From Private Data to Public Commons: Open Access/Open Data and the Public Good

Jenifer Sunrise Winter
Associate Professor
School of Communications
Co-Director, Pacific ICTD Collaborative
University of Hawai‘i at Mānoa
jwinter@hawaii.edu / @jenifersunrise
Overview

• The emerging IoT datasphere
• Big data repositories and the public good
• Big data repositories as public goods/Commons
• Data governance/privately held data sets of high societal value
  – Values-driven Data Governance project
The emerging IoT datasphere

• As a growing number of ordinary objects are redesigned to include digital sensors, computing power, and communication capabilities become part of the Internet, the **physical and virtual worlds are merging**.
  - As many as 75 billion “Things” may be connected to the Internet by 2020, with room for growth (Proffitt, 2013).

• Increased instrumentation, tracking, and measurement of both natural and social processes.

• Data about ‘trivial’ daily tasks federated and used to reveal associations or behaviors.
If You Have a Vizio TV, It’s Watching You Too

Ethan Wolff-Mann  @ewolffmann | Nov. 9, 2015

The default setting is on, but you can turn it off.

If you have a Vizio TV, it might be watching your every move—and selling those moves to advertisers so they can target you more efficiently by analyzing your behavior. According to a report by ProPublica, the unadvertised feature which appears deep in its 3,500-word privacy policy, called “Smart Interactivity,” is turned on by default.

Almost every new-fangled feature comes with an implicit but fairly obvious cost of privacy, but Vizio goes a step further. Unlike its competitors LG and Samsung, which have similar features, the Vizio version has an opt-out model with the sharing as the default. (The other manufacturers use an opt-in system.) Even more importantly, Vizio steps ever further by allowing the behavior analysis to be connected to individuals IP addresses for precise targeting. As ProPublica notes, data brokers like Experian (whose name you recognize from your credit report) frequently sell comprehensive
What are the sources of these data?

- Digital traces (Bottom-up)
- Govt. data (Top-down)
- Corporate data
- Natural world
- M2M data (Data begets more data)
IoT datasphere

- As much data will flow this year as in the entire history of the Internet since its creation in 1969.
- “Big Data” describes large, complex data sets that require novel data management tools.
- Allows modeling and predicting natural phenomena, as well as human behaviors in their rich contexts of relationships, groups, and social influence.
Governments release (or create) valuable data sets in order to build public-facing tools that allow citizens to better understand government processes via increased transparency.

Within the sciences, big data sets are probing new frontiers through exploratory analysis in astronomy and physics, environmental sciences.

Goo.gl/Turql  Goo.gl/vYZCnM
Medical data sets are being mined to create targeted genomic therapy, increase patient care, and survival rates.

Smart cities/smart grid initiatives focus on energy conservation, reduction of reliance on fossil fuels.
Open Data/Open Access?

• **Open government data** describes a growing movement to make certain data freely available for citizen use
  – Goal is to release (or create) valuable data sets in order to build public-facing tools that allow citizens to better understand government processes

• **Open research data** refers to making publicly funded research data available for reuse
  – Access to publicly funded data to advance science, confirm integrity of results
  – *What about other, privately-held data of high societal value?* (medical claims, energy use, etc.)
Data as a public good or Commons (Values-driven Data Governance)

• Should access to reasonably secured medical claims data (or other big data repositories of great social value) be provided without profit to all members of a society? (researchers, citizen scientists)

• How can we navigate multiple value claims about these data resources in order to lower the barrier between public and private goods?

• We are developing a research project (VDDG) that draws from theories of collective action and public goods to investigate big data marketplaces as Public Commons
Values-driven Data Governance

• UH TASI-Pacific Health Informatics and Data Center (PHIDC) project developing an “all claims data warehouse.”

• Act 139 (16) requires Medicare/Medicaid/EUTF claims data to be provided to the State Health Planning and Development Agency and to the PHIDC for management and analysis (repository).

• Intended to support policy research related to healthcare costs, quality, service gaps and disparities.
Values-driven Data Governance

• We argue that data governance/stewardship policies are needed to negotiate the multiple socioeconomic and political values arising from creation and use of these big data repositories.

• Big data repositories are in fact a public good, and should be governed as a Commons (that is, a resource affecting the entire community) rather than held privately (siload).
  – We all contribute to the Commons, often without even realizing it.
Values-driven Data Governance

• Goal: Theorize and conceptually model data governance of Big Data Commons

• Develop a conceptual model for values-driven data governance (from discourse analysis and text mining) related to healthcare claims data

• Develop/empirically demonstrate a methodology for applying the model to other big data domains.
Values-driven Data Governance

• Broader implications of the project include translation of the model as a tool for data governance and policy making into other domains:
  – may include: smart cities (governance, environment); education; other areas related to conservation/environment.
  – Lowering the barrier between public and private goods.

• *In the IoT datasphere, what other data repositories should be governed as a Commons?*