manner of research and administrative questions. Conversely, if no inventory of available data resources on campus exists, it is difficult to draw boundaries around what data must be governed. Not all requests for access to data or uses of data fall within the scope of the charge, however. The proposed governance process will distinguish between access to data and appropriate uses of data, where relevant.

Law, Policies, and Contracts

We first considered how data about UCLA individuals are subject to laws, university policies, and contractual requirements. Most of these address safeguarding the privacy of individuals through constraints placed on collecting and using personal information. For example, data collected for research projects are governed by human subjects regulations, funding agency policies, and community practices; patient and health records are governed by HIPAA and California Confidentiality of Medical Information Act regulations; and student records are

generally governed by FERPA. Overlaid on these sectorial laws and policies are those that cut across domains. For example, the UC Electronic Communications Policy (2005) provides a uniform approach to maintaining the privacy of a community member's electronic communications within the institution. Federal and State open records laws also span domains, but are intended to provide transparency of public institutions by requiring disclosure of email (and more) upon formal request. These laws and policies are in flux, necessitating a model that allows the University to remain flexible in light of changes in laws, societal norms, technological change, individual expectations, and University needs (University of California Privacy and Information Security Steering Committee, 2014, p. 18).

The Task Force initially concluded that we should be most concerned with data about individuals not covered by existing law, policy, or contract. For example, if FERPA applies, no further analysis would be needed. However, even when laws or policies define boundaries, the discretionary space within those boundaries can be vast. Few "big data" predictive analytics run afoul of current law and policy, for example, even as the expectations of our community may delineate different boundaries. In other cases, multiple laws or policies apply, and part of the analysis is to understand that intersection in the specific context.

Data Types

Using the scenarios and the consideration of laws, policies, and contracts outlined above, we generalized data types that would outlive today's examples. These included data collected about our community, regardless of the source. Our assessment resulted in examples that fell within or outside the scope for data governance.

Types of data within our scope of governance:

- Data the campus possesses about any UCLA person; i.e., staff, faculty, students
- Data that are identifiable by name or easily linked to a person
- Data the campus possesses on any person generated during the scope of the person's business with the University, including data sent to someone at the University

Types of data outside our scope of governance:

- Research data under the purview of IRB regulations
- Protected Health Information (PHI) governed by HIPAA, or individually identifiable health information in campus student healthcare facilities

In summary, we found that many forms of data about individuals in our community are sensitive, whether alone or in combination with other data; independent of whether those data meet a legal or policy definition of PII; and, more generally, independent of whether they are currently governed by any existing law, policy, or contract. The traditional Fair Information

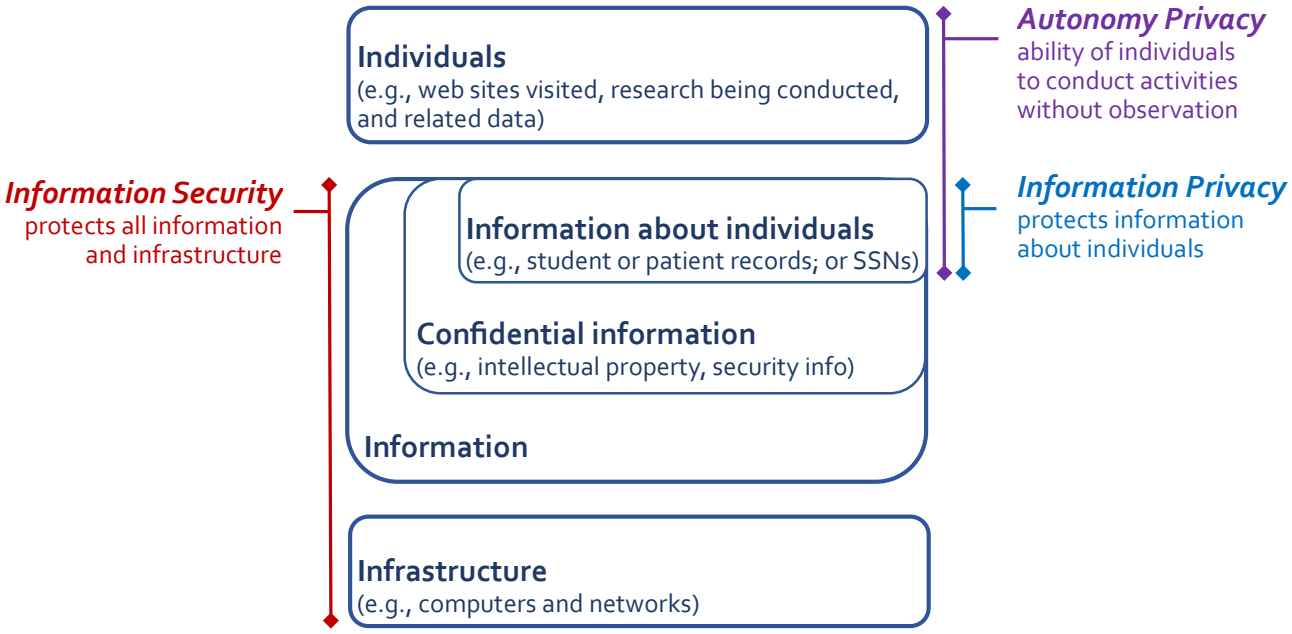
Practices principles of notice and consent continue to be necessary, but are no longer sufficient. Campus data governance mechanisms must address the uses and reuses of data about individuals in our community.

2. Principles

Given the work the University of California has devoted to establishing principles for privacy and information security in recent years, we concluded that data governance for UCLA should build upon principles already established.

We began by inventorying relevant publications, reports, and policy documents. From these materials we identified foundational definitions of, and relationships between, information security (confidentiality, integrity, and availability), information privacy (Fair Information Practices, or FIPs), and autonomy privacy (safeguards from surveillance and the monitoring of behavior) (University of California Privacy and Information Security Steering Committee, 2014) as shown in Figure 1 below.

Figure 1. Relationships between autonomy privacy, information privacy, and information security.



Next, we identified existing statements and principles already adopted or in use by UC or UCLA that the Task Force believes can serve as a foundational basis for data governance. These documents (for which bibliographic references and links can be found in Appendix C) include:

- Belmont Report
- Code of Fair Information Practices

- UC Statement of Privacy Values and UC Privacy Principles
- UCLA True Bruin (UC Statement of Ethical Values and Standards of Ethical Conduct)
- UCLA Principles of Community (Regents Policy 4400: Policy on University of California Diversity Statement)
- UCLA Principles of Scholarly Research and Public Records Requests

Two noteworthy themes emerged from an environmental scan for other principles that could be applied to data governance. First, given the difficulties in implementing traditional notice and consent in today's data-intensive environments, debate focuses on constraining the use of data rather than on constraining data collection (Pfeifle, 2014). Second, despite many efforts in the education sector to develop analytic frameworks for student learning and success, few documents provide practical guidance in balancing data governance decisions (see Appendix D).

