# Ethical Use of Data Beyond Privacy

**Facilitators:**

**Julie Carpenter-Hubin**, Assistant Vice President, Office of Academic Affairs, Office of Institutional Research & Planning, The Ohio State University

**Don Hubin**, Director, Center for Ethics and Human Values, The Ohio State University

# Internal Data Sources

| | |
|---|---|
| Pre-application and pre-enrollment browsing | Campus web browsing |
| Demographic data from applications | Healthcare records |
| SAT/ACT scores, HS GPA, etc. from applications | Living situation and financial situation |
| Degree program and declared major | Activity records: campus dining choices; gym use; student club memberships; etc. |
| Overall performance in previous courses | Student surveys and other quantitative data like focus groups |
| Performance within courses from LMS | Alerts from front line faculty and staff |
| Disciplinary records | Email content and metadata |
| Library records and e-reader records | Student movement based on wireless connection |

# External Data Sources

- National Student Clearinghouse
- Academic Analytics (for faculty)
- U.S. Census Data
- U.S. Department of Labor
- Social Media Data

# Potential Benefits

- For Students
- For Institutions (beyond those for students)
- For Researchers
- For Vendors
- For Society

# Potential Risks

- Privacy Violations
- Threats to Autonomy
- Algorithmic Bias
- Profiling
- Data Ownership Issues
- Threats to Faculty/Instructor Intellectual Freedom

# Current & Near-Future Uses of Predictive Analytics

- Current Uses
    - Enrollment Management
    - Early Intervention Systems
    - Course of Study Recommender Systems
    - Learning Analytics
    - Alignment of Course Offerings with Student Demand
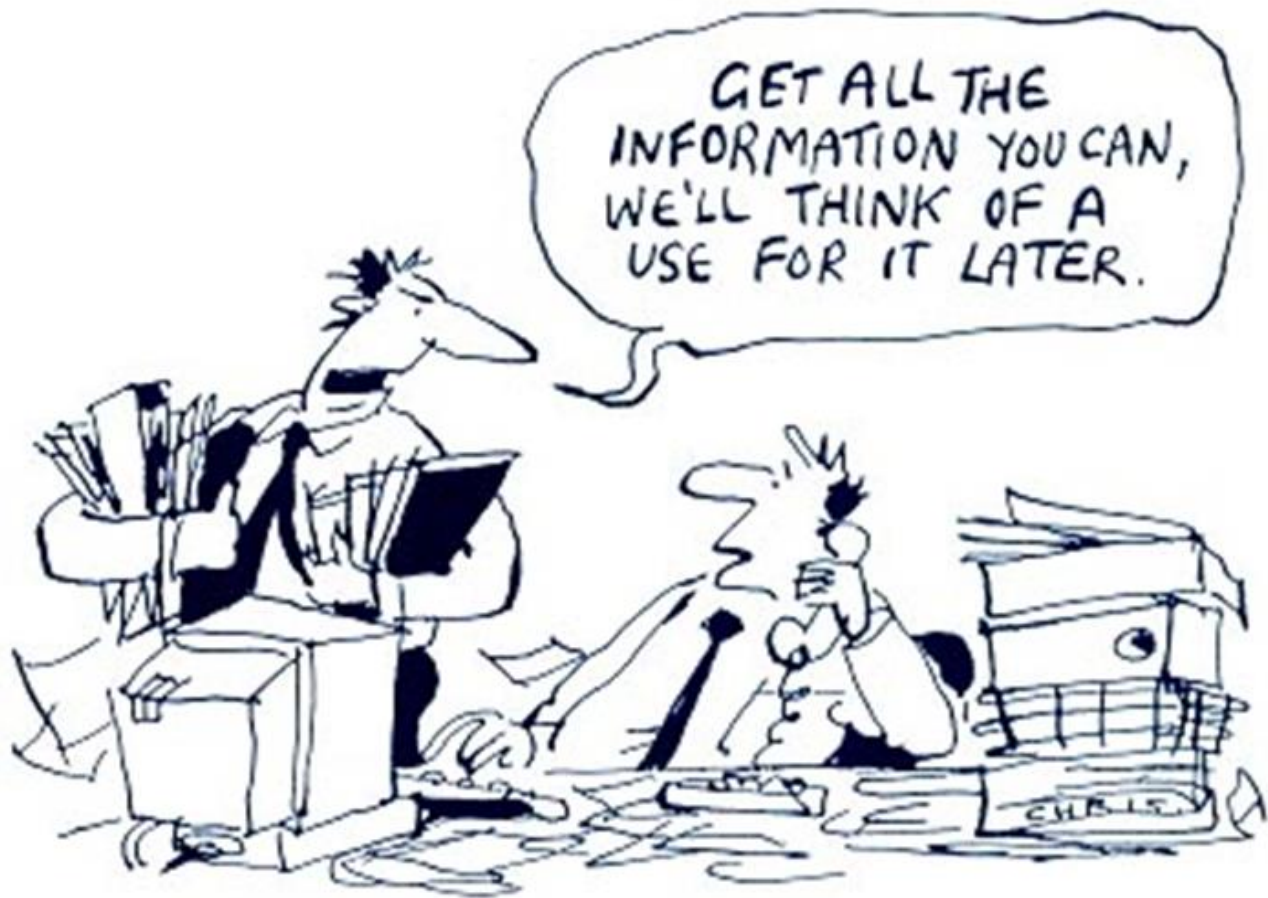- Near-Future Uses

