Charleston Conference

Roll with the Times…

November 4, 2016

Big Data 2.0

Critical Roles for Libraries and Librarians

MOODS!

Sheila Corrall

scorrall@pitt.edu

Big Data 2.0: Critical Roles for Libraries and Librarians

Outline

- The data shift – the evolving (library) data landscape
- Current + emergent data roles of libraries and librarians
  - projects, partnerships, communities, and concerns
- The next frontier – Big Data 2.0 global megaprojects
  a.k.a. **MOODS** = **M**assive **O**pen **O**nline **D**ata-driven **S**tudies
- Issues and roles for libraries and librarians in the BD 2.0 arena

Charleston Conference™

ISSUES IN BOOK AND SERIAL ACQUISITION

October 31 - November 5, 2016

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

Data as the new currency

Government's role in facilitating the exchange

BY WILLIAM D. EGGERS, ROB HAMILL AND ABED ALI

> ILLUSTRATION BY JOHN HERSEY

Deloitte Review | DELOITTEREVIEW.COM

The Evolving (Library) **Data** Landscape

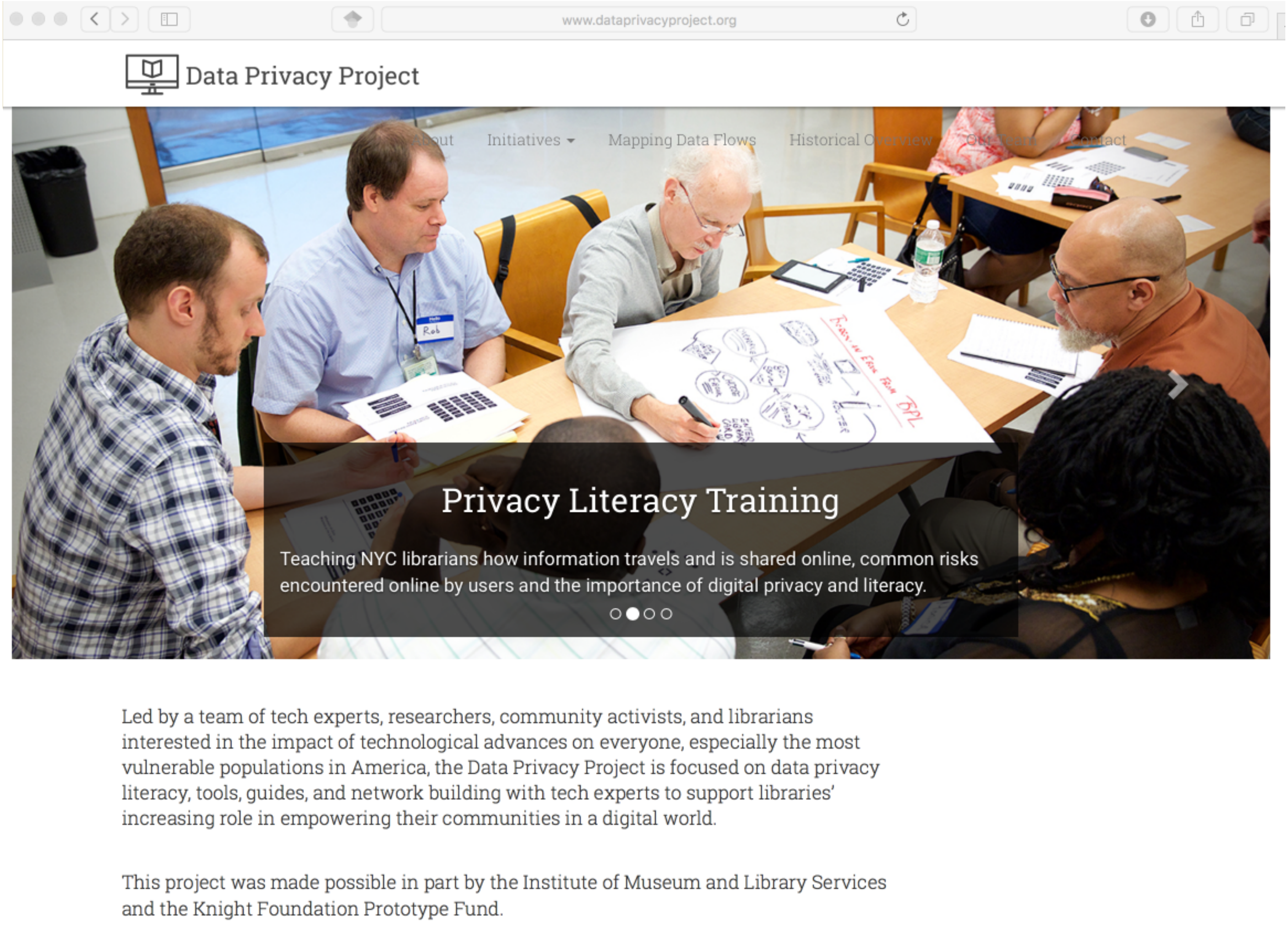
- Social science data archives and geospatial data resources
- Networked data-intensive science and research data services
- Linked open collections data for cultural heritage institutions
- Data analytics and visualization for research and assessment
- Creating local data infrastructures and metadata 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 copyright and IPR issues arising from big data

