Ethical Issues of Big Data 2.0 Collaborations

Roles and Preparation of Information Specialists

Sheila Corrall
Kip Currier

UNIVERSITY OF PITTSBURGH
VERITAS 1787 VIRTUS

Ethical Issues of Big Data 2.0 Collaborations: Roles and Preparation of Information Specialists

Outline

- The data shift and evolving information services landscape
  - current and emergent roles for libraries and librarians
- Big data 2.0: Massive Open Online Data-driven Studies (MOODS) – examples of projects, identification of issues
- Developing info pros as data ethicists
  - roles: responsibilities and relationships
  - competencies: components and contexts
  - education: modules and methods
- Review and summary

Technology trends – **the data shift**

"…**technology** has enabled **data** to become the prevalent material and **currency** of research. **Data**, not information, not publications, is rapidly becoming the accepted deliverable of research"

Graham Pryor, DCC (2012)

Data is the new currency

- the new form of exchange in the business world
- \+ in the public sector where governments are key players

The evolving (library) data landscape

- Social science data archives and geospatial data resources
- Networked data-intensive science and research data services
- Linked open collections data for cultural heritage institutions
- Data analytics and visualization for research and assessment
- Creating local data infrastructures and metadata schemes
- Regional open data centers for municipal governments
- Digital privacy and data literacy training in public libraries
- Providing support for researchers using text and data mining
- Advising on IPR and ethics conflicts arising from Big Data

Libraries and librarians are already engaging with data as resources for research, public participation, and assessment

A measured approach: **The conscience of the big data world?**

"Currently most libraries seem to be (accidentally) providing a huge hoard of private user data to virtually anyone who wants it, but not actually using any of it themselves. If we are to credibly claim to be defenders of intellectual freedom and responsive to our communities, we need to use data more cleverly – and protect member privacy while we do so."

Hugh Rundle, information flâneur (2015)

"Should we moderate our traditional defense of privacy to enable data-driven processes…? Or should we play a more active role in defending privacy in a digital age?

Barbara Fister, Gustavus Adolphus College (2015)

"Learning analytics in the library and the emergence of professional ethics conflicts"

Big Data has arrive in college campuses and classrooms by way of learning analytics (LA) initiatives. LA measures, collects, analyzes, and reports on student behaviors in order to improve learning environments and outcomes. Now, academic libraries are considering their role in tracking and acting on analyzable flows of student data. However, LA comes with moral and ethical problems related to students' intellectual privacy and intellectual freedom. ... Libraries must therefore consider how LA initiatives contradict ethical principles set forth in the American Library Association's "Code of Ethics."

Kyle Matthew Lauer Jones, Indiana University &
Dorothea Salo, University of Wisconsin-Madison (Session 4.3 C)

Data roles of libraries and librarians

Current roles

- **Access management**
  - facilitating external dataset use
- **Data literacy**
  - educating students and faculty about managing research data
  - preparing graduates for working with data in employment
- **Collection building**
  - auditing and appraising data assets, cataloging and archiving datasets
- **Publishing support**
  - advising on identifying, citing, sharing, licensing and impact
- **Policy development**

Emergent roles

- **Data literacy**
  - preparing frontline staff to advise on digital privacy or data profiling
- **Metadata consultancy**
  - providing specialist expertise to support open data sharing
- **Infrastructure development**
  - serving as local platforms for open data in smaller communities
  - hosting data hackathons and facilitating data deposit
- **Data protection**
  - promoting responsible use of personal patron/student data

Big Data 2.0 – The next frontier

- Converges e-science with business intelligence, crowdsourcing, big data analytics, social media and Web 2.0 technologies
- Enables broader and deeper applications of analytical tools
- Takes data-driven research to new levels of technical and organizational complexity
- Located in academic/research institutions, but based on public participation

Global megaprojects

- Very large scale
- Interdisciplinary
- Human subjects
- Inter-state/international
- Multiple jurisdictions
- Cross-sector partners
- Different cultures

- Advancing knowledge to benefit society, but raising multiple issues of concern…

Research questions

RQ1 What ethical issues are raised by Big Data 2.0 initiatives?

RQ2 What best practices are emerging to inform responses to the challenges presented?

RQ3 What roles can information professionals play in promoting ethical practices in the big data environment?

RQ4 What competencies (knowledge, skills, and abilities) are required to perform the envisioned roles?

RQ5 How can LIS educators prepare new and experienced information professionals to perform as data ethicists?

Research strategy

Multiple-case study using existing documents as data sources

Pittsburgh Health Data Alliance

Carnegie Mellon University | University of Pittsburgh | UPMC

World-class CS/ machine learning + Medical research expertise + Deep data, clinical setting, commercialization

What roles can libraries and librarians play in such endeavors?