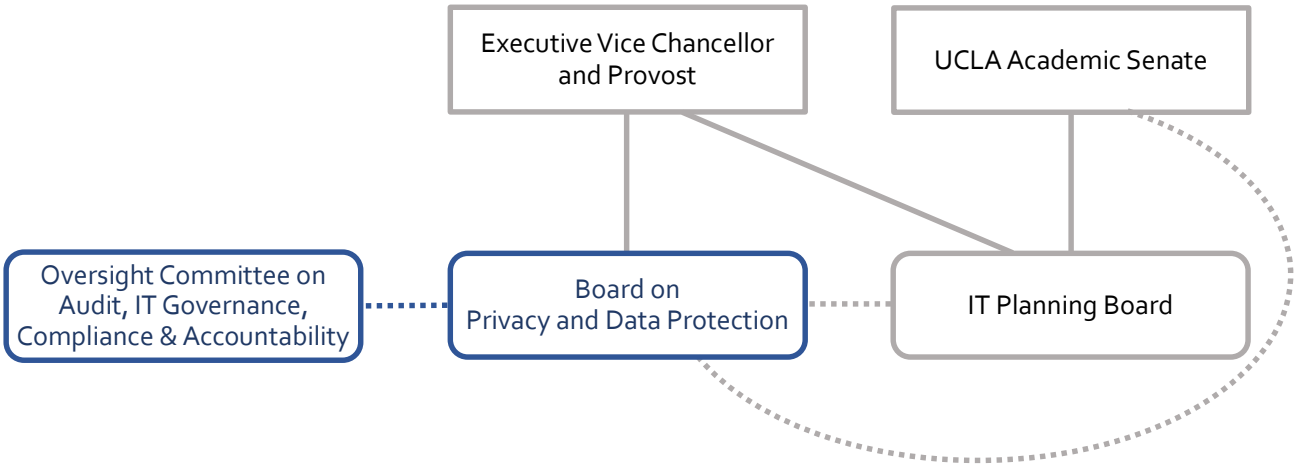
3. Governance Structure

The three UCLA scenarios present difficult institutional questions about data governance that cut across the concerns of the Academic Senate, University administration, and student governing bodies. Indeed, these questions are hard because different parts of the institution have legitimate but competing interests. UCLA's success in joint Academic Senate and Administration governance of information technology provides the foundation to address the campus's strategic data governance needs.

Our goal is to leverage existing structures and expertise, and thus align data governance with the privacy balancing process established in the report of the UC Privacy and Information Steering Committee (2014, p. 18). These governance structures, functional offices, and processes already address issues of autonomy, privacy, information privacy, information security, and data (see Appendix E).

As currently configured, the practices do not holistically address campus governance of data collection and use. The gaps are especially notable when contracts are involved, whether with other universities or with private sector companies; and when the value of our data comes to the fore. To address our data governance needs, the campus should build upon existing joint Academic Senate and Administration governance of information technology, as shown in Figure 2 below.

Figure 2. Campus entities responsible for IT governance, their existing reporting relationships, and the proposed framework for governing appropriate uses of data.



The Board on Privacy and Data Protection (Appendix F) is suited to take on the analysis role for the institution for these reasons:

- The Board has experience in the complex balancing processes for privacy and data protection akin to those arising in the UCLA scenarios on p. 6.
- The Board's membership has the broad representation necessary to address data governance issues. These include a balanced number of faculty members and administrators, plus undergraduate and graduate representatives. Key administrative offices represented include legal counsel, privacy, information security, ethics and compliance, and the health sciences.
- The composition of the board is determined jointly by the Academic Senate and the Provost. Senate leadership approves appointments of faculty members to the Board. One of these faculty members chairs the board; both the Senate leadership and the Provost approve the chair appointment. The Provost appoints staff members to the board, in consultation with the board chair.
- The Board reports to the Executive Vice Chancellor and Provost and has dotted line relationships to the Oversight Committee on Audit, IT Governance, Compliance and Controls (Appendix F, p. 28) and the IT Planning Board (Appendix F, p. 29).

The Board must ensure on a case-by-case basis that experts knowledgeable in the domain of a case are included in the decision-making process.

The Oversight Committee on Audit, IT Governance, Compliance and Accountability has key membership and authority to approve recommendations from the Privacy and Data Protection

Board with respect to appropriate data use. The senior executive of any unit impacted by a recommendation should be present for Committee discussion and approval processes.

This governance process should:

- Resolve legitimate disagreements and provide a path forward
- Promote transparency
- Promote open discussion

4. Governance Operations

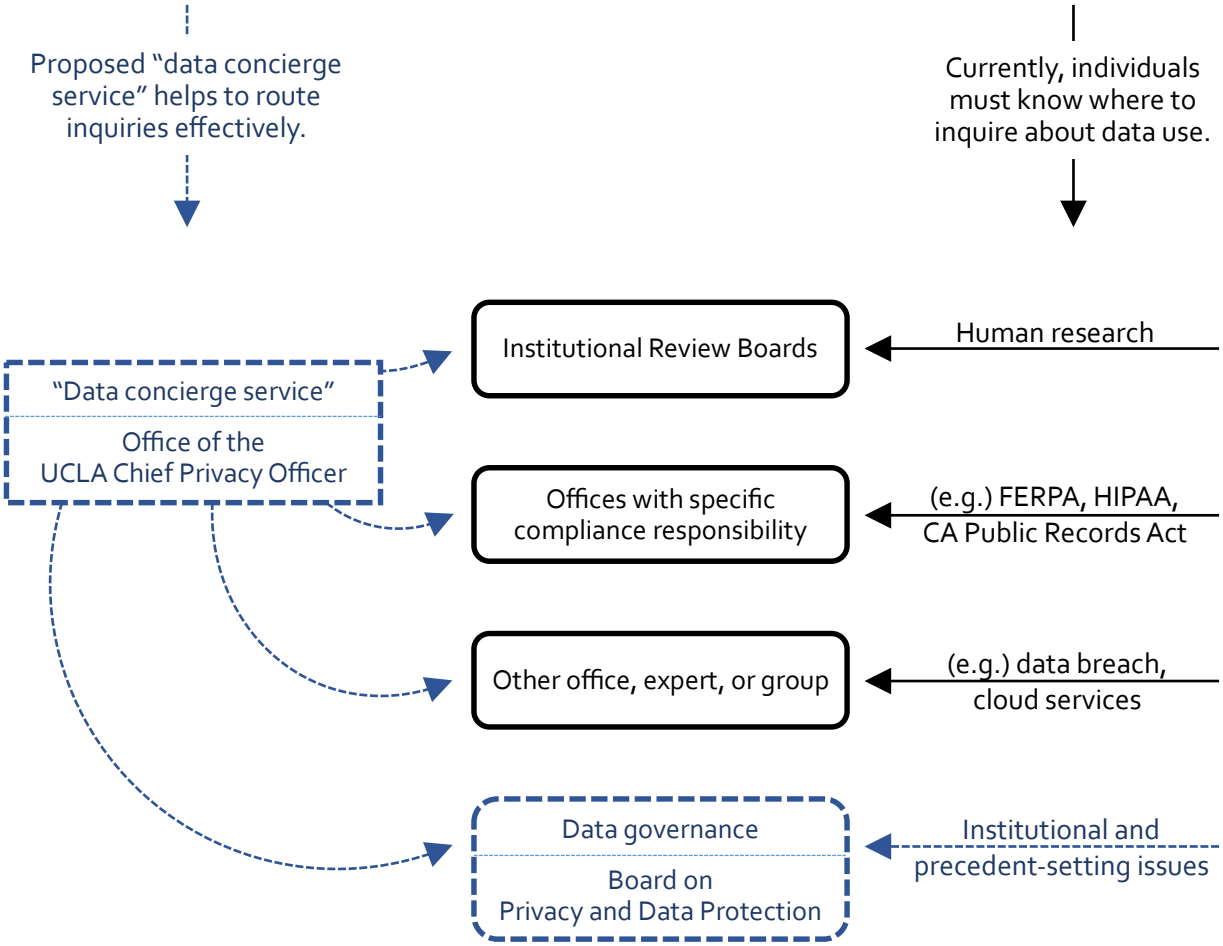
Similarly, data governance should build on existing governance operations, such as the Privacy and Data Protection Board, Information Technology Planning Board, and the overall operational structure of Figure 2 above.

