

45th LIBER Annual Conference
Libraries Opening Paths to Knowledge
Wednesday, June 29, 2016

Legal, Ethical, and Policy
Issues of “**Big Data 2.0**”
**Collaborative Ventures
and Roles for Info Pros**

Sheila Corrall
Kip Carrier





Legal, Ethical, and Policy Issues of “Big Data 2.0” Collaborative Ventures and Roles for Info Pros

Outline

- Background

Library literature – Definition of key terms

- Case studies

Pittsburgh Health Data Alliance – UK Biobank – Big Data Europe –
Personal Genome Project – Precision Medicine Initiative –
Oncology Research Information Exchange Network

- Implications

Legal, ethical, policy

- Conclusions

Roles and competencies





Library literature

- Educating students about how companies use big data and advising users on how to find datasets for research (Bieraugel, 2013; Hoy, 2014)
- Moving beyond research data management to define and discuss other specialized data-related roles (Lyon & Brenner, 2015; Lyon et al., 2016)
- Exposing (linked) library collections data and making them reusable for resource discovery (Campbell & Cowan, 2016; Teets & Goldner, 2013)
- Carrying out their own big data projects to analyze collection use and conduct cross-disciplinary comparisons (Huwe, 2014; Tattersall, 2016)
- Helping communities create local data infrastructures and make big data more useful, by creating taxonomies, designing metadata schemes, and systematizing retrieval methods, and also assisting with policy concerns (Bertot et al., 2014; Bieraugel, 2013; Reinhalter & Wittman, 2014)
- Serving as authorities on copyright and intellectual property issues arising from big data (Gordon-Murnane, 2012)



What are data? (*When are data?*)

Data are forms of information that may be defined by example, processing level, origin, and preservation value

“In addition to digital manifestations of literature (including text, sound, still images, moving images, models, games, or simulations), [the term] refers as well to forms of data and databases that generally require the assistance of computational machinery and software in order to be useful, such as various types of laboratory data including spectrographic, genomic sequencing, and electron microscopy data; observational data, such as remote sensing, geospatial, and socioeconomic data; and other forms of data either generated or compiled, by humans or machines.”

(Uhlir & Cohen in Borgman, 2015, p. 19)



Critical questions

- What are the chief **legal, ethical, and policy issues** triggered by Big Data (and Little Data)?
- What **best practices** can be identified to address these kinds of legal, ethical, and policy issues?
- What are the **roles** that information professionals and research libraries can and will assume in contributing to considerations of the legal, ethical, and policy issues raised?
- What are the competency implications in terms of the **knowledge, skills, and abilities** libraries need to acquire or develop for the Big Data world?

Three Pittsburgh institutions.

One goal.

Pittsburgh Health Data Alliance

Carnegie Mellon University

University of Pittsburgh

UPMC

The future of health care is in the data.



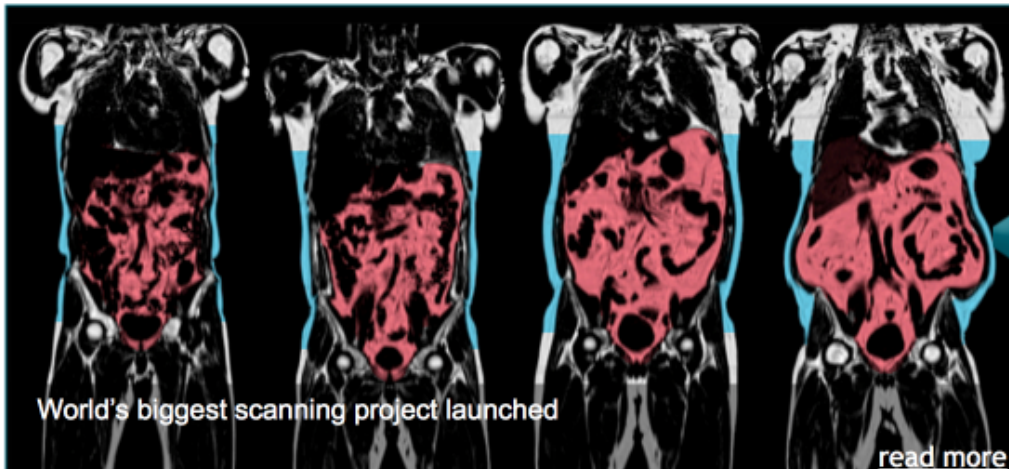
Pittsburgh Health Data Alliance

Carnegie Mellon University University of Pittsburgh UPMC

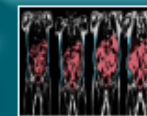
World-class CS/machine learning + Medical research expertise + Deep data, clinical setting, commercialization

What roles can information professionals and research libraries 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.”