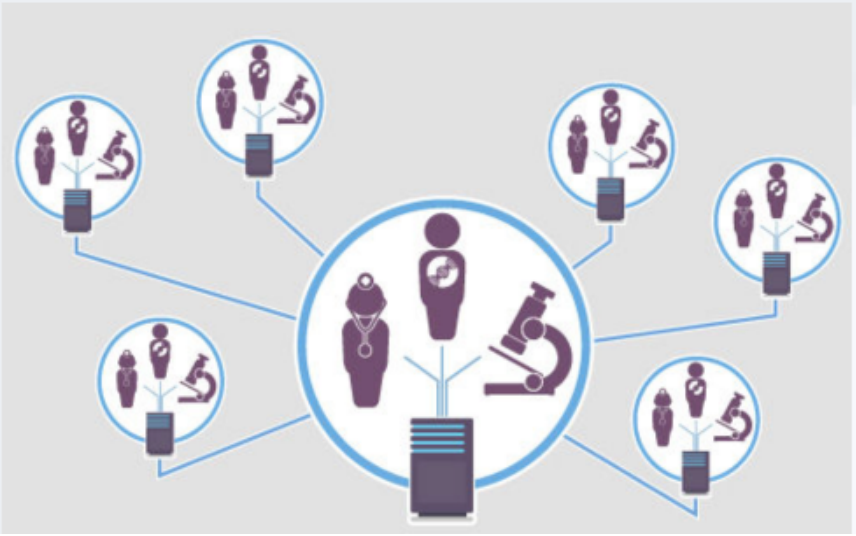
"The health care field generates an enormous amount of data every day. There is a need, and opportunity, to mine this data and provide it to the medical researchers and practitioners who can put it to work in real life, to benefit real people. Many organizations can fulfill part of this process, but none of them are equipped to begin with raw data, develop an idea and move that idea directly into a practice setting."

A NEW KIND OF RESEARCH ALLIANCE

The Oncology Research Information Exchange Network (ORIEN) is a unique research partnership among North America's top cancer centers that recognize collaboration and access to data are the keys to cancer discovery. Through ORIEN, founders Moffitt Cancer Center in Tampa and The Ohio State University Comprehensive Cancer Center – Arthur G. James Cancer Hospital and Richard J. Solove Research Institute in Columbus leverage multiple data sources and match patients to targeted treatments.

DATA SHARING TO GUIDE DISCOVERY

ORIEN partners utilize a common protocol: Total Cancer Care®. Established by Moffitt in 2006, Total Cancer Care provides a standard system for tracking patient molecular, clinical and epidemiological data and follows the patient throughout his or her lifetime. Partners have access to one of the world's largest clinically annotated cancer tissue repositories and data from more than 100,000 patients who have consented to the donation for research.



About **PGP**

Harvard PGP is "an open science research project…designed to create **public** scientific resources that everyone can access by bringing together genomic, environmental, and human trait data donated by our participants"

- Founded at Harvard Medical School in 2005, now a Global Network involving Canada (University of Toronto), the UK (UCL) and Austria (Austrian Academy of Sciences)
- Harvard PGP is staffed by a small, largely volunteer group of researchers, engineers, and ethicists who are all pioneers in their fields.
- Members of the Global Network follow a common set of guidelines, but the quantity and quality of information on national sites varies significantly

"Privacy, confidentiality and anonymity
are impossible to guarantee in a...research study
where public sharing of genetic data is an explicit goal"

Personal Genome Project

Guidelines of the **Global PGP Network**

a) **Public Data**. Participants are invited to share genomic and trait data using a CC0 waiver

b) **Non-anonymous**. Risks of participant re-identification are addressed upfront as part of the consent and enrollment process
   - **Neither anonymity nor confidentiality of their data is promised to participants**