The operational structure needs to create mechanisms that address the following:

- When data are used to make decisions about people
- When data are collected about people without their knowledge or consent
- When data about people are used in unexpected ways without subjects' knowledge or consent (e.g., new applications of data or systems; mining, analysis, and aggregation)
- When data are used for evaluation purposes
- When data are shared with external entities, whether with research partners or through service contracts with the private sector

Given the array of campus authorities, mechanisms, and entities that govern data in specific circumstances, a "concierge service" should direct inquiries to appropriate units. This service will ensure efficiency, build institutional memory when setting precedents, and avoid overloading the governance structures proposed below. The concierge can provide answers directly, when possible and appropriate; but otherwise will direct inquiries to appropriate campus entities (Figure 3 below). The concierge will direct inquires to the Board on Privacy and Data Protection when competing interests, goals, values, or obligations exist, and when new data services are proposed. The intent of the concierge service is triage and navigation, rather than to duplicate the functions of existing campus entities.

Figure 3. Workflow for proposed "data concierge service" and requests for access to data that fall within the scope of the governance process.



Multiple campus entities should be involved in educating the community about appropriate data use and the data concierge service. Faculty, research staff, or other members of the community whose normal activities do not intersect with IRB, HIPAA, or PHI may inadvertently become data stewards. Increasingly, graduate and undergraduate students also become data stewards. These stakeholders often are unaware of relevant campus policies or support infrastructures. Conversely, the campus may be unaware of data collected for research, teaching, learning, or administrative projects that are sufficiently sensitive to be worthy of governance and protection.

IV. Recommendations

The Task Force has devoted nearly 18 months to developing recommendations for how UCLA can balance tremendous opportunities to collect, analyze, interpret, and act upon data about our community with the potential for eroding trust and the relationships built upon that trust.

Implicit in these recommendations is the continuing need to bring together the many campus stakeholders in data and to recognize that stakeholders in data change over time. As collaborative initiatives emerge, whether with internal or external partners, the data governance process should be deployed early, and perhaps often. The IT Planning Board and the Privacy and Data Protection Board will be important resources when further input, deliberation, or formal Academic Senate–Administration channels are needed.

Recommendation 1. Establish data governance at UCLA on the following statement of scope.

The scope of data to be governed includes:

- Data the campus possesses about any UCLA person; i.e., staff, faculty, students
- Data that are identifiable by name or that can easily be linked to a person
- Data that the campus possesses on any person that was generated during the scope of the person's business with the University, including data that were sent to someone at the University

Data within the above parameters should be considered in scope even if they are not personally identifiable information (PII) as defined by existing law or policy. Such data should also be considered in scope whether they are identified as "records" or as "non-records" by University of California policy ("UC RMP-2: Records retention and disposition," 2004, p. 2).

The scope of data to be governed excludes:

- Research data under the purview of IRB regulations
- Protected Health Information (PHI) governed by HIPAA, or individually identifiable health information in campus student healthcare facilities

Recommendation 2. Extend data management work already undertaken by campus to include data that are in the stated scope of data governance.

The Task Force was charged with identifying best practices for data governance, which in turn requires best practices for managing those data. The UCLA "Book of Record" initiative (Appendix G), conducted in parallel with the development of the Opus Faculty Information System, established an inventory of data collections and responsible data stewards for Opus.

This framework should be extended to encompass all data sources within the scope. While the initial "Book of Record" project focused on Opus, the resulting inventory was intended for general campus use. The recommendations articulated in the project's report (Appendix H) are a good starting point for understanding the issues involved in identifying data collections and assigning responsibility for them.

Recommendation 3. Build upon established fair information practices principles for privacy and extend these principles to account for appropriate uses of the data as technology, practice, and policy evolve.

The principles from the UC Privacy and Information Security initiative report, Fair Information Practices principles, and the Belmont Report remain foundational. These principles encompass both the traditional approach of notice and consent and the use of data, whether alone or in combination with other data elements or sources.

The principles of notice and consent remain necessary, but are no longer sufficient. New principles for data governance must also address appropriate usage of data in ways that respect autonomy privacy, information privacy, and information security, as defined in Figure 1 above. Appropriate use must also consider other policies and practices that may apply in a given context such as requirements for data sharing and reuse, licensing of data resources, and domain-specific practices such as embargoes. These factors and other legitimate institutional concerns for data assets of the university must be balanced. Further background useful for such balancing is provided in Appendix D.

Recommendation 4. Extend existing structures and practices for governing information technology at UCLA to the operational framework for data governance.

Expand the scope of two existing committees to determine appropriate uses of data:

- Charge the Board on Privacy and Data Protection (Appendix F) as the primary campus entity to analyze, deliberate, and make recommendations about appropriate uses of data. The Board must ensure that on a case-by-case basis, experts knowledgeable in the domain of a case are included in the decision-making process.
- Charge the Oversight Committee on Audit, IT Governance, Compliance, and Controls to make final decisions regarding recommendations from the Board on Privacy and Data Protection and to use the Committee's existing authority to direct functional offices to make policy or operational changes, as appropriate. The executive head of any unit potentially impacted by a decision should be present and involved for the Committee's deliberation.

These governance mechanisms should be invoked when competing privacy interests, goals, University values, or obligations in the application or uses of these data exist and for which no statutory provision, law, or university policy is directly applicable.

Recommendation 5. Develop programmatic activities necessary to support effective data governance.

Given the array of campus authorities, mechanisms, and entities that govern data in specific circumstances, a "data concierge service" should direct inquiries to appropriate units. This service will ensure efficiency, build institutional memory when setting precedents, and avoid overloading the governance structures proposed below. The concierge can provide answers directly, when possible and appropriate, but otherwise will direct inquiries to the relevant campus entities (Figure 3 above). The concierge will direct inquires to the Board on Privacy and Data Protection when competing interests, goals, values, or obligations exist, and when new data services are proposed. The intent of the concierge service is triage and navigation, rather than to duplicate the functions of existing campus entities.

