

WHOSE DATA IS IT ANYWAY?
EXPLORING THE COMPLEXITIES OF DATA OWNERSHIP, ACCESS, AND UTILIZATION IN AN ERA OF BIG DATA
SACSCOC 2016
DR. JULIANA LANCASTER
DR. LINDA GILBERT



WELCOME AND INTROS

- Collaborative effort across campus
 - Juliana Lancaster – Office of Plans, Policies, and Assessment
 - Linda Gilbert – Educational Technology
 - Laura Ledford – Enrollment Management
- Members of "Data Stewards Team"
- Georgia Gwinnett College
 - Public, baccalaureate granting college
 - Opened for classes in 2006 with 118 students, 11 faculty, and 3 majors
 - Currently have 12,000 students, over 500 faculty, and 15 majors

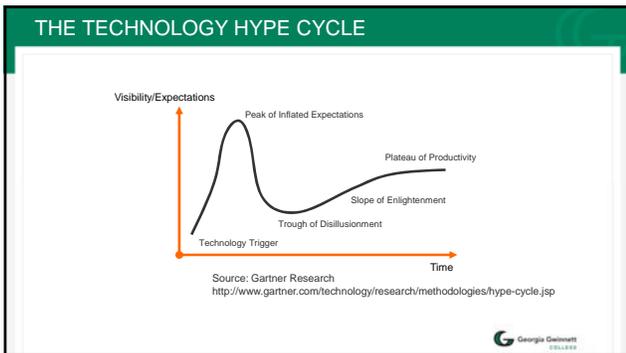
• How MANY different data elements did you list during the opening exercise?
• What types of data do we collect? From what sources?



OVERVIEW

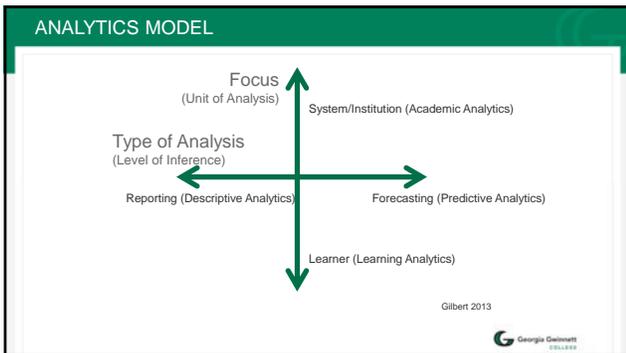
- Background and context
- Related research and practice
- Concerns and issues
- A few ethical questions
- Best practices and recommendations

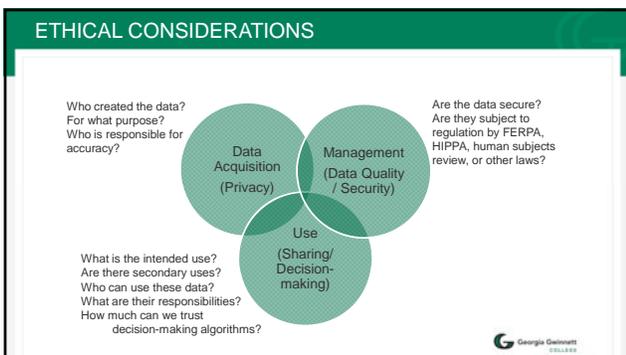




- ### BACKGROUND / CONTEXT
- Increasing pressures toward data use in education
 - Mirroring “big data” in business
 - Learning Analytics – personalization
 - Institutional Analytics - decision-making at multiple levels
 - Technology/capability outstripping understanding of implications
 - Legalities lagging practice
 - Business “transactional” model getting increasing push-back
 - Most discussion about data use is “binary” (right/wrong).
 - More exploratory: What does this mean for (education/society, etc.)?
- Georgia Gwinnett College

- ### KEY DEFINITIONS
- Big Data
 - Used to indicate huge quantities of data
 - 3V's: Volume, Variety, Velocity
 - Also used (informally) to reflect data that is highly connected across systems
 - Not so much “more” data, but more accessible/usable for analytics.
 - Ethics
 - “In the digital context, we define ethics as the systematization of correct and incorrect behavior in virtual spaces according to all stakeholders.” Pardo & Siemens 2014
- Georgia Gwinnett College