# Fair Information Practice Principles (FIPPs) Approach to Privacy Protection

- **Collection Limitation:** There should be limits imposed by lawful and fair means.
- **Purpose Specification:** The purposes for which personal data are collected should be specified no later than at the time of data collection and subsequent use limited to the fulfillment of these purposes …
- **Use Limitation:** Personal data should not be disclosed or otherwise used for purposes other than those specified except with the consent of the data subject, or by the authority of law.

# Purpose Specification

"The big data business model is antithetical to data minimization. It incentivizes collection of more data for longer periods of time. It is aimed precisely at those unanticipated secondary uses, the "crown jewels" of big data. After all, who could have anticipated that Bing search queries would be used to unearth harmful drug interactions?"

Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property*, 11, no.5 (2013): 259-260

# Use Limitation & Consent

"Today, the widespread and perpetual collection and storage of personal data have become practically inevitable. … Indeed, there is hardly any part of one's life that does not emit some sort of "data exhaust" as a byproduct. And it has become virtually impossible for someone to know exactly how much of his data is out there or where it is stored."

Craig Mundie, "Privacy Pragmatism: Focus on Data Use, Not Data Collection," *Foreign Affairs*, March/April 2014

# Data Stewardship

## Commercial/Transactional Model

## vs.

## Fiduciary Model

# Dimensions of the Ethics of Data Analytics

***Ethics of Data:*** generation, recording, curation, processing, dissemination, sharing, use

***Ethics of Algorithms:*** artificial intelligence, machine learning, robots/chatbots

***Ethics of Practices:*** responsible innovation, programming, hacking, professional codes, consent, privacy

# Scenario One

The University of Erewhon is adopting a new predictive analytics system to support advising for undergraduate students. Based on algorithms developed from several prior years of student data, the system calculates, based on each students' high school record, standardized test scores, and college coursework to date, how likely a student is to be successful in a particular course.

The University is considering several plans about how to use the results of these analyses to best support student success. The following proposals are under consideration:

# Scenario One (cont'd.)

The University is considering several plans about how to use the results of these analyses to best support student success. The following proposals are under consideration:

- Provide academic advisors an assessment of every student's likelihood of success in all classes and major programs they are enrolled in or considering enrolling in.
- Notify the academic advisors of a student's likelihood of success in a course or major program only when students are at risk of failure, allowing advisors to proactively address the issues.
- Notify both academic advisors and instructors of students who are at risk for failure in a given course to allow them to provide extra resources to help those students succeed.

What advice would you provide to the University of Erewhon as they move to adopt this system to support student success?