The data concierge service, in concert with other campus entities, should promote education and outreach efforts in support of data governance. Individuals, whether faculty, staff, or students, need to be aware of their responsibilities as data stewards. Members of the community too often become inadvertent data stewards through participation in research, evaluation, or administrative projects. Collaboration is needed between campus entities responsible for privacy, information technology, and security (Appendix E); the Academic Senate; and stakeholders such as data collectors, data stewards, subjects of data collection, and aggregators who acquire data from the campus.

Recommendation 6. Develop processes, policies, and tools to protect community data – including metadata, algorithms, and analytics – when relationships are established with third parties.

The University imposes requirements for security, privacy, ownership, and management practices on third parties with whom data are shared. The procurement process articulates these obligations as standard systemwide contract language. Equivalent protections for metadata, algorithms, and analytics have yet to be addressed in a similar uniform manner. Examples include the ability of a third party to infer behavior of individuals, to make unauthorized uses of data in the design of research applications or services, or to monetize insights based on such uses of our data. The campus should employ data governance mechanisms to implement comparable requirements for the use of data about our community.

Similar concerns arise when individual members of the community use third party services in the absence of standard university contracts. For example, if a faculty member requires students to use a specific Internet-based service to participate in a course, and that service is not contracted through the university, students are then obligated to agree to the service's standard terms and conditions as individuals. Such terms may contradict data governance principles implemented in university contracts. The campus should develop policies, procedures, tools, or guidance as appropriate to assure that data governance principles are observed consistently. Concerns about privacy and data use frequently arise with social media and online technologies (i.e., storage, cloud based services and analytic tools) in teaching and research.

The Office of Information Technology should be charged with facilitating the necessary campus discussions, governance reviews and approvals, and implementation of these recommendations as part of its mandate for addressing academic data issues (Appendix E, item 6).

Recommendation 7. Review and assess the scope, principles, and operations for data governance approximately 12 to 18 months after the recommendations of this report are implemented.

A review of cases, actions, and operations is an important opportunity to look holistically at what the campus has achieved, and what we may need to change to do better.

Appendix A. Charge Letter

UCLA Office of the Executive Vice Chancellor and Provost

August 29, 2014

Senior Campus Counsel Amy Blum
Professor Christine Borgman (co-chair)
Project Director Meg Buzzi
Professor Christina Christie
Director Mike Lee
Professor Vickie Mays
Director Kristen McKinney
Director Kent Wada (co-chair)
Director Kelly Wahl
Professor Neil Wenger
Student representatives (TBD)

Dear Colleagues:

UCLA continues to generate an ever-expanding amount of data about the members and activities of our campus community. These data offer new opportunities for increasing the efficiency of campus daily operations; for instructional enhancement; and for accreditation and governmental review. Commercial entities with other purposes also are interested in such data. On the other hand, we have numerous examples from outside UCLA about both the misuse and unintended consequences of big data. This points to the need for UCLA and members of the UCLA community to have a consistent set of expectations about ethical and appropriate use of such data, as well as security, accuracy, and compliance obligations. We also need a governance structure that ensures those expectations are met.

Good data governance gives assurance to the UCLA community that the campus is acting with credibility and trust with respect to data about its faculty, students, and staff. While *ad hoc* discussions on data policies and practice have occurred at UCLA, they have been limited in scope. As the demand for data increases, it is incumbent upon us to articulate underlying principles about the use of these data and to ensure that we have the capacity through our governance mechanisms to consider campus data holistically and address both current and potential issues.

We are therefore charging a joint Academic Senate–Administration task force to define those principles and to recommend how institutional data governance should be structured and operated for UCLA, including mechanisms to address new data requests, resolve conflicts, and align policy and practice among data stewards.

We would like you to serve as members of this new Task Force. We ask the Data Governance Task Force to provide recommendations for an initial set of principles by which governance

decisions can be considered, a UCLA data governance structure, and necessary surrounding processes by March 1, 2015. The Task Force should begin with existing policy and practice regarding campus data usage and should consider models from comparable institutions. Please address the following:

1. *Principles:* What are the initial operating principles that will guide UCLA when considering how data about members of its community should be used?
2. *Governance structure:*
    a. Is UCLA's existing governance—such as the IT Planning Board, the Board on Privacy and Data Protection and the Oversight Committee on Audit, IT Governance, Compliance and Accountability—robust enough to address and administer the growing concerns regarding data governance and usage?
    b. If so, where in the existing structure should this responsibility lie? If not, what should be the scope and authority of any new entity charged with those responsibilities; what would be the criteria for membership on this new governance structure; and how would it be appointed?
    c. How should this structure be initiated to ensure it will scale?
    d. How should the governance structure be staffed and resourced to do its work?
3. *Governance operations:*
    a. How should the governance structure work with individual data stewards?
    b. How can individual students, faculty, and staff participate in decisions regarding the use of data about themselves?
    c. How should the campus identify and address concerns, conflicts, and opportunities?

We hope you will agree to serve on this important task force. Please contact Assistant Provost Maryann Gray if you are unable to do so or have questions. Otherwise you will be contacted soon for scheduling. Thank you for this important service to UCLA.

Sincerely,

Scott L. Waugh
Executive Vice Chancellor & Provost

Janice L. Reiff
Chair, Academic Senate

cc: Vice Provost Jim Davis
Assistant Provost Maryann Gray

Appendix B. Task Force Members

L. Amy Blum	Interim Vice Chancellor, Legal Affairs
Christine L. Borgman co-chair	Distinguished Professor and Presidential Chair in Information Studies, GSEIS
Meg Buzzi	Opus Project Director, Office of the Vice Chancellor, Academic Personnel
Kristen Chamberland	Graduate Student Representative, GSEIS
Christina A. Christie	Professor of Education, GSEIS Division Head, Social Research Methodology, GSEIS Member, North General IRB
Mike Lee	Associate Director, Social Sciences Computing
Vickie M. Mays	Professor, Department of Psychology Professor, Department of Health Services, Fielding School of Public Health Director, UCLA BRITE Center for Science, Research and Policy
Kristen McKinney	Director, Student Affairs Information and Research Office
Kent Wada co-chair	Director, Strategic IT Policy, Office of Information Technology UCLA Chief Privacy Officer
Kelly Wahl	Director, Statistical Analysis, Office of Academic Planning and Budget
Neil S. Wenger	Professor of Medicine, General Internal Medicine and Health Services Research Director, UCLA Health System Ethics Center Chair, Ethics Committee, Ronald Reagan UCLA Medical Center
(Unassigned)	Undergraduate Student Representative

Anna Joyce, Manager, Administrative Policies and Delegations, was asked by the co-chairs to join the Task Force as policy staff.

