BIG DATA ETHICS
Lessons from UCLA’s Approach to Big Data Governance

Evan Selinger
Associate Professor of Philosophy & MAGIC Center Head of Research Communications, Community & Ethics, RIT

Kent Wada
UCLA Chief Privacy Officer
Director, Strategic IT Policy

These slides have been edited for readability when “builds” were used and to replace certain material with links to original sources for copyright purposes.
(Big) Data Governance: Why Now?

Scenarios
**Scenario: The Boyfriend Dilemma**

**Step Six (Student):** Indicate which colleges you want to receive your FAFSA information.

Enter the six-digit federal school code and your housing plans. You can find the school codes at [www.fafsa.gov](http://www.fafsa.gov) or by calling 1-800-4-FED-AID. If you cannot obtain the code, write in the complete name, address, city and state of the college. For state aid, you may wish to list your preferred college first. To find out how to have more colleges receive your FAFSA information, read *What is the FAFSA?* on page 10.

<table>
<thead>
<tr>
<th>103. a</th>
<th>1st FEDERAL SCHOOL CODE</th>
<th>NAME OF COLLEGE</th>
<th>ADDRESS AND CITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>STATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. b</td>
<td>on campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with parent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>off campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. c</td>
<td>2nd FEDERAL SCHOOL CODE</td>
<td>NAME OF COLLEGE</td>
<td>ADDRESS AND CITY</td>
</tr>
<tr>
<td>OR</td>
<td>STATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. d</td>
<td>on campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with parent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>off campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. e</td>
<td>3rd FEDERAL SCHOOL CODE</td>
<td>NAME OF COLLEGE</td>
<td>ADDRESS AND CITY</td>
</tr>
<tr>
<td>OR</td>
<td>STATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. f</td>
<td>on campus</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Some colleges are denying admission and perhaps reducing financial aid to students based on a single, non-financial, non-academic question that students submit to the federal government on their applications for student aid.”

Smart – or – sleazy?

NB. This issue has been resolved and is only being used to illustrate a point.

See: Using FAFSA Against Students
Scenario: Advances in Learning

– or –

The Sanctity of the Classroom

Harvard secretly photographed students to study attendance
November 6, 2014 | The Chronicle of Higher Education

See: Harvard secretly photographed students to study attendance
chronicle.com/article/Harvard-Researchers-Used/149865/

See: The Sanctity of the Classroom, Continued
harvardmagazine.com/2014/12/more-disclosures-of-surveillance-in-harvard-classrooms
Detour: So Much Data, So Little Time...

cf. transparency, purpose specification, and use limitation

– precisely to avoid misalignment between expectation and reality
Still shot from a classroom project used with permission from Professor Leah Lievrouw, Information Studies, UCLA. Classroom privilege and student confidentiality apply.
A ‘City within a City’

A map of a generic campus, showing many of the functions and venues found in a city: libraries, dental clinic, police, hospital, fire department, rec centers, hotel, restaurants, private club, elementary school.
Consider relationships we have with many students:

- Attorney – Client
- Employer – Employee
- ISP – Customer
- Landlord – Tenant
- Physician – Patient
- Team – Athlete
Data Exhaust

• Campus single signon use, IP address logs, campus wi-fi location data
• BruinCard: purchases, res hall entry, meal swipes
• Dean of Students/misconduct and campus police records
• Library circulation records (privacy-friendly)
• Campus-controlled Gmail, Google Drive, Box accts
• Snaps, vines, tweets, posts, yaks
Student Data We Have

- Age, gender, ethnicity, height and weight (athletes)
- Health status, psychological counseling, disabilities disclosures
- SSN, drivers license numbers
- Credit card info
- Student visa info
- Financial aid
- Parental finances
- Academic records
Scenario: Online Learning

University of Phoenix logo

coursera logo

udacity logo

thiel fellowship logo

2U logo

edX logo

khan academy logo

Minerva Schools at KGI logo
Helps students succeed

– or –

Creepy 1984 profiling?