- ### PREDICTION: CONSIDERATIONS AND CONCERNS
- Messy Data
 - Availability
 - Data choices
 - WYSIATI
 - "Map, not territory"
 - Metric issues
 - Human Biases
 - Limits to Prediction
 - "Black Swan" events (low probability/high impact)
 - Humans = Autonomous Agents
 - Feedback Loops / Self-fulfilling prophecies
- "The problem starts when smart people in nice suits and lab jackets proclaim that "the data says..." In truth, the data never says anything. We interpret it in one way or another and there are lots of ways to interpret it incorrectly." - Greg Satell

<http://www.digitaltoronto.com/2011/the-pitfalls-of-prediction/>
- Georgia Gwinnett College

PREDICTION: CONSIDERATIONS AND CONCERNS

- Metric issues (beyond poor choice and mis-interpretation)
 - "...once an indicator is chosen as reference, behaviours affecting the measured phenomenon may change as a result."
 - OECD, (2005) *Statistics, Knowledge, and Policy: Key Indicators to Inform Decision-Making*, p.404.
 - "KPIs are powerful tools...However, if the KPIs become the goals, then they turn into toxic material that will inhibit performance improvement."
 - Marr, 2014 <https://www.linkedin.com/pulse/20140326073422-64875646-caution-when-kpis-turn-to-poison>
- Human Biases → Algorithm Biases
 - Facebook / Twitter examples
- Potential "engines of inequality"
 - "Networked systems tend to create ...asymmetric accumulations. That's an important regulatory challenge."
 - <http://www.chronicle.com/article/How-Colleges-Should-Adapt-in-a/237842>

Georgia Gwinnett COLLEGE

HEADLINES

The image shows a collage of news headlines. At the top is 'the guardian' with a headline about Facebook's algorithm. Below it is 'Intel' with a headline about Facebook apologising for a fake article. To the right is 'THE WRAP' with a headline about Microsoft's 'Artificial Intelligence' Twitter Bot Tay being shut down. The Georgia Gwinnett College logo is in the bottom right corner.

PRIVACY CONSIDERATIONS AND CONCERNS

- Monitoring/Surveillance
 - Monitoring is a general term describing collection of data; surveillance "more narrowly refers to a relationship between a control agent and those being monitored." Knox 2010
- Privacy
 - "While it has numerous definitions, in this document, privacy is defined as the regulation of how personal digital information is being observed by the self or distributed to other observers." Pardo & Siemens 2014

Georgia Gwinnett COLLEGE

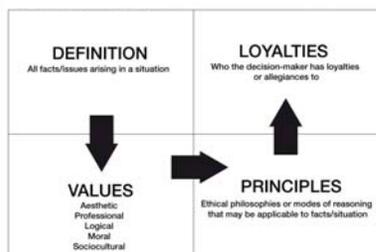
HIGHER EDUCATION CONCERNS AND ISSUES

- Data Governance often not clear
 - Are people managing/using data aware of their responsibilities?
- Proprietary black boxes
- “Ownership” – raw data, derived data
 - Your data, their algorithm – who owns it?
 - Plagiarism software
- Computerizing biases



Georgia Gwinnett COLLEGE

“POTTER BOX”



Georgia Gwinnett COLLEGE

A FEW ETHICAL QUESTIONS

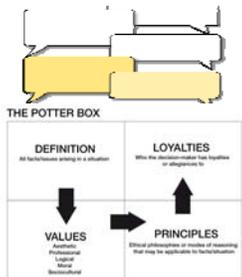
- To what extent should learning/institutional analytics continue to take cues from business analytics?
- If we “know something”, what is our obligation to “do something”?
- What are the rules about data access? What about copyright? Legal access to data?
- Who decides what feedback to students is valid and how often it should be delivered?
- What technical infrastructure exists to handle something unexpected in the data?



Georgia Gwinnett COLLEGE

PRACTICAL EXAMPLES!

- In groups, review one of the scenarios, basing your discussion on the Potter Box model.
- Be prepared to report what you would do and why.