Appendix C. Existing Principles

1. The **UC Statement of Privacy Values** (University of California Privacy and Information Security Steering Committee, 2014, p. 14) declares privacy—of both autonomy and information—as an important value of the University, as this is not explicitly done elsewhere; and clarifies that privacy is one of many values and obligations of the University. This Statement is directly based upon the UCLA Statement developed some years earlier (UCLA Board on Privacy and Data Protection, 2011, p. 2).
2. The **UC Privacy Principles** (University of California Privacy and Information Security Steering Committee, 2014, pp. 15–17) are derived from the UC Statement of Privacy Values and established privacy principles. They are intended to guide policies and practice in conjunction with well-understood information security objectives of protecting the confidentiality, integrity, and availability of information resources. The UC Privacy Principles consist of principles that address both autonomy privacy and information privacy (Figure 1, p. 11).
3. The **Code of Fair Information Practices** (United States Department of Health, Education, and Welfare Secretary's Advisory Committee on Automated Personal Data Systems, 1973) long underpins international best practice (and laws such as HIPAA and FERPA) with its core notions of notice and consent.
4. The **UCLA Principles of Community** speak to the campus community's commitment "to discovery and innovation, creative and collaborative achievements, debate and critical inquiry, in an open and inclusive environment that nurtures the growth and development of all faculty, students, administration and staff." Regents Policy 4400: Policy on University of California Diversity Statement (University of California, 2010) is the systemwide statement. Both provide a cultural context within which institutional decision-making should occur.
5. The **Belmont Report** (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) is the basis for Institutional Review Boards to protect human subjects in research—a balancing act between enabling research that is our core mission and ensuring to the greatest extent possible that we do no harm.
6. **UCLA True Bruin** ("UCLA True Bruin," 2009) is the campus's program built around the core of the **UC Statement of Ethical Values and Standards of Ethical Conduct** (University of California, 2005). The UC Standards provide a foundational context for ethical decision-making, including a specific section on "Records: Confidentiality/Privacy and Access" (University of California, 2005, p. 5). This section also makes explicit the inherent balancing needed because of sunshine laws: "The legal provisions and the policies are based upon the principle that access to University of California information concerning the conduct of the people's business is a fundamental and necessary right of every person, as is the right of

individuals to privacy." It does not, however, give guidance as to how to adjudicate between these two fundamental and necessary rights.

7. The **UCLA Principles of Scholarly Research and Public Records Requests** (Joint UCLA Senate–Administration Task Force on Academic Freedom, 2012), announced by Chancellor Block in January 2014 (Block, 2014), provides the guidance not given in the previous item, articulating the reasoning used by UCLA when material related to scholarly research is requested under the California Public Records Act.

Appendix D. Principles Beyond FIPs

1. The **Asilomar Convention for Learning Research in Higher Education** (Asilomar Convention, 2014) speaks to six principles that "should inform the collection, storage, distribution and analysis of data derived from human engagement with learning resources. The principles are stated here at a level of generality to assist learners, scientists, and interested citizens in understanding the ethical issues associated with research on human learning." Their basis in FIPs and the Belmont Report is evident in the language of the document.
2. The **Responsible Data Analytics draft by the Department of Finance of the Australian government** (Australian Government, 2014) provides thoughtful and practical guidelines for considering data analytics.
3. An **analysis by Professor Paul Schwartz** for the Centre for Information Policy and Leadership, Hunton & Williams LLP Data Protection Law and the Ethical Use of Analytics (Schwartz, 2010) points out where FIPs falls short in today's "big data" world, and what may be done to close the gap.
4. **NIST Internal Report 8062 (draft): Privacy Risk Management for Federal Information Systems** (Garcia, Lefkovitz, & Lightman, 2015) is a recent draft document that offers new terminology that changes the way we think about the standard FIPs model.
5. **Control Use of Data to Control Privacy** (Landau, 2015) discusses a new approach to privacy as the traditional controls on collection of data (notice and consent) become meaningless or unrealistic in the era of "big data." One of a collection of articles in a special issue of **Science** on The End of Privacy (Enserink, Chin, Wible, & Jasny, 2015).

Appendix E. Campus Entities that Address Privacy, Information Security, and Data

1. The **Oversight Committee on Audit, IT Governance, Compliance and Controls** (Appendix F, p. 28) approves the campus annual audit and compliance plans and reviews results of the programs; provides policy direction and oversight to the campus on IT governance and financial controls and accountability programs; and advises the Chancellor on emerging audit, accountability, and ethics and compliance issues.
2. The **Information Technology Planning Board** (Appendix F, p. 29) is a joint Senate–Administration committee that is the primary governance and oversight body for IT at UCLA. The ITPB takes leadership in establishing UCLA's IT vision, strategic plan, and policy guidelines.
3. The **Board on Privacy and Data Protection** (Appendix F, p. 27) is the campus nexus for consideration of institutional privacy and data protection needs, when these needs must be balanced with the campus's many other values and obligations and must account for external trends in technology and individual expectations.
4. The UCLA **Institutional Review Boards** are required by federal regulations to review all human subjects research activities conducted, ensuring adherence to all federal, State, local, and institutional regulations concerning the protection of human subjects in research.
5. Various **functional offices with compliance responsibility** (e.g., Office of the Registrar for FERPA compliance, Office of the Human Research Protection Program that provides guidance and administrative support to the five IRBs, and the HIPAA privacy officers).
6. The portfolio of the **Vice Provost, Information Technology** (Waugh, 2015), through the Office of Information Technology and the Institute for Digital Research and Education, includes several overarching elements related to data, including: campus academic data, academic decision and shared data practices, and appropriation and technology transfer of data assets; educational technology platforms and analytics infrastructure alignment, integration, R&D and investment; and the Office of the UCLA Chief Privacy Officer.
7. The **Research Informatics Strategic Planning Board** is charged by the Vice Chancellor for Research and Vice Provost, Information Technology, to ensure cross-campus orchestration of the draft RISP plan. Goals include identifying "informatics areas that are transformative and can capture the imagination of private and public donors and corporate partners" and identifying, investigating and proposing "campus policies and practices that are currently unaddressed through existing campus structures, e.g. data sharing plans, practices around meta-data, etc."

Appendix F. Descriptions of Selected Campus Entities

1. Board on Privacy and Data Protection

The Board is charged with articulating institutional positions on privacy and data protection reflecting the campus's values and cultural expectations to guide policy development and decision-making. It is the campus nexus for considering initiatives, proposals, and stances that must balance privacy, data protection, and the campus's other values and obligations.

Membership

The Board maintains a balanced number of faculty and administration, plus one graduate and one undergraduate student representative. Faculty appointments should ensure social, cultural, technical, and management aspects of privacy and data protection. Administrative appointments should have direct involvement with institutional management of privacy matters.

The Chair and/or the Vice Provost, Information Technology represents the Board on interactions with the Oversight Committee on Audit, IT Governance, Compliance and Accountability as appropriate to the topic.