See: Big Data on Campus
nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html?_r=0
Already, he said, “we have shown some pretty interesting results in being able to detect different [emotional] states from keystroke data.”
Student Learning Analytics

• Who owns these data? The algorithms?
• Who benefits? Why?

‘Facebook and You’ cartoon
Key Insights

• May be highly constrained by law and policy even as there is significant leeway for safety, evaluative initiatives, and more.

• The model we use as a foundation for analysis – the IRB – is often not involved in these situations.

• These data are highly desirable to many, even if their value is not always directly quantifiable.
Timing: 3 Complications
#1 Neo-liberal Pressures

**Corporation Neo-Liberalism Created The 3rd World**

[Sign held by a person protesting against neo-liberal policies.]
#2 Big Data Envy

BIG DATA
#3 Online Ed Envy

WHAT'S UR DIGITAL ID?
Ethics
Ethical Challenges

1. There’s no clear set of ideals or policies for preventing big data creep from going too far.
2. Insufficient transparency prevents students from having a clear sense of all the ways they are being observed, profiled, and analyzed.
3. Insufficient pathways exist for students to correct distorted or incorrect data, should they happen to discover that it exists.
4. Manipulative techniques are being used on students that threaten their autonomy, raise questions concerning whether universities are appropriately committed to respecting personhood and fostering mature character development, and suggest that student interests might be inappropriately circumvented.

5. Information about students is being gathered and shared that they might wish to keep private, or at least more obscure.
Ethical Challenges

6. Disciplinary systems are being created that can threaten faculty freedom and lead them to adopt big data practices they have reservations about.

7. The ideology of “scientism” is being promoted. It can help fallacious views about information be perceived as credible.
With big data invading campus, universities risk unfairly profiling their students

Obama's proposed Student Digital Privacy Act aims to limit what schools can do with data collected from apps used in K-12 classrooms. But college students are just as vulnerable to privacy violations.

By Evan Selinger, Contributor | JANUARY 13, 2015

See: With big data invading campus, universities risk unfairly profiling their students
csmonitor.com/World/Passcode/Passcode-Voices/2015/0113/With-big-data-invading-campus-universities-risk-unfairly-profiling-their-students
With big data invading campus, universities risk unfairly profiling their students

Obama's proposed Student Digital Privacy Act aims to limit what schools can do with data collected from apps used in K-12 classrooms. But college students are just as vulnerable to privacy violations.

By Evan Selinger, Contributor  |  JANUARY 13, 2015

See: *With big data invading campus, universities risk unfairly profiling their students*

csmonitor.com/World/Passcode/Passcode-Voices/2015/0113/With-big-data-invading-campus-universities-risk-unfairly-profiling-their-students
See: Your College May Be Banking on Your Facebook Likes
nytimes.com/2015/01/25/technology/your-college-may-be-banking-on-your-facebook-likes.html
See: Obama to Call for Laws Covering Data Hacking and Student Privacy

nytimes.com/2015/01/12/us/politics/obama-to-call-for-laws-covering-data-hacking-and-student-privacy.html?_r=0
UCLA Work

Early Results on (Big) Data Governance

1. A framework for putting this talk in context
2. Principles and insights
3. Leveraging existing pieces
UCLA: $5B, 42K students

Mission:
- Education
  - 112K freshman applicants for AY2015-16
  - "customers"
- Research
  - RU/VH
  - Academic freedom
  - Shared governance
- Service
  - "teach for California, research for the world"
  - "FOIA"
“We keep secrets because we get to do our work free from judgment — until we’re ready to share it,” he wrote. “We keep secrets because keeping secrets gives you space to change your mind until you’re really sure you’re right.”