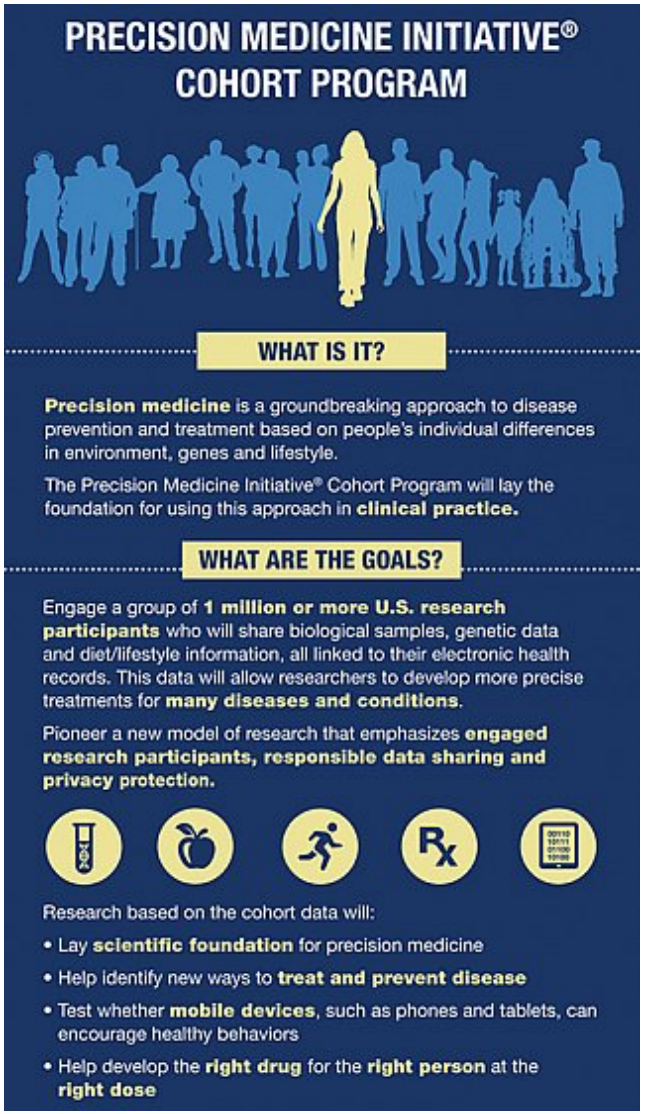
c) **Equal access.** Participants are given timely and complete access to their individual data

   i.e., raw data and not just summary results "where feasible"

d) **Oversight.** Each member must maintain current Institutional Review Board [Research Ethics] or **local equivalent approval**

e) **Not for profit.** Managed or sponsored by a non-profit organization (or local equivalent).
   - A member shall not sell or license participant data or tissues **"other than purposes of reasonable cost recovery"**

Pretty Good Privacy?