Voting Membership

- Faculty members equal to the number of administrative voting members (staggered three-year terms), including the faculty chair (two-year term)
- Chief Compliance Officer, UCLA Health System and David Geffen School of Medicine
- University Librarian
- University Registrar
- Vice Provost, Information Technology and Chief Academic Technology Officer (administrative vice chair)
- Associate Vice Chancellor, Information Technology Services
- Designee from Campus Human Resources
- Designee from the Office of the Campus Counsel
- Designee from the Office of the Vice Chancellor, Research
- One undergraduate student designated by the Undergraduate Students Association Council (one year term)
- One graduate student designated by the Graduate Students Association (one year term)

Non-Voting Membership

- UCLA Chief Privacy Officer
- UCLA Chief Information Security Officer
- Chief Information Officer, David Geffen School of Medicine
- Chief Data Officer, Health System
- Designee of the Executive Vice Chancellor and Provost
- Designee from Audit & Advisory Services

2. Oversight Committee on Audit, IT Governance, Compliance and Controls

The Oversight Committee on Audit, IT Governance, Compliance and Controls approves the campus annual audit and compliance plans and reviews results of the programs. The Committee also provides policy direction and oversight to the campus on information technology governance and financial controls and accountability programs. Finally, the Committee advises the Chancellor on emerging audit, accountability and ethics and compliance issues. Its responsibilities:

- Examine emerging audit IT governance/security compliance and financial controls and accountability issues and determine the overall direction that should be taken to address these issues on a campus-wide basis.
- Provide guidance in developing, clarifying and promulgating campus policy and where indicated, recommending policies and best practices for system-wide implementation with audit, IT governance/security, compliance and financial controls and accountability implications in areas where policy is determined to be non-existent, weak, or poorly communicated.
- Define the management concerns and issues that should be considered in developing the campus audit, IT governance and security, compliance and financial controls and accountability programs.
- Assess the effectiveness and consistency of compliance practices throughout the campus.
- Report compliance risk areas of high priority and propose risk mitigation activities to the UC Ethics and Compliance Risk Council.

Members

- Chancellor
- Executive Vice Chancellor and Provost
- Administrative Vice Chancellor
- Vice Chancellor and Chief Financial Officer (chair)
- Vice Chancellor, Legal Affairs

- Vice Chancellor, Research
- Vice Chancellor, Health Sciences
- Vice Provost, Information Technology and Chief Academic Technology Officer
- Dean, Life Sciences
- Senior Vice President and Chief Compliance and Audit Officer, UCOP

Non-voting participants

- Associate Vice Chancellor and Controller, Corporate Financial Services
- Associate Vice Chancellor, Information Technology Services
- Chief Information Officer, Health Sciences
- Director, Administrative Policies and Compliance
- Director, Audit and Advisory Services and Campus Ethics and Compliance Officer

3. Information Technology Planning Board

The Information Technology Planning Board (ITPB) is a joint Administration–Senate committee and the primary governance and oversight body for information technology and services at UCLA. The Board has purview over UCLA's information technology strategies, which include consideration of technology infrastructure, the increasingly interrelated realms of academic and administrative computing, and central vs. distributed deployment of resources.

The ITPB informs and advises UCLA's executive management and Senate leadership regarding information technology, including budget and policy matters, and strategic opportunities regarding investment in and deployment of information technology. In furtherance of these responsibilities, the Board reviews and makes recommendations concerning technology-related budget requests and opportunities stemming from a variety of corporate and industrial relationships. The Board's oversight activities — concerning local, central and partnership technology plans, budgets, policy, and progress — will both inform decision-making and ensure internal coordination. The ITPB takes the lead in establishing UCLA's information technology (IT) vision, strategic plan, and policy guidelines.

Members

The Board will be chaired by a member of the faculty. The Chair will be selected jointly by and maintain regular communication with the Executive Vice Chancellor and Provost and the Chair of the Academic Senate. The Board will be composed of current senior academic administrators, administrators, and additional faculty, with the faculty predominant. The Chair and members will be jointly determined by the Executive Vice Chancellor and Provost and the Chair of the Academic Senate.

- Faculty (12)
- Administration (8)
- Undergraduate student representative
- Graduate student representative

Appendix G. Opus Book of Record Initiative: Information for Data Stewards

Sponsors

- *Carole Goldberg—Vice Chancellor, Academic Personnel*
- *Steve Olsen—Vice Chancellor and Chief Financial Officer*
- *Jim Davis—Vice Provost, Information Technology*

Background: Opus

Opus is UCLA's faculty information system project. Opus has been funded to solve two fundamental business problems: UCLA's knowledge around, access to, and context for faculty data is insufficient, and the academic review process (currently paper-based), lacks transparency and efficiency. The core deliverable of the Opus project is both a faculty information system of record for UCLA and a reengineering of the academic personnel processes and workflow at all levels of the organization. It will provide campus with a central enterprise system that increases the accuracy and efficiency of academic advancement and reporting processes, while decreasing the administrative burden normally associated with these activities. The scope of the system encompasses the entire faculty lifecycle: from recruitment to separation.

As a faculty information system, Opus must provide current data about faculty teaching, research, and service activities. Rather than recreate data for which authoritative sources already exist, Opus will rely on data feeds from existing Books of Record to populate its data fields. For data that currently exists only in paper form, or within a system that will be subsumed by Opus, Opus will likely become the Book of Record.

Background: Book of Record Initiative

A Book of Record is an authoritative source for a data element or a set of data elements. As stated above, Opus will source faculty data from campus Books of Record wherever possible. However, to date the campus has never developed institutional policies around the management and stewardship of institutional data. Instead each data-owning unit manages its own set of regulatory policy requirements, and any associated UC and UCLA policy requirements and practices. This network of policies and practices impact how data is accessed, shared and/or coordinated across data-owning units, as well as the structure of the campus data warehouse. This collection of policy and practice-driven protocols has never been institutionally reconciled and documented, and no unit has documented authority over any set of data.

The key driver behind the Book of Record Initiative is the fact that for Opus to be successful, both administrative and academic users MUST trust the data in the system. As such, it is critical to assign and document campus Books of Record, and to coordinate and document the policies and practices surrounding the Books of Record that act as sources for Opus.

The outcomes of the Book of Record Initiative will include:

- List of Books of Record and Data Stewards mapped to specific Opus data elements to provide transparency to end users and ensure that Opus is using the most accurate/authoritative data
- Documentation of the data lifecycle for data elements within Opus to further data transparency, and to assist with error resolution
- Documented agreement regarding the responsibilities of data stewards to downstream systems such as Opus in order to ensure data quality and to protect data stakeholders

Workflow

1. High-level Business Process & Dataflow

In this initial interview, the Opus team will meet with the Data Steward and business technical lead. The goal of the interview will be to understand the context of the data at a high-level. This discussion tells the "story" of the data and provides a snapshot of the data lifecycle, data creators and users, important systems, and key contacts. Throughout the following phases, the Opus team will contact specified personnel, where additional clarification is needed.

2. System Architecture

After the initial interview, the Opus team will need to familiarize themselves with any prepared technical documentation about the systems and processes surrounding the specified data elements. Data dictionaries, data models, system architecture diagrams, screenshots, etc. will all be useful as we try to better understand the data. The outcome of this phase will be a vetted high-level model of data flow through the organization.