— Snapchat CEO Evan Spiegel
breach protocol
data protection policies
cloud contracts
awareness initiatives
cyber insurance

Institutional positions
© police
data classification
cloud is here to stay?

volume
velocity
variety
variability
veracity
big data privacy hazards

- indiscriminate collection of data
- data generation
- predictive analysis
- indefinite storage
- data breaches

- volunteered data
- observed data
- inferred data

- descriptive – summarize what happened
- predictive – forecast what may happen in the future
- prescriptive – recommend one or more courses of action

Source: Based on materials from Doron Rotman, KPMG
assumptions
• security and privacy (as part of ethics) are a basic given
• appropriate use and ethics are the next frontiers
• governance will need to frame a data ethics model (a la IRB??) for big data
“Give us/we have your data. We’ll analyze it for you and give you the insights you want.”
Institutional values
Value of the data

= E ?

iso _________ cpo _________ cdo
Task Force Charge

• A consistent set of expectations/principles for the campus and the campus community about ethical and appropriate use of such data.

• A mechanism for big data governance to assure these expectations are met.

See: UCLA Data Governance Task Force
https://ccle.ucla.edu/course/view/datagov

• (We have the chair of the UCLA Medical Center’s Ethics Board as a Task Force Member to help)
FIPs is Not Enough

IV. The Fit with Fair Information Practices

V. Revisiting FIPS in the Ethical Use of Analytics

See: Data Protection Law and the Ethical Use of Analytics
huntonfiles.com/files/webupload/CIPL_Ethical_Underpinnings_of_Analytics_Paper.pdf
A Gathering of Principles

- FIPPs/OECD Privacy Framework (2013 revision)
  [oecd.org/sti/ieconomy/oecd_privacy_framework.pdf](oecd.org/sti/ieconomy/oecd_privacy_framework.pdf) PART I, Chapter 1, Parts One and Two (pp 13-15)

- The Belmont Report of 1979: Ethical Principles and Guidelines for the Protection of Human Subjects of Research
  [hhs.gov/ohrp/humansubjects/guidance/belmont.html](hhs.gov/ohrp/humansubjects/guidance/belmont.html)

- Big data analytics
  - Asilomar Convention for Learning Research in Higher Education (2014) [asilomar-highered.info](asilomar-highered.info)
  - Australian Government's draft Responsible Data Analytics paper (2014)
See: Code of practice for learning analytics: A literature review of the ethical and legal issues
repository.jisc.ac.uk/5661/1/Learning_Analytics_A_-_Literature_Review.pdf
Copyright © Jisc November 2014, licensed under Creative Commons Attribution 4.0.
Leveraging Existing Policy

• UC Statement of Ethical Values and Standards of Ethical Conduct
• UC Statement of Privacy Values and Privacy Principles
• UC Diversity Statement
• UC Principles of Community
• Faculty Code of Conduct
• Student Conduct Code
• Academic freedom and shared governance
information security protects all information

IT security protects technical infrastructure

information

individuals
(e.g., web sites visited, research being conducted and related data)

information about individuals
(e.g., student/patient records; SSNs)

confidential information
(e.g., intellectual property, security info)

autonomy privacy
ability of individuals to conduct activities without observation

information privacy
protects information about individuals

See: University of California Privacy and Information Security Report
ucop.edu/privacy-initiative
Leveraging Existing Organizational Structure

- board on privacy and data protection
- chief privacy officer
- privacy framework
- privacy program
- breach protocol
- policy interpretation
- data protection initiatives
- community awareness
- statement of privacy values
- privacy principles
- (shared) privacy governance
- strategic direction setting
- privacy balancing process
Final Takeaways

• Even without volume, velocity, variety, variability, and veracity raise the ethical complexity of big data.

• The sharing of data between stakeholders layers even greater complexity and is roughly proportional to the difference in missions, values, and obligations of the two organizations.
Q & A

Evan Selinger
eselinger@gmail.com | @EvanSelinger

Kent Wada
kent@ucla.edu | @KentWada