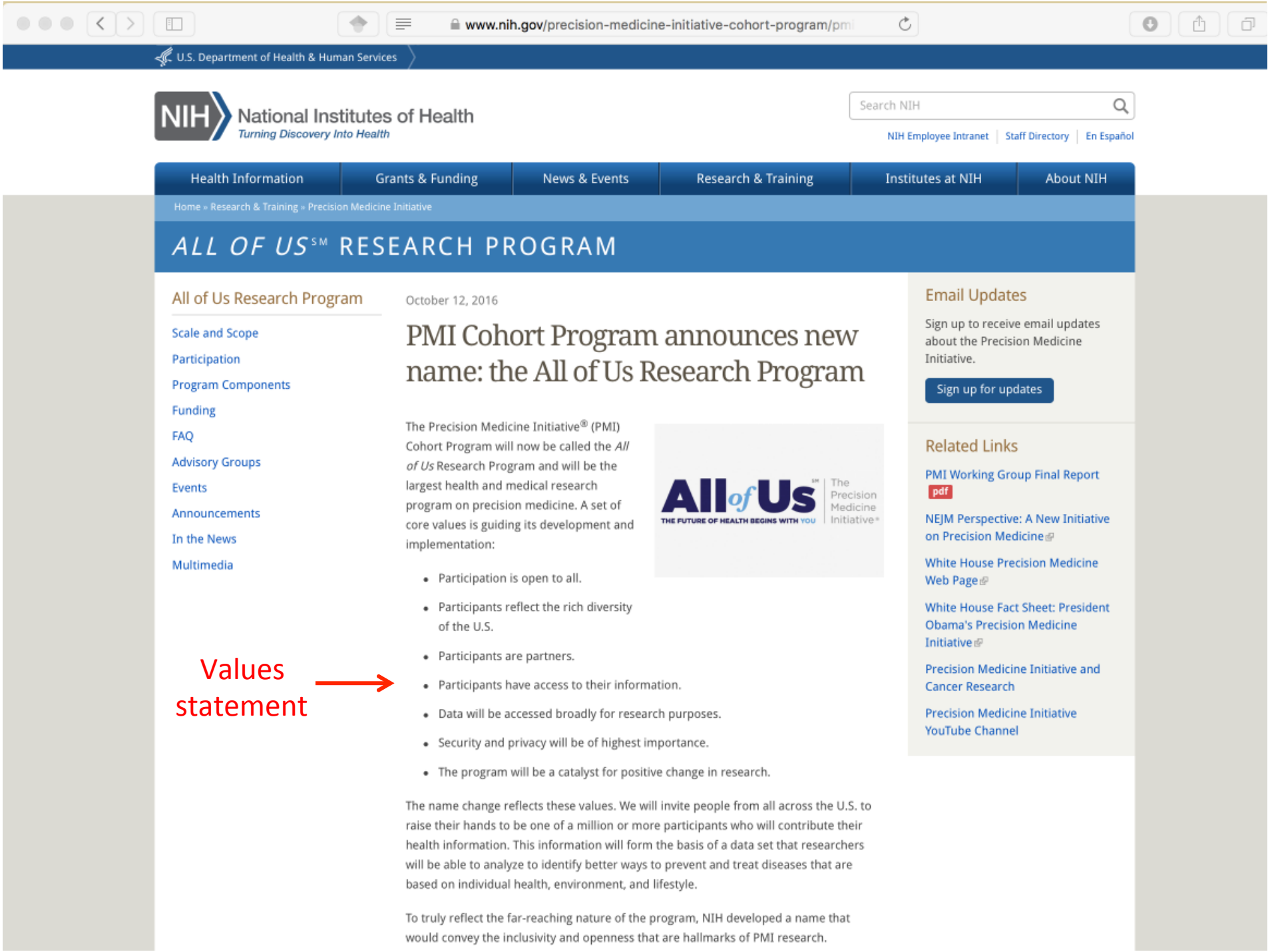


Precision Medicine Initiative

- Launched by President Obama in his January 2015 State of the Union address
- Aims to leverage advances in genomics, emerging methods for managing and analyzing large data sets, and health ICTs to accelerate biomedical discoveries
  - while protecting privacy
- Plans to enroll one million or more volunteers and may include children

"committed to engaging multiple sectors and forging strong partnerships with academic and other non-profit researchers, patient groups, and the private sector to capitalize on work already underway"

All of Us


www.nih.gov/precision-medicine-initiative-cohort-program/pm
U.S. Department of Health & Human Services
NIH National Institutes of Health
Turning Discovery Into Health
Search NIH
NIH Employee Intranet | Staff Directory | En Español
Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH
Home » Research & Training » Precision Medicine Initiative
ALL OF US℠ RESEARCH PROGRAM
All of Us Research Program
Scale and Scope
Participation
Program Components
Funding
FAQ
Advisory Groups
Events
Announcements
In the News
Multimedia
October 12, 2016
PMI Cohort Program announces new name: the All of Us Research Program
The Precision Medicine Initiative® (PMI) Cohort Program will now be called the All of Us Research Program and will be the largest health and medical research program on precision medicine. A set of core values is guiding its development and implementation:
All of Us℠ THE FUTURE OF HEALTH BEGINS WITH YOU | The Precision Medicine Initiative®
Values statement
Participation is open to all.
Participants reflect the rich diversity of the U.S.
Participants are partners.
Participants have access to their information.
Data will be accessed broadly for research purposes.
Security and privacy will be of highest importance.
The program will be a catalyst for positive change in research.
The name change reflects these values. We will invite people from all across the U.S. to raise their hands to be one of a million or more participants who will contribute their health information. This information will form the basis of a data set that researchers will be able to analyze to identify better ways to prevent and treat diseases that are based on individual health, environment, and lifestyle.
To truly reflect the far-reaching nature of the program, NIH developed a name that would convey the inclusivity and openness that are hallmarks of PMI research.
Email Updates
Sign up to receive email updates about the Precision Medicine Initiative.
Sign up for updates
Related Links
PMI Working Group Final Report pdf
NEJM Perspective: A New Initiative on Precision Medicine
White House Precision Medicine Web Page
White House Fact Sheet: President Obama's Precision Medicine Initiative
Precision Medicine Initiative and Cancer Research
Precision Medicine Initiative YouTube Channel

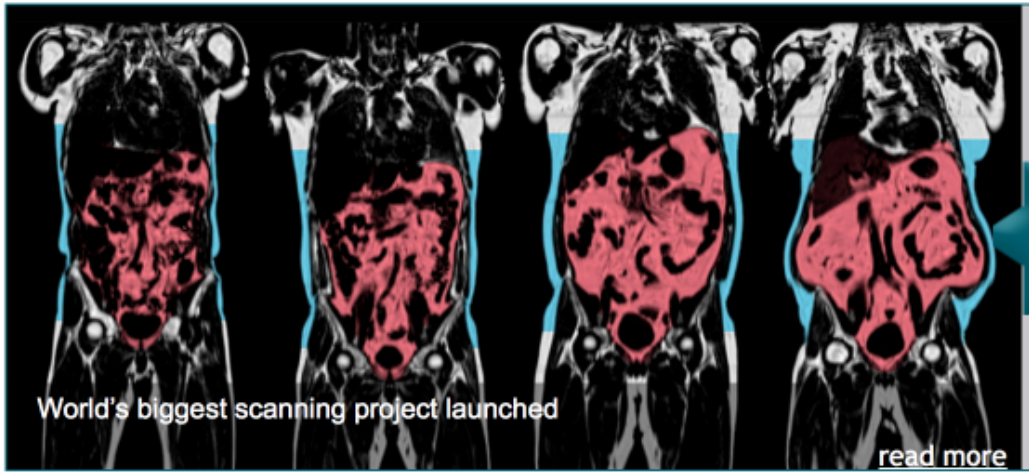
ukbiobank.ac.uk

biobank uk

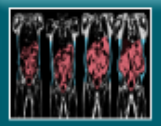
Call Us On: **0800 0 276 276**

Your feedback is important to us, tell us what you think

About | Participants | Resources | Scientists | Data Showcase | Register & Apply | Approved Research | Publications