3. Data Mapping

During this phase, the Opus team will map each Opus data element to the tables and columns in the source system based on the data model and data dictionary. Where such technical documentation is not available, additional interviews may be necessary to ensure accuracy. The Opus team will likely need to meet with business technical staff at this phase to clarify definitional differences and ensure an accurate mapping. The outcome of the data mapping process will be shared with and vetted by the business technical team.

4. Data Lifecycle Documentation

Concurrent with the above phases, the Opus team will form an increasingly detailed understanding of the data management policies and processes throughout the lifecycle of each data element. Documentation of specific aspects of the data lifecycle (see below) are necessary to provide transparency to data stakeholders, ensure data quality, and to create a holistic understanding of faculty data across organizational boundaries. The outcome of this phase will be a model of the data lifecycle for each element as well as documentation of policies and practices where applicable. The Opus team will work with the business technical team as necessary to ensure shared understanding.

5. Data Stewardship Agreement

The final step in the process will be the development of a data stewardship agreement. The Book of Record Data Steward Workgroup designed the template for the agreement, which documents the responsibilities of a data steward to the campus as a whole as well as to downstream systems that use the data under their care. The template will be adjusted as necessary to accommodate differences among Book of Record units.

Data Lifecycle Documentation

For each identified campus unit, we hope to work with both the data steward (business unit) and managing unit (IT) to discover the following for each data element:

1. Data characteristics
2. Data creation/updating process
3. Data deletion/archiving process
4. Data usage/access policies
5. Data maintenance policies
6. Data interface
7. Data relay

Appendix H. Outcomes from the Book of Record Initiative as of September, 2014

(This draft represents the latest version of the document available.)

Completed

1. Identified data stewards for most major datasets needed by Opus
2. Identified gaps in data stewardship for remaining datasets needed by Opus
3. Identified where Opus could serve as a Book of Record for faculty data
4. Identified policy and process gaps

Upcoming

5. Define what data stewardship means in the context of Opus (UCLA?)
6. Agree on an error resolution process
7. Create SLAs with data stewards outlining responsibilities of stewardship/serving as Book of Record.

Issues and Policy Gaps

1. The University is lacking clear and accessible information about Data Stewardship and Book of Record systems.

Opus serves as an aggregator of faculty information sourced from several existing campus systems. Upon launching the project, it was unclear how the project team could ensure that the data source was authoritative. The project team frequently learned of existing systems only in conversation with other system owners. Other applications have encountered this issue as well, often resulting in the creation of shadow systems when no source can be found. The Book of Record Initiative was created as a first step in solving this problem.

Recommendation: Starting with data gathered from the Book of Record Initiative, create a central information resource linking specific datasets to data stewards, book of record systems, and current documentation.

2. Documenting data stewards and books of record from a project rather than an institutional perspective reveals that what is viewed as "authoritative" is dependent upon the context of the data/analysis.

The Book of Record process has been Project (Opus) focused rather than Institutionally focused, the result may not be a definitive list of Books of Record and Data Stewards, but rather a list of Data Stewards/Books of Records in the context of a faculty information system.

In Opus everything is viewed through the lens of the person, the faculty member and their activities, a very different perspective than one would expect of the institution as a whole.

Example: Enrollment data—AIM reports to UCOP but enrollment numbers do not include all students in the room (for purposes of financial reporting). Additionally, the algorithm used by AIM to create the faculty workload report for UCOP, while equitable, does not necessarily reflect the experience of teaching the course. On the other hand, the Registrar presents another view of enrollment based on the number of people in the room at any one time.

Recommendation: Develop an institutional perspective on master data management that takes into account the fact that authority is context-specific.

3. Fine distinctions in data ownership make it difficult to create a clean map of Data Stewards and Books of Record.

Responsibility for a set of data is frequently shared across organizational lines. This may be due to disconnected business processes, funding arrangements, informal agreements, etc. Rather than clear lines of authority, the result is a spiderweb of systems, and offices that may or may not be working ins

Example: Faculty Salary - Faculty salaries are approved by the Vice Chancellor for Academic Personnel on paper, but entered into the payroll system by Department Staff perhaps weeks after the paper approval. Processing errors, retroactive payments, as well as legitimate salary augmentations due to stipends, school-specific agreements, etc. also lead to a divergence between the academic personnel record of salary and the payroll record of salary. The distinction made by the Opus in concert with the data stewards was to assign stewardship for "approved" salaries to the Academic Personnel Office, and stewardship for "actual" salaries to the Office of the Controller.

Recommendation: Documentation of data stewardship and books of record will need to include an explanation of how the data is bounded and why. Such documentation can pave the way for improvements in the efficiency of data flow and clarity in determine stewardship of specific data.

4. Determining stewardship for self-reported data calls into question the concepts of stewardship and books of record.

Opus will be a source for faculty data not captured in existing systems. Examples include service or professional activities, conference presentations, and creative works. This data will be manually entered by the faculty or their proxies. It is not practical to confirm the accuracy of this data, therefore is it possible to speak of Opus as a Book of Record for this data? Who is the data steward in these cases? What responsibility does Opus have to make transparent the non-verifiability of the data?

Example: Faculty will enter data about their service activities, some of which may support local communities. The UCLA Office of Government and Community Relations would like to create a

report based on Opus data about service activities that took place in Los Angeles. The Office will share the data with the UC Regents and State Legislature to support increased funding. Is it problematic that the institution is making its case on self-reported data?

Recommendation: Ensure data sources are transparent to any downstream system or user, particularly when the data are self-reported. Data consumers should use caution when basing decisions on these data.

5. There is no standardized process for addressing data access questions and issues.

Data Stewards have voiced concerns about access that are unique to their data. For each data source it was necessary to uncover and address potential access issues. These concerns will be part of the Data Stewardship SLAs between Opus and the Data Stewards.

Example: The Office of Research Administration feels that grant proposal information is sensitive, subject to intellectual property concerns, and should be protected. Opus will only make the data available to the individual(s) listed on the proposal and give them an option to add the proposal to their academic review dossier, but will NOT give the individual the option to share the data publically. Additionally the data will not be included in reports available to administrators

Recommendation: As part of a larger effort to standardize data documentation, guide data stewards to include existing access constraints on the data. Data consumers may also want to make these constraints apparent to their own downstream users. Additionally, create a centralized process that standardizes data access policies according to the data classification standard and reviews allegations of inappropriate access and misuse.

6. Without institutional policy and process to support the initiative, there is no enforcement ability and little incentive for cooperation.

The Data Stewards we have worked with to date have been extremely cooperative and supportive of the Book of Record Initiative. However, as we proceed to create SLAs that outline the responsibilities of being a data steward and document the error resolution process there is little incentive to cooperate beyond good will - we are asking for investment of time and resource while offering little in return. Additionally, we currently lack an institutional process to determine data stewardship, to resolve disagreements among data stewards, and to hold data stewards accountable for good data practices.

Example: Opus will source data from several campus systems. These data may be amalgamated with other faculty data and then consumed by these same or other systems. If the Registrar found that Opus was adjusting enrollment data based on faculty complaints, then passing this altered data to other campus systems, what recourse would the Registrar have as the Data Steward for these data? How will using data from authoritative sources be enforced?