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](#)

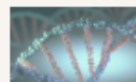
News



Annual Meeting presentations available now



World's biggest scanning project launched



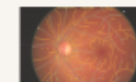
Inherited gene changes take years of life expectancy



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



Background

“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

- 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?



BIG DATA EUROPE

Empowering Communities with Data Technologies

HOME

WEBINARS

ABOUT

SC NETWORKS

EVENTS

CONTACTS



Health



Food & Agriculture



Energy



Transport



Climate



Social Sciences



Security



Our workshops provide key opportunities for stakeholders in the Societal Challenges to influence the design of BigDataEurope's Big Data platform.

[Join one of our workshops!](#)

Upcoming Events

MON
27

The ICPSR Summer Programme in Quantitative Methods of Social Research

June 20 @ 09:00 - July 29 @ 18:00

MON
27

"Behind the scenes of the BigDataEurope Transport Pilot" Hangout

June 27 @ 16:00 - 17:00



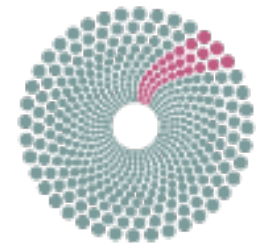
Big Data Europe

Who is Big Data Europe for?

- “Small, Medium and large-sized entities coming from any sector within industry, research or the public sector, that have much to gain from making sense of large volumes of data (of both static or dynamic nature, and from various sources) to realise new and innovative use-cases, not just within their domain but also across different sectors”
- 16 European partners at present, representing a diverse range of academic, for-profit, and government entities in 10 countries

Big Data partnership projects – A key question

- Given current political uncertainties (e.g., BREXIT), what can be done to ensure stability and continuity of Big Data partnerships (like Big Data Europe), while providing leeway for accommodating changes and course corrections that may be periodically warranted?



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Sharing Personal Genomes

Genom Austria, as part of the Global Network of Personal Genome Projects, aims to create a dataset of openly available human genomes. It also contributes to the public discussion about genomes in science, medicine and society. Sharing genome data is critical to scientific progress, but has been hampered by traditional research practices. Our approach is to invite volunteering participants to publicly share their personal genome data for the greater good.

[Learn more >](#)



Participation

Genom Austria invites participation of volunteers who are willing to share their personal genomes with the public. Making personal genome publicly and freely available is a great way to help advance our understanding of human genetics, biology, and medicine.

[Learn about participating >](#)

Open Data

Open data sharing is very important for science. But because personal genomes are identifiable and predictive, many studies restrict the access to genomic data. In contrast, Genom Austria's personal genomes are openly and freely available for both scientists and the general public.

[View Genom Austria data >](#)

Global Network

Genom Austria is a member of the Global Network of Personal Genome Projects. Since the Personal Genome Project was launched at Harvard Medical School in 2005, the network has grown to include researchers at many leading institutions around the globe.

[Find out about the network >>](#)

PGP Global Network

[United States](#) • [Canada](#) • [United Kingdom](#) • [Austria](#)
[Learn about our network](#)

Website information

[Contact Us](#) • [About PersonalGenomes.org](#) • [Terms of Service](#) • [Privacy Policy](#)

About PGP UK

Founded in 2013 by Stephan Beck, the United Kingdom Personal Genome Project is located at University College London.

PGP UK is a member of the Global Network of Personal Genome Projects (PGP), a group of research studies creating freely available scientific resources that bring together genomic, environmental and human trait data donated by volunteers. Initiated by George Church at Harvard Medical School in 2005, the Personal Genome Project has pioneered ethical, legal, and technical aspects related to the creation of public resources involving highly identifiable data like human genomes.

