Dialogue on Privacy, Security, Ethics

In Preparing for this Panel...

- Folks from very different areas of research and practice
- All interested in the issues of privacy, security, ethics, research
- Disciplinary differences have emerged in our discussions
- We feel this debate is healthy and necessary

Elizabeth's World

- Coming from Philosophy, Applied Ethics, and Research Ethics: Hoping to prioritize the role of ethics in our considerations (Meta Considerations, such as research methods and ethics; IRBs' stake in big data discourse; specific issues, such as privacy, anonymity, de-re-identification science);

    - Trying to make connections between the language of privacy harms and risks (from research regulation) with emotive harms cause by "inadvertent algorithmic cruelties," and "spurious correlations" vis-à-vis big data
    - Today, thinking about the epistemology of big data and research and our "aspirational visions"
    - What is our common goal? Is ethics one of them?
    - Predictive analytics, targeting, profiling: Are we ready to embrace these for a social good, and who is/is not on board? Is this TOO deterministic?
    - What is not quantifiable? When did we accept the quantification of self, experience, as the only or the best, realities?
    - Digital diploma mills ideology (Nobel, 1990s)
    - When we look at nations, eg, Finland, how is the educational system using BD and analytics?
    - Differential systems in access to data?
    - Longitudinal look at return of results to students
    - Work versus education; education versus training (Wisconsin Idea demise)
    - How transparent are the analytics?

The Original Plan

- Privacy is a right
- Privacy is contextual
- Privacy is subjective
  - The quality or condition of being secluded from the presence or view of others
  - The state of being free from unsanctioned intrusion
  - The state of being concealed

- Natural and normative privacy protections
- Control/Restriction; contextual integrity theories (Moor, Nissenbaum)
- Instrumental versus intrinsic value

- Growing concerns around the ethics of use of core and derived data;
- Considerations of public data
- Balancing data sharing with personal information;
- What are students consenting to? Students as learners, research participants, customers, members of groups

There are very few Moby-Dicks. Most of us are sardines. The individual sardine may believe that the encircling net is trying to catch it. It is not. It is trying to catch the whole shoal. It is therefore the shoal that needs to be protected, if the sardine is to be saved. An ethics addressing each of us as if we were all special Moby-Dicks may be flattering and it is not mistaken, but needs to be upgraded urgently. Sometimes the only way to protect the individual is to protect the group to which the individual belongs. Preferably before any disaster happens.

multiple disciplines. As set out by **Kuhn (1962)**, a paradigm constitutes an accepted way of

interrogating the world and synthesizing knowledge common to a substantial proportion of
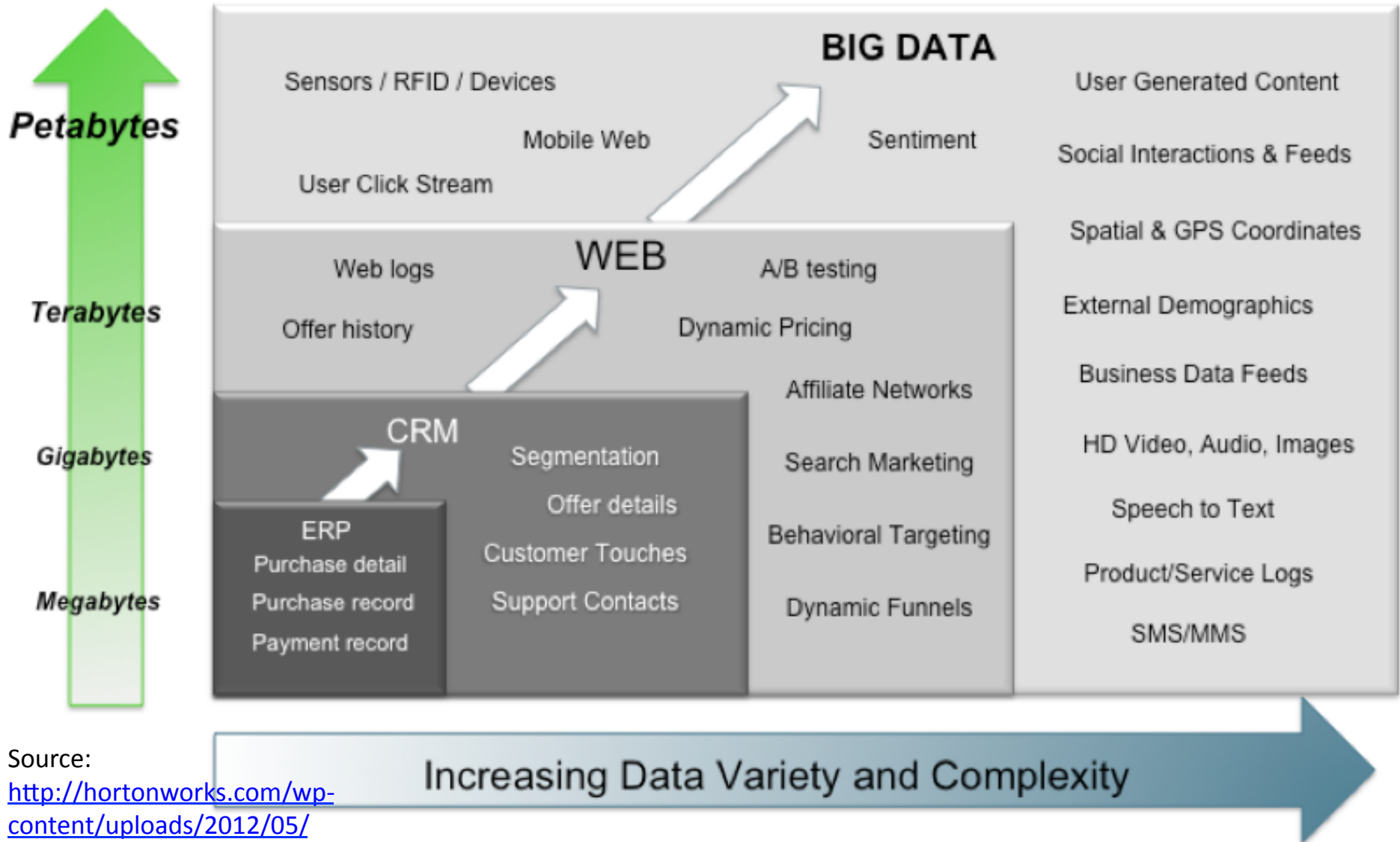
Abstract