SCENARIO 1

- Board of Regents requires that faculty and staff information be entered into the person file in Banner just like student information. There is a conflict with having faculty and staff information stored in the same place as student information.
 - Faculty and staff *have* to have their information in Banner in order to allow access to Banner web, as well as identifying faculty who teach classes at the institution.
 - Some staff can then see ALL personnel information – including faculty/staff, not just student. This is considered inappropriate by HR.
 - Also, there are (some) student workers who can see student info (background checked, signed confidentiality agreements). And even some who can see faculty/staff info, depending on role.
- How does an institution deal with the conflicting data needs for those two groups?



SCENARIO 2

- A researcher wants to explore the level of discourse of students when commenting issues posed as part of an online course. Students express their views in a closed platform.
- May the data be collected and used for publication?
- Is user consent needed to perform the analysis?
- Should other users be made aware of information related to the overall discussion?
- Can individuals be identified based on the derived data?



BEST PRACTICES / RECOMMENDATIONS

- Transparency.
- Student control over data use
- Security/clearly defined access
- Accountability and Assessment



BEST PRACTICES / RECOMMENDATIONS

- Transparency. Communicate...
 - What information are you collecting about students
 - Why are you collecting this information
 - How the information is protected
 - What information is shared with third parties and, if so, with whom and for what purpose(s)
 - Who should be contacted with questions
- Student control over data use
 - Terms of use of student data
 - Who owns data; under what circumstances?
 - When is consent for use required?
 - What process exists for students to review and request corrections?
 - “Portable legal consent”



BEST PRACTICES / RECOMMENDATIONS

- Security/clearly defined access
 - Maintain awareness of all applicable law
 - Maintain a current inventory of all third-party educational services being used
 - Require written contracts that address privacy and data security
 - Recognize trade-offs between access and benefit
- Accountability and Assessment
 - Define responsibilities
 - Governance / stewardship
 - Track changing legal and socially-acceptable requirements
 - BALANCE privacy and usefulness



INTERESTING DIRECTIONS

- Educational initiatives at institutions
 - Including ethics in classes for students in computer science and related fields.
 - Lecture series on data ethics for faculty and staff
 - “Data clinic” attached to statistics consulting unit
 - At some institutions, Library becoming a hub for “research data management”
- Cross-institution attention
 - Summer 2016: Ithaka and Stanford “Responsible use of student data” <http://ru.Stanford.edu>
 - Model policies under development

Bottom line: As we move into increased use of analytics, we need to be thoughtful about how it impacts education.

THE CHRONICLE of HIGHER EDUCATION



CITED REFERENCES AND CREDITS

- Eppling, M., Timmons, S. and Wharrad, H. (2003) An educational panopticon? New technology, nurse education, and surveillance. *Nurse Education Today* (23), 412-415.
- Gilbert, L. (2013). Analytics: The Hope and the Hype. Part of a group presentation on *Moving into Analytics: Practical Steps* by Gilbert, L., Nagel, K. and Ledford, L. at the Educause Southeast Conference, Atlanta, 5/30/2013.
- Gilbert, L. and Lancaster, J. (2014). *The Road to Analytics: Laying the Foundation*. Pre-conference workshop at the Educause Annual Conference, Orlando, 9/29/2014.
- Jones, K. M.L., Thomson, J., and Arnold, K. (2014). *Questions of Data Ownership on Campus*. *Educause Review Online*. <http://er.educause.edu/articles/2014/8/questions-of-data-ownership-on-campus>
- Knox, D. (2010). A Good Horse Runs at the Shadow of the Whip: Surveillance and Organizational Trust in Online Learning Environments. *Canadian Journal of Media Studies* (7), June 2010.
- Pardo, A. and Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45(3), 438-450
- Slade, S. and Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57(10) pp.1509–1528.
- The Aspen Institute. (2010). *The Promise and Peril of Big Data*.
- Federal Privacy Technical Assistance Center (PTAC) <http://ptac.ed.gov/>
 - Note: For P-12, but useful.



QUESTIONS, COMMENTS, DISCUSSION

Juliana Lancaster jlancaster@ggc.edu
 Linda Gilbert lgilbert@ggc.edu