Public Data, Methods, and Materials

We believe sharing is good for science and society. Our project is dedicated to creating public resources that everyone can access. Privacy, confidentiality and anonymity are impossible to guarantee in a context like this research study where public sharing of genetic data is an explicit goal. Therefore, our project collaborates with participants who are fully aware of the implications and privacy concerns of making their data public. Volunteering is not for everyone, but the participants who join make a valuable and lasting contribution to science.

Ongoing Participatory Research

We respect the people behind the data, and we aim to maintain strong relationships with participants. We want to collaborate on tracking health and other traits as they unfold over time. We also want to better understand the benefits and risks related to accessing and sharing personal genomes and other types of data.

Genomes, Environments, and Traits

The genome is just a part of the story: genes interact with the environment to form traits. Participants may choose to contribute other public data to build public records of their health and traits. We also try to connect participants with research, education, and citizen science projects that are connected to personal genome data.

PGP Global Network

[United States](#) • [Canada](#) • [United Kingdom](#) • [Austria](#)
[Learn about our network](#)

Website information

[Contact Us](#) • [About PersonalGenomes.org](#) • [Terms of Service](#) • [Privacy Policy](#)



About PGP *genom austria*

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”





Guidelines of the Global PGP Network


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

**PRECISION MEDICINE INITIATIVE®
COHORT PROGRAM**



WHAT IS IT?


Precision medicine is a groundbreaking approach to disease prevention and treatment based on people's individual differences in environment, genes and lifestyle.

The Precision Medicine Initiative® Cohort Program will lay the foundation for using this approach in **clinical practice**.

WHAT ARE THE GOALS?

Engage a group of **1 million or more U.S. research participants** who will share biological samples, genetic data and diet/lifestyle information, all linked to their electronic health records. This data will allow researchers to develop more precise treatments for **many diseases and conditions**.

Pioneer a new model of research that emphasizes **engaged research participants, responsible data sharing and privacy protection**.



Research based on the cohort data will:

- Lay **scientific foundation** for precision medicine
- Help identify new ways to **treat and prevent disease**
- Test whether **mobile devices**, such as phones and tablets, can encourage healthy behaviors
- Help develop the **right drug** for the **right person** at the **right dose**

- 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”

ORIEN

ONCOLOGY RESEARCH
INFORMATION
EXCHANGE NETWORK

Accelerating Cancer Discovery And
Delivering Hope Through Collaborative
Learning And Partnerships





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.

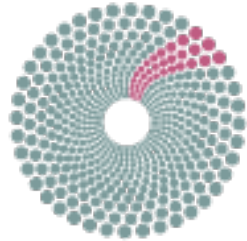
DATA SHARING TO GUIDE DISCOVERY

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.





biobank^{uk}



BIG DATA EUROPE

Empowering Communities
with Data Technologies



ORIEN ONCOLOGY RESEARCH
INFORMATION
EXCHANGE NETWORK

Big projects, Big problems

- Very large scale
- Interdisciplinary
- Human subjects
- Inter-state/international/global
- Multiple jurisdictions
- Cross-sector partners (public/private)
- Cultural differences

Pittsburgh Health Data Alliance

.....
Carnegie Mellon University University of Pittsburgh UPMC



Legal issues arising from Big Data

Compliance with

- Privacy laws
- Data protection/security laws
- Genetic information laws
- Freedom of information
- Right to be forgotten
- Intellectual property
 - e.g., patenting of human genes/synthetic human genes
 - cf. EU and US (*Myriad Genetics* case, 2013)
- Licensing and contractual issues
- Publishing



Ethical issues arising from Big Data

- Privacy
 - of donors
 - how to comply with privacy laws of different nations/groups
- Maintaining anonymity of specimen donors
 - protection against bad actors, e.g., cybercriminals, hactivists
 - triangulation of data from multiple sources used to circumvent anonymization of donors
- Monetization, Commodification
 - selling of health data to commercial interests
 - use of indigenous knowledge/traditional knowledge
 - should specimen donors share in any potential profits?