Annual Meeting presentations available now



World's biggest scanning project launched

Scanning: looking at the whole person

Participants

- Update your contact details
- World's biggest scanning study launched
- Find out how the resource is being used
- Participant Events

- General Practice linkage →
- Data Showcase →
- Annual Meeting: watch again

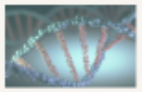
Scientists

- Data Showcase
- Activity data available in Data Showcase
- Video: How to Register and Apply
- Annual Meeting: watch again here

News

Annual Meeting presentations available now

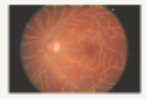
World's biggest scanning project launched



Inherited gene changes take years off life expectancy

Active commuting helps lower body fat and BMI

Short man or overweight woman? Your size could make you poorer



Assessment measures provide insight into common eye disorders

biobank uk

"A major national health resource"

- Registered charity
- Est. by Wellcome Trust, MRC, Dept. of Health, Scottish Gov., and NW Regional Dev. Agency; funded by Welsh Dev. Agency, BHF, and Diabetes UK)
- Hosted by U. Manchester, supported by NHS
- Open to *bona fide* researchers anywhere in the world, including those funded by academia and industry

Background

- Aims to improve prevention, diagnosis and treatment of life-threatening illnesses
- Recruited 500,000 people aged 40-69 in 2006-2010
- Participants have undergone measures, provided blood, urine and saliva samples, and detailed personal information
  - and agreed to have their health followed

"…to help scientists discover why some people develop particular diseases and others do not"

Best **Ethical** Practice?

UK Biobank wants to be "a model not only for best science but for best ethical practice too, in relation to these big biobank projects"

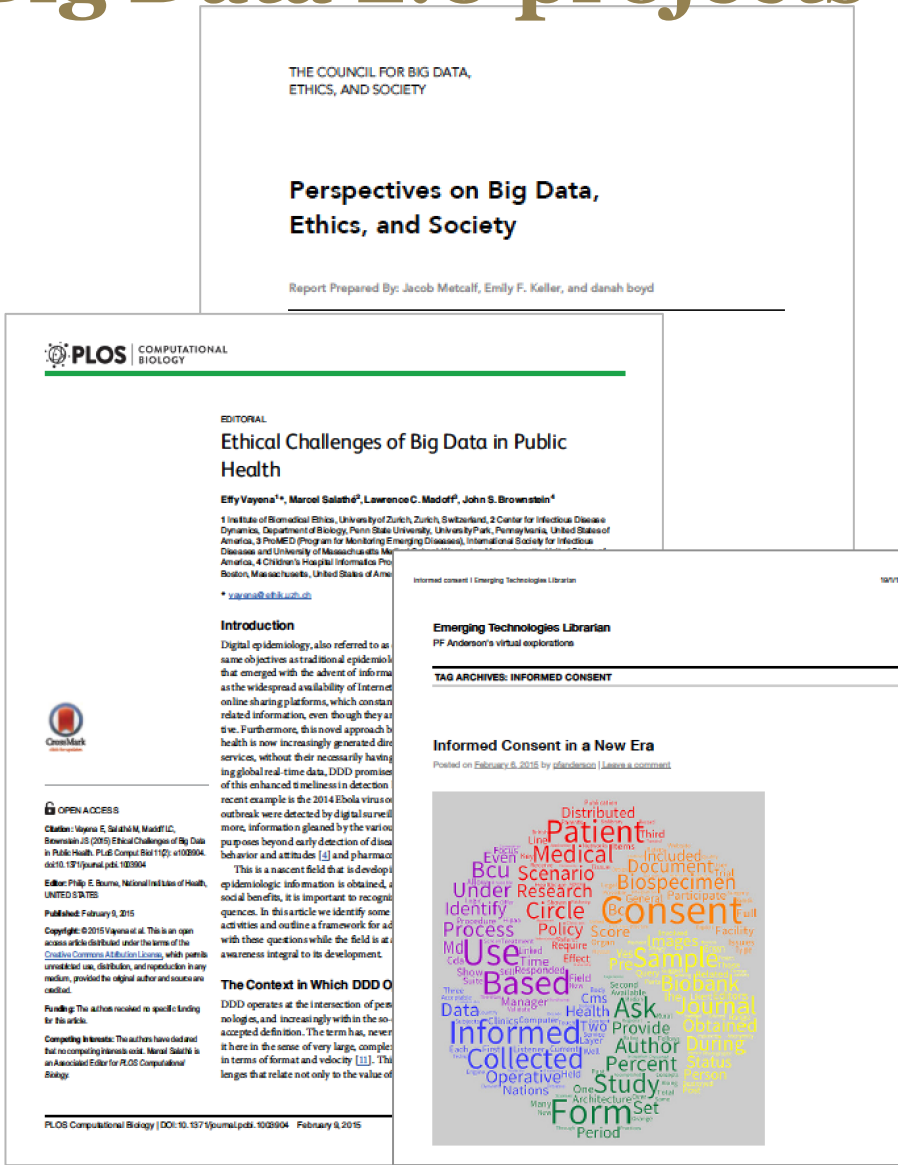
Professor Roger Brownsword, Chair (2011-2015)
UK Biobank Ethics and Governance Council (UKEGC)