In what circumstances? What are the responsibilities of data consumers such as Opus to render the data as received from books of record?

Recommendation: Explicitly define the responsibilities of data stewardship and create a mechanism for holding data stewards accountable for data quality and access.

7. Definitions for high-level concepts are not consistent across offices and systems.

One aspect of master data management is standardization of definitions for core business concepts. For the university this might include the following entities "student", "faculty", "grant", "course", "employee", etc. We found that with each system it was necessary to redefine these terms and take note of where they were inconsistent. Occasionally it was necessary to come up with an alternative word or phrase that could encompass the multiple meanings that had been assigned to a concept.

Example: we were not able to find an authoritative definition for a UC employee. It was difficult to determine whether all sets of Academic Appointees were considered employees (as some of these serve in Voluntary or Without Salary positions). It was therefore difficult to ascertain whether we would be able to collect data on our full population.

Recommendation: Standardize definitions of conceptual data elements that are used across organizations and systems.

8. System updates/upgrades are occurring with greater frequency necessitating increased communication with downstream systems.

Over the last two years, several campus offices have upgraded their systems resulting in new data, logic, and definitions. The Opus team (and others downstream systems) must adjust its data mapping, data definitions, and other metadata and documentation for each system as these updates occur. Currently, not all systems have shareable documentation. Extra time and effort must be spent to ensure data stewards and consumers have a shared understanding of the data.

Recommendation: Create a campus-wide standard for data documentation and a shared repository where authorized users can access the documentation.

9. The University does not provide guidance on best practices for data management, nor does it educate data consumers about institutional data.

Data management practices were frequently ad hoc, changeable, and at the discretion of the data steward. Additionally, UCLA faculty and staff were often ignorant of data access restrictions, the level of risk in disclosing certain data, the availability of the data for a public records request, which data was already maintained in campus systems, how data was currently being used, etc.

Recommendation: Create or make visible expectations around data accuracy and reliability, standards related to documentation, storage, security, accessibility, etc. If possible, provide user-facing metadata or help text that discloses data source(s), retention, access, acceptable uses, definitions, and contact information for the steward.

10. While access policies sometimes exist for specific data about an individual, the University has not yet developed guidelines for dealing with aggregated data.

Recommendation: Develop a process/tool for data aggregation risk assessment (based on what?...)

Appendix I. Bibliography

Asilomar Convention. (2014). The Asilomar Convention for Learning Research in Higher Education. Retrieved from http://asilomar-highered.info/asilomar-convention-20140612.pdf

Australian Government, D. of F. (2014). Australian Government — Responsible Data Analytics. Retrieved from http://www.finance.gov.au/sites/default/files/Responsible%20Data%20Analytics%20Draft.pdf

Block, G. (2014, January 10). Announcement of the UCLA Principles of Scholarly Research and Public Records Requests by UCLA Chancellor Gene Block. Retrieved February 13, 2015, from http://chancellor.ucla.edu/updates/principles-of-scholarly-research-and-public-records-requests

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.

Enserink, M., Chin, G., Wible, B., & Jasny, B. (Eds.). (2015). The end of privacy. *Special Issue, Science*, *347*(6221), 490–491. http://doi.org/10.1126/science.347.6221.490

Garcia, M., Lefkovitz, N., & Lightman, S. (2015). *NISTIR 8062 (Draft): Privacy Risk Management for Federal Information Systems* (Internal Report No. 8062). National Institute of Standards and Technology. Retrieved from http://csrc.nist.gov/publications/PubsDrafts.html#NIST-IR-8062

Gellman, R. (2014). Fair Information Practices: A Basic History. Retrieved from http://www.bobgellman.com/rg-docs/rg-FIPShistory.pdf

Joint UCLA Senate–Administration Task Force on Academic Freedom. (2012, September). Statement on the Principles of Scholarly Research and Public Records Requests. Retrieved from https://www.apo.ucla.edu/resources/academic-freedom

Landau, S. (2015). Control use of data to protect privacy. *Science*, *347*(6221), 504–506. http://doi.org/10.1126/science.aaa4961

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Pfeifle, S. (2014, May 27). DPI Dispatch, The Future of Data Collection and Use; Notice and Choice. Retrieved from https://iapp.org/news/a/dpi-dispatch-the-future-of-data-collection-and-use-notice-and-choice/

Schwartz, P. M. (2010). *Data Protection Law and the Ethical Use of Analytics*. The Centre for Information Policy and Leadership, Hunton & Williams LLP. Retrieved from http://www.huntonfiles.com/files/webupload/CIPL_Ethical_Undperinnings_of_Analytics_Paper.pdf

The Organisation for Economic Co-operation and Development. (2013). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Retrieved July 31, 2014, from http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

UCLA Board on Privacy and Data Protection. (2011, May). UCLA Statement on Privacy and Data Protection—For campus discussion. Retrieved from http://privacyboard.ucla.edu/documents/privacystatement-20110531.pdf

UCLA Office of the Human Research Protection Program. (2013, July 2). Guidance: Determining Which Activities Require UCLA OHRPP/IRB Review. Retrieved from http://ora.research.ucla.edu/OHRPP/Documents/Policy/3/Activities_Requiring_Review.pdf

UCLA True Bruin. (2009). Retrieved September 11, 2015, from http://www.truebruin.ucla.edu/

United States Department of Health, Education, and Welfare Secretary's Advisory Committee on Automated Personal Data Systems. (1973). *Records, Computers and the Rights of Citizens*. Retrieved from http://epic.org/privacy/hew1973report/default.html

University of California. (2005). UC Statement of Ethical Values and Standards of Ethical Conduct. Retrieved from http://www.ucop.edu/ethics-compliance-audit-services/_files/stmt-stds-ethics.pdf

University of California. (2010, September 16). Regents Policy 4400: Policy on University of California Diversity Statement. Retrieved from http://regents.universityofcalifornia.edu/governance/policies/4400.html

University of California Business and Finance Bulletin RMP-2: Records retention and disposition: principles, processes, and guidelines. (2004, September). Retrieved from http://policy.ucop.edu/doc/7020454/BFB-RMP-2

University of California Electronic Communications Policy. (2005, June). Retrieved from http://policy.ucop.edu/doc/7000470/ElectronicCommunications

University of California Privacy and Information Security Steering Committee. (2014). *Privacy and Information Security Initiative, Steering Committee Report to the President*. University of California. Retrieved from http://ucop.edu/privacy-initiative/

Waugh, S. (2015). Reappointment of Vice Provost Jim Davis. Retrieved from http://evc.ucla.edu/announcements/reappointment-of-vice-provost-jim-davis-1

Appendix J. Acknowledgements

- James F. Davis, Vice Provost, Information Technology and Chief Academic Technology Officer
- Janice L. Reiff, Past Chair, Academic Senate
- Anna Joyce, Manager, Administrative Policies and Delegations and policy staff to the Task Force
- George Mood, for editorial assistance
- Julian Gautier, for bibliographic assistance