Ethical issues arising from Big Data

- Peaceful/Public Good/Public Interest uses vs. Military/
National Security uses vs. Terrorist applications
 - who will determine the societally acceptable/desirable uses and applications for health data/big data?
- Psychological well-being/Informed consent of donors
 - fully advising donors of their rights and of the obligations of the respective data-gathering and data-using entities to donors
 - taking account of the best interests of donors in making their data available to them
- Solicitation of specimen donors for participation in studies

In 2015 the UK BioBank Ethics and Governance Council faced a policy issue over its proposed use as a recruitment platform by researchers who wanted to identify people for a separate study



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



Policy issues arising from Big Data

- How and by whom will health data/big data be preserved and made retrievable for and by future stakeholders?
- What guidelines and requirements are needed for publishing related to health data/big data?
- Who needs to have a voice in policy-setting and policy-making, and who should craft the governing policies and codes of ethics?
 - ☞ Given the pace of change, how often should policies and codes be reviewed and updated?
- What oversight and enforcement mechanisms are needed to ensure compliance?
 - ☞ What are the penalties for piracy of health data or malfeasance, negligence, willful blindness, and harmful impacts on human subjects?
 - ☞ What protections are available or need to be developed and codified for whistleblowers who report lapses and breaches of compliance?



Library Roles and Competencies

- Data are forms of information requiring stewardship
 - like the many other knowledge resources libraries manage
- Big data 2.0 initiatives pose particular challenges
 - because of their scale, variety, complexity, and openness
- Libraries are well positioned to assume a proactive role
 - building on their existing work in scholarly communication
- Potential roles for libraries in the big data arena require professional, technical, organizational, managerial, personal, and interpersonal knowledge, skills, and abilities
 - including expertise associated with other professions and enhanced competencies in relationship management



Potential Library Roles in Open Domains

<i>Types</i>	Open Content				Open Process				Open Infrastructure			
<i>Domains</i>	OA	OData	OER	OBib	OSS	OD	OEP	OPR	OSci	OI	OStd	OSys
<i>Roles</i>												
Use												
Educate												
Advocate												
Facilitate												
Mediate												
Collaborate												
Coordinate												
Integrate												
Lead												

(Corrall, 2016, In press)



References

- Bertot, J., Butler, B., & Travis, D. (2014). Local big data: The role of libraries in building community data infrastructures. *dg.o 2014: Proceedings of the 15th Annual International Conference on Digital Government Research* (pp. 17-23). doi:10.1145/2612733.2612762.
- Bieraugel, M. (2013). Keeping up with... big data. ALA, ACRL. Retrieved from http://www.ala.org/acrl/publications/keeping_u_with/big_data.
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Campbell, D. G., & Cowan, S. R. (2016). The paradox of privacy: Revisiting a core library value in an age of big data and linked data. *Library Trends*, 64(3), 492-511. doi:10.1353/lib.2016.0006.
- Gordon-Murnane, L. (2012). Big data: A big opportunity for librarians. *Online*, 36(5), 30-34.
- Hoy, M. B. (2014) Big data: An introduction for librarians. *Medical Reference Services Quarterly*, 33(3), 320-326. doi:10.1080/02763869.2014.925709.
- Huwe, T. K. (2014, March). Big data and the library: A natural fit. *Computers in Libraries*, 34(2), 17-18.



References

- Huwe, T. K. (2014, March). Big data and the library: A natural fit. *Computers in Libraries, 34*(2), 17-18.
- Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. *International Journal of Digital Curation, 10*(1), 111-122. doi:10.2218/ijdc.v10i1.349.
- Lyon, L., Acker, A., Mattern, E., & Langmead, A. (2016). Applying translational principles to data science curriculum development. *iPres 2015: Proceedings of the 12th International Conference on Digital Preservation* (pp. 109-117). Retrieved from <https://phaidra.univie.ac.uk/view/o:429552>.
- Reinhalter, L., & Wittmann, R. J. (2014). The library: Big data's boomtown: *The Serials Librarian, 67*(4), 363-372. doi:10.1080/0361526X.2014.915605.
- Tattersall, A. (2016). Big data – what is it and why it matters. *Health Information & Libraries Journal, 33*(2), 89-91. doi:10.1111/hir.12147.
- Teets, M., & Goldner, M. (2013). Libraries' role in curating and exposing big data. *Future Internet, 5*(3), 429-438. doi:10.3390/fi5030429.

Any Questions?

Sheila Corral

scorrall@pitt.edu

Kip Currier

kcurrier@sis.pitt.edu

**Department of Information Culture
& Data Stewardship**
School of Information Sciences