http://www.ukbiobank.ac.uk/ethics/

What are some of the "best science" and "best ethical practice" lessons that can be learned from UK Biobank?

Ethical issues from Big Data 2.0 projects

- Privacy
- Anonymity
  - protection from bad actors e.g., cybercriminals, hactivists
- Monetization
  - selling of health data
- Conflicts of interest
- Informed consent
  - online informed consent, portable online/legal consent, dynamic consent, broad consent, presumed consent
- Solicitation of donors
  - for participation in other (independent) research studies



Participant Centered Consent Toolkit

Sage Bionetworks developed and maintains the *Participant-Centered Consent (PCC) toolkit*.

The goal of the PCC toolkit is to transform the concept of consent from a signature on a legal form to a process that educates, and to allow as many people as possible to engage in consented studies. The PCC toolkit is openly licensed and may be used in any context, by any user, for any purpose, as long as attribution is provided back to Sage Bionetworks.

The PCC toolkit contains a variety of resources for developing a consent process:

- PCC visual dictionary of icons and concept animations
- eConsent workflows
- Design documents and templates
- FAQ

biobank uk "…a precedent-setting case"

- Researchers wanted to use UK Biobank to identify people to invite into a separate study
- They asked UK Biobank to send an introductory email to its participants pointing to the website of the new study
- Offering such a recruitment mechanism could benefit the research community
  - But take time and resources that could be used elsewhere
- In what circumstances would it be acceptable for Biobank to divert resources in this way?
  - How should *ad hoc* third-party re-contacts be accommodated?
- UKBEGC proposed two options
  - Create a dedicated webpage to provide neutral information about (approved) studies
  - Provide a withdrawal category allowing Biobank participants opt-out from email invitations

The project was approved as a pilot subject to fitting with Biobank's timetable of re-contacts and will be used to draw up a framework for future requests