# Scenario One (cont'd.)

The University is considering several plans about how to use the results of these analyses to best support student success. The following proposals are under consideration:

- Provide all students with access to information about their likelihood of success in any courses and majors they are considering, recommending that students discuss this information with their advisors.
- Notify students of their likelihood of success in a course or major program only when they are at risk of failure, recommending that students discuss this information with their advisors.
- Making information about a student's likely success or failure in a course or major program available to a student only upon meeting with an academic advisor.

What advice would you provide to the University of Erewhon as they move to adopt

# Scenario Two

Suicide is a leading cause of death among college and university students in the United States (https://www.sprc.org/settings/colleges-universities). A student survey administered on your campus with the promise of confidentiality asks questions about connectedness to other students, staff, and faculty on campus; hours per day spent sleeping; and levels of alcohol and drug use. Some combinations of responses might indicate that a student could be at risk of suicide. Should you share the responses from individuals with such combinations with mental health professionals on campus? Does that violate the promise of confidentiality?

# Scenario Two (cont'd.)

Suppose this information comes from sources other than surveys. Students' ID card swipes might tell you that a particular student is in his dorm room most of the time he is not in class and that he is spending that time alone. When he does leave the dorm, it's often late at night and he goes off campus for a short period of time. His grades have dropped, but are still in the B- to C range. This behavior is very different from his behavior in the previous semester. Is it your responsibility to share these data with a counselor? Does that violate his right to privacy?

# Scenario Three

Professor Xavier has been teaching calculus at Graymalkin University for 20 years, with reasonable success with students. Results from analysis of the learning management system indicate that a different approach using a different textbook is likely to be more successful with women and/or minority students. However, Professor Xavier thinks the alternative approach/textbook is inferior. Professor Xavier has had a higher percentage of his students go on to be successful in graduate school. Graymalkin University has set increasing the number of women and minorities in math and physical sciences as a strategic goal. How would you advise administrators at the University to address Professor Xavier's choice of how to teach the calculus course? Would it be appropriate for the University to use the results of this analysis to assign Professor Xavier to teach different courses instead of the calculus course?

# Scenario Four

The Board of Trustees for the University of South Nooz has challenged the institution's president to improve its graduate school rankings. To select their incoming graduate class, U.S. Nooz has long relied on a short essay discussing the applicant's interest in the relevant discipline and long term goals, along with GRE scores and GPA data.  The new dean of the graduate school has suggested eliminating the essays, citing the time it takes reviewers to read each of them and the potential bias in such a nonstandardized measure.  Instead, the dean suggests that graduate student success can more likely be predicted by GRE scores, overall GPA, GPA in the major, number of years to finish the bachelor's degree, and zip code of the undergraduate institution. Several data scientists develop a sophisticated algorithm for use in admissions and your institution is now involved in discussions of whether to institute the sort of program the graduate dean has proposed.  What questions would you raise during the discussions?  What cautions would you share with the data science team?  With the graduate school dean?

# Scenario Five

A for-profit company, National Success Associates (NSA), is offering colleges and universities a high-quality learning management system, *Student Success*, at a substantially lower price than competing programs from Blackboard and Desire2Learn. The agreement provides that NSA will do detailed analysis using advanced data-analytic methods on student progress and success, sharing the results openly with institutions. NSA will also provide each institution that adopts *Student Success* analyses comparing performance of students at that institution with performance of students at other institutions using the program, with the identities of the other institutions concealed. In order to facilitate the delivery of these services, NSA's contract with institutions provides that NSA is the owner of the data collected by the learning management system. The Dean of Undergraduate Education at your college is considering entering into this sort of contract with NSA and has asked for your advice. What should you tell her?

# Scenario Six

Dr. Messen, Vice President for Research at Aspire University, is rolling out a plan to improve the research productivity and profile of the University. He proposes that A.U. use the extensive data it holds on faculty, including publication records, impact measures, grants, awards, membership in distinguished academies, and so forth, to predict the career trajectories of younger faculty. This would be done using analytic tools developed by Research Impact, Inc., a private company that has developed advanced algorithms to predict research success. Dr. Messen proposes that A.U. redeploy its resources to foster the research of the most promising young faculty, providing them with internal grants, course releases, and extensive leave time to support their research. How would you advise Dr. Messen?