Paradigm	Nature	Form	When
First	Experimental science	Empiricism; describing natural phenomena	pre-Renaissance
Second	Theoretical science	Modelling and generalization	pre-computers
Third	Computational science	Simulation of complex phenomena	pre-Big Data
Fourth	Exploratory science	Data-intensive; statistical exploration and data mining	Now

Compiled from Hey et al. (2009).

Think about where we are in our teaching and education. What are we teaching about the ethics of data science

Big Data = Transactions + Interactions + Observations

BIG DATA

Sensors / RFID / Devices

Mobile Web

User Click Stream

Sentiment

User Generated Content

Social Interactions & Feeds

Spatial & GPS Coordinates

WEB

Web logs

Offer history

A/B testing

Dynamic Pricing

External Demographics

Business Data Feeds

CRM

Segmentation

Affiliate Networks

Search Marketing

HD Video, Audio, Images

ERP
Purchase detail
Purchase record
Payment record

Offer details

Customer Touches

Support Contacts

Behavioral Targeting

Dynamic Funnels

Speech to Text

Product/Service Logs

SMS/MMS

Petabytes

Terabytes

Gigabytes

Megabytes

Increasing Data Variety and Complexity

Source: Contents of above graphic created in partnership with Teradata, Inc.

Ethics, Privacy, BD

"Algorithms are worthless (and also valueless) without data to run on. Tapping into personal info is actually foundational to the connection to privacy in the ethics/rights context. The whole enterprise of predictive analytics is predicated on tapping into personal info, fitting people into categories, and feeding that back to users in various ways. So, we can think of data as a "fuel source" for algorithms AND we can think of it as inextricably tied to one's private zone, one's personal identify, space, etc.

Privacy is something that can be compromised when information appropriate

to one context is distributed in another; consent and privacy are

intermixed here as well since it's hard to say that someone who turned

over their info after absent-mindedly agreeing to some EULA or ToS

actually consented to their info being used in certain ways. There can be

a contextual conflict here, I think - and that is also relevant for

privacy, since privacy is also important for allowing us to navigate different settings and understanding the social world around us."

Consider the Ethics of:

- Algorithms providing course selection advice, algorithm stereotyping via student clustering,
- Algorithms providing encouragement/discouragement, the cultural norms of advice delivery and whether they impact the student (Una-May)
- The incompleteness or incorrectness of data that might be used to feed algorithms (Ari)
- What potential harms do researchers have to think about / design / disclose / account for when using algorithmic (specifically: machine-learning) approaches to educational data? (Ari)
- Reconsideration of risks and harms
    - Secondary subjects
    - Research bystanders
    - Collateral subjects
    - Group and individual emotive harms

New Dynamics, New Norms?

Medicine Today

Reactive, population-based, one-size-fits-all model of care

Personalized Medicine

Predictive, preventive, patient-centric model of care