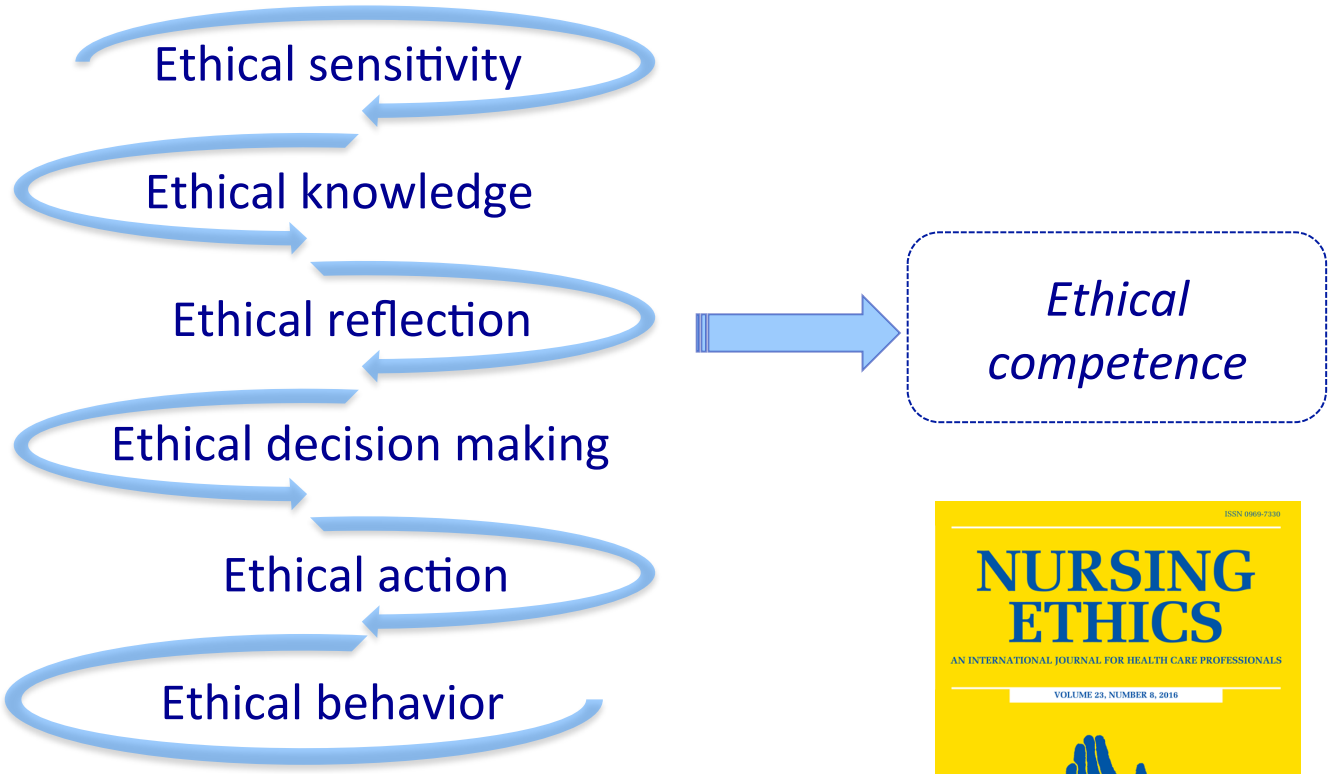
UK BIOBANK ETHICS AND GOVERNANCE COUNCIL ANNUAL REVIEW 2015

Roles for info pros as data ethicists

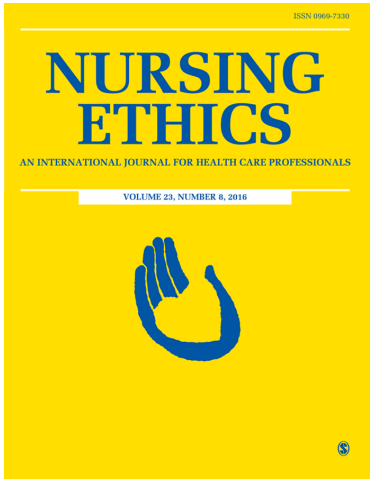
- Data advisor/consultant – dealing with queries about data access, privacy/protection, sharing, and use at reference desks or point of need
- Data literacy educators – extending information literacy instruction to data use in diverse contexts: learning, the workplace, and everyday life
- Interdisciplinary facilitators – helping multidisciplinary research teams negotiate unfamiliar information resources and intellectual norms
- Policy developers – contributing to or leading on formulation of data management policy and production of guidelines on implementation
- Research ethics partners – supporting ethics and RCR/RDM training, developing online resources, and serving on institutional review boards
- Cross-cultural mediators – anticipating cultural differences and alerting stakeholders to issues arising from plural and conflicting values

– an ethical responsibility to engage!

Components of ethical competence



(Lechasseur et al., 2016)



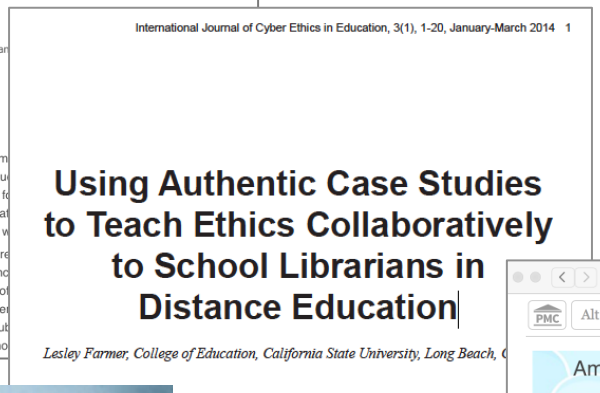
Competencies for data ethicists

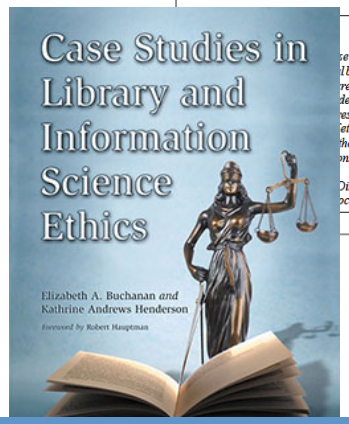
- Awareness of big data trends and developments
  - including local initiatives, programs and projects
- Knowledge of key ethical principles, codes and policies
  - both professional (library and information) ethics and research ethics (including Internet/online/virtual research ethics)
- Moral sensitivity, judgment, motivation, and character
  - the ability to recognize ethical tensions, to reflect/think critically and make informed decisions, to understand professional obligations and responsibilities, and to act and behave with respect for others
- Skills in critical thinking, reflection, analysis, and advocacy
  - already part of the information specialist's professional skillset