Current Roles of Libraries and Librarians

- **Access management** – facilitating use of external datasets
- **Data literacy** – helping users exploit existing data resources, educating students and faculty about managing research data + preparing graduates for working with data in employment
- **Collection building** – auditing and appraising data assets, selecting and creating metadata/cataloging selected datasets
- **Digital curation –** capturing, organizing, preserving, and archiving research data generated by students and faculty
- **Publishing support** – advising researchers on identifying, citing, sharing, licensing, and demonstrating impact of data
- **Policy development** – consulting stakeholders, drafting and revising documents, advocacy of take-up and implementation

Emergent Roles of Libraries and Librarians

- **Data literacy** – preparing frontline library staff to respond to patron needs regarding digital privacy or data profiling
  - + training youth librarians to deliver technology-supported informal learning programs for teens in public libraries
- **Metadata consultancy** – providing specialized expertise to support open data sharing by municipal government agencies
- **Infrastructure development** – serving as local platforms for gathering, using, and developing data in smaller communities (e.g., hosting data hackathons and facilitating data deposit)
  - + building, testing, and evaluating custom infrastructure for storing, transferring, and processing Big Data for re-use
- **Data protection** – promoting responsible use of personal data

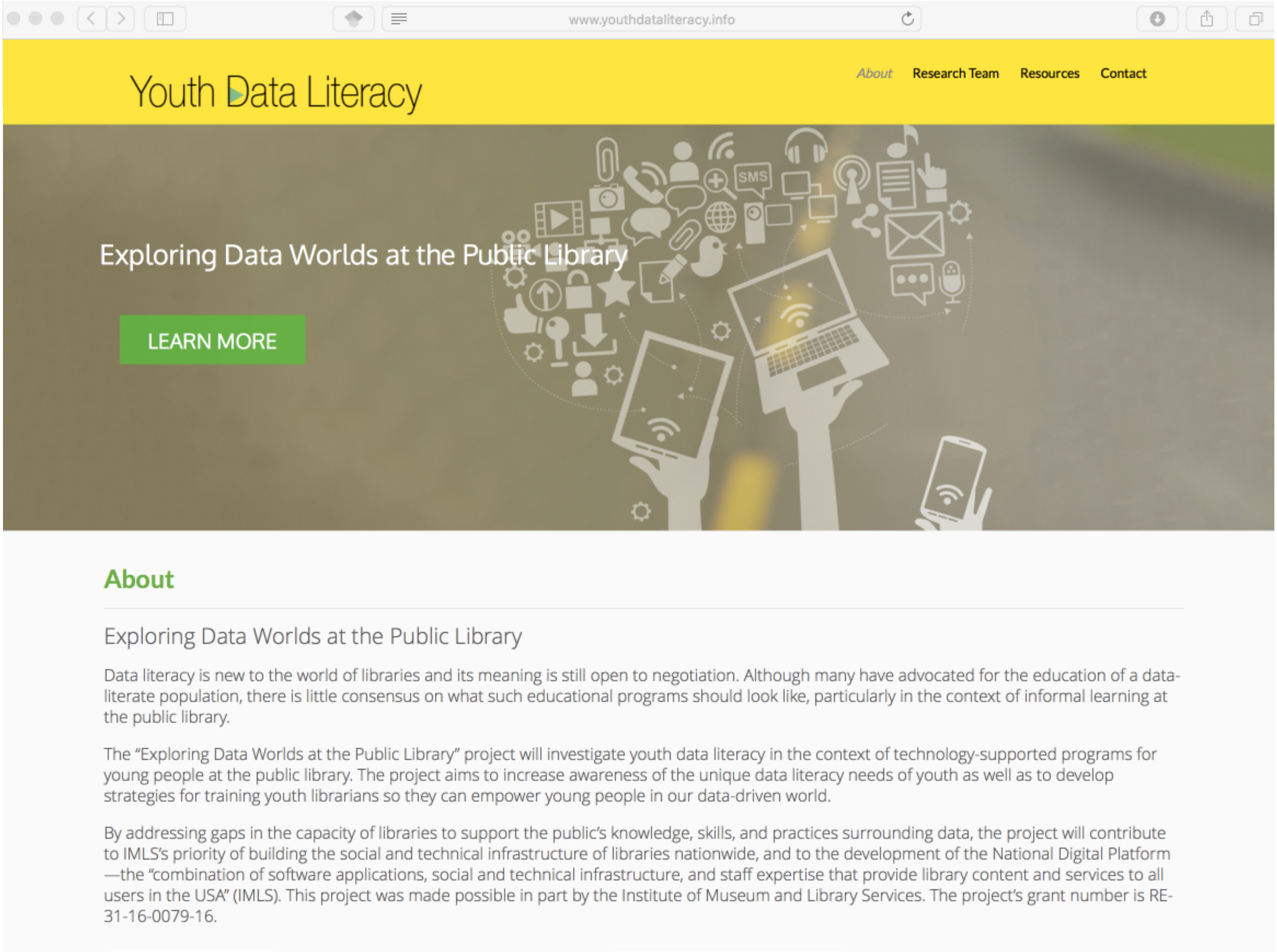


Privacy Literacy Training

Teaching NYC librarians how information travels and is shared online, common risks encountered online by users and the importance of digital privacy and literacy.

Led by a team of tech experts, researchers, community activists, and librarians interested in the impact of technological advances on everyone, especially the most vulnerable populations in America, the Data Privacy Project is focused on data privacy literacy, tools, guides, and network building with tech experts to support libraries' increasing role in empowering their communities in a digital world.

This project was made possible in part by the Institute of Museum and Library Services and the Knight Foundation Prototype Fund.


www.youthdataliteracy.info
Youth Data Literacy
About
Research Team
Resources
Contact
Exploring Data Worlds at the Public Library
LEARN MORE
SMS