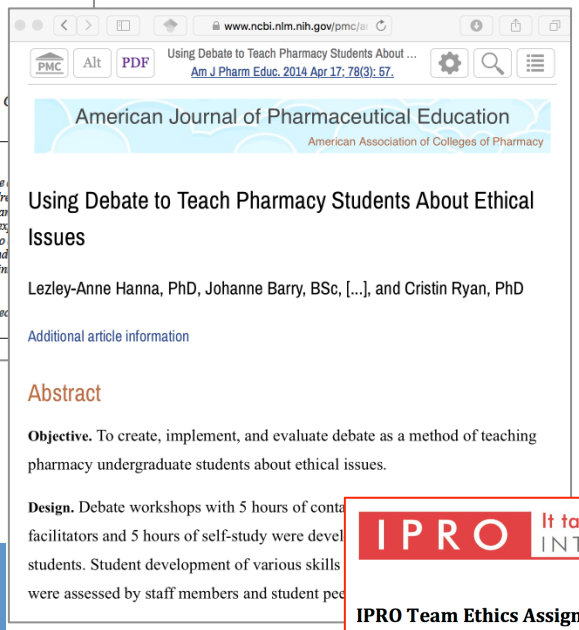
Generic and contextually situated know-how

Pedagogical strategies

- Part of a foundational course/required curriculum
- Concentrated in a required or elective standalone course
- Distributed across multiple courses
  - Integrated into specialized electives for contextually-situated learning







immersiveInformatics

Teaching methods

- Readings and cases
- Debates, role plays, and observations – shadowing *in situ*

IPRO It takes a team! INTERPROFESSIONAL PROJECTS PROGRAM

IPRO Team Ethics Assignment
Role-Play

Education for data ethicists

- Guided reading, listening, and viewing
  - Blog posts, Books, Journal articles, Magazine features, News stories, Videos…
- Case studies/reports/materials
  - Original or published sets of paradigmatic cases for analysis and discussion, with questions and/or commentaries to encourage critical reflection
- Formal debates (with voting)
  - Divide students into groups and teams, provide a debate motion, assign to particular sides (proposing or opposing), supply instructions/resources, and ask team members to develop arguments supporting the views assigned
  - Onsite/online (synchronous web conference or asynchronous discussion board)
- Role plays (and observation)
  - Ask students to play a character in a case study portraying a realistic, difficult ethical situation that illustrates divergent perspectives on the same problem
  - Provide descriptions of the situation and script/leading lines to get them started
- Immersive experiences
  - Students spend a day observing researchers working with data in a laboratory

Framework for Ethical Decision Making

Recognize an Ethical Issue

1. Could this decision or situation be damaging to someone or to some group?
2. Does this decision involve a choice between a good and bad alternative, or perhaps between two "goods" or between two "bads"?

Get the Facts

3. What are the relevant facts of the case? What facts are not known? Can I learn more about the situation? Do I know enough to make a decision?
4. What individuals and groups have an important stake in the outcome? Are some concerns more important? Why?
5. What are the options for acting? Have all the relevant persons and groups been consulted? Have I identified creative options?

Framework for Ethical Decision Making (continued)

Evaluate Alternative Actions

6. Evaluate the options by asking the following questions:
   - Which option will produce the most good and do the least harm? (The Utilitarian Approach)
   - Which option best respects the rights of all who have a stake? (The Rights Approach)
   - Which option treats people equally or proportionately? (The Justice Approach)
   - Which option best serves the community as a whole, not just some members? (The Common Good Approach)
   - Which option leads me to act as the sort of person I want to be? (The Virtue Approach)

Make a Decision and Test It

7. Considering all these approaches, which option best addresses the situation?
8. If I told someone I respect-or told a television audience-which option I have chosen, what would they say?

Act and Reflect on the Outcome

9. How can my decision be implemented with the greatest care and attention to the concerns of all stakeholders?
10. How did my decision turn out and what have I learned from this specific situation?

Review and summary

- Information professionals can act as big data 2.0 advisors/ consultants, interdisciplinary facilitators, policy developers, research ethics partners, and cross-cultural mediators
- Big data workers need to develop **awareness** of trends and developments in the field, **knowledge** of ethics principles, codes, and policies, **moral sensitivity**, **judgment**, **motivation**, and **behavior**, and **skills** in critical thinking and reflection, analysis, and advocacy.
- We can prepare students for working in the big data arena with **readings** and **viewings**, **case studies**, **debates**, **role plays**, and immersive experiences of **shadowing** researchers in labs
  - by embedding data ethics education in foundational and other courses and our outside-the-classroom curriculum

Questions and comments

Sheila Corrall

Department of Information Culture
& Data Stewardship

scorrall@pitt.edu
www.ischool.pitt.edu/people/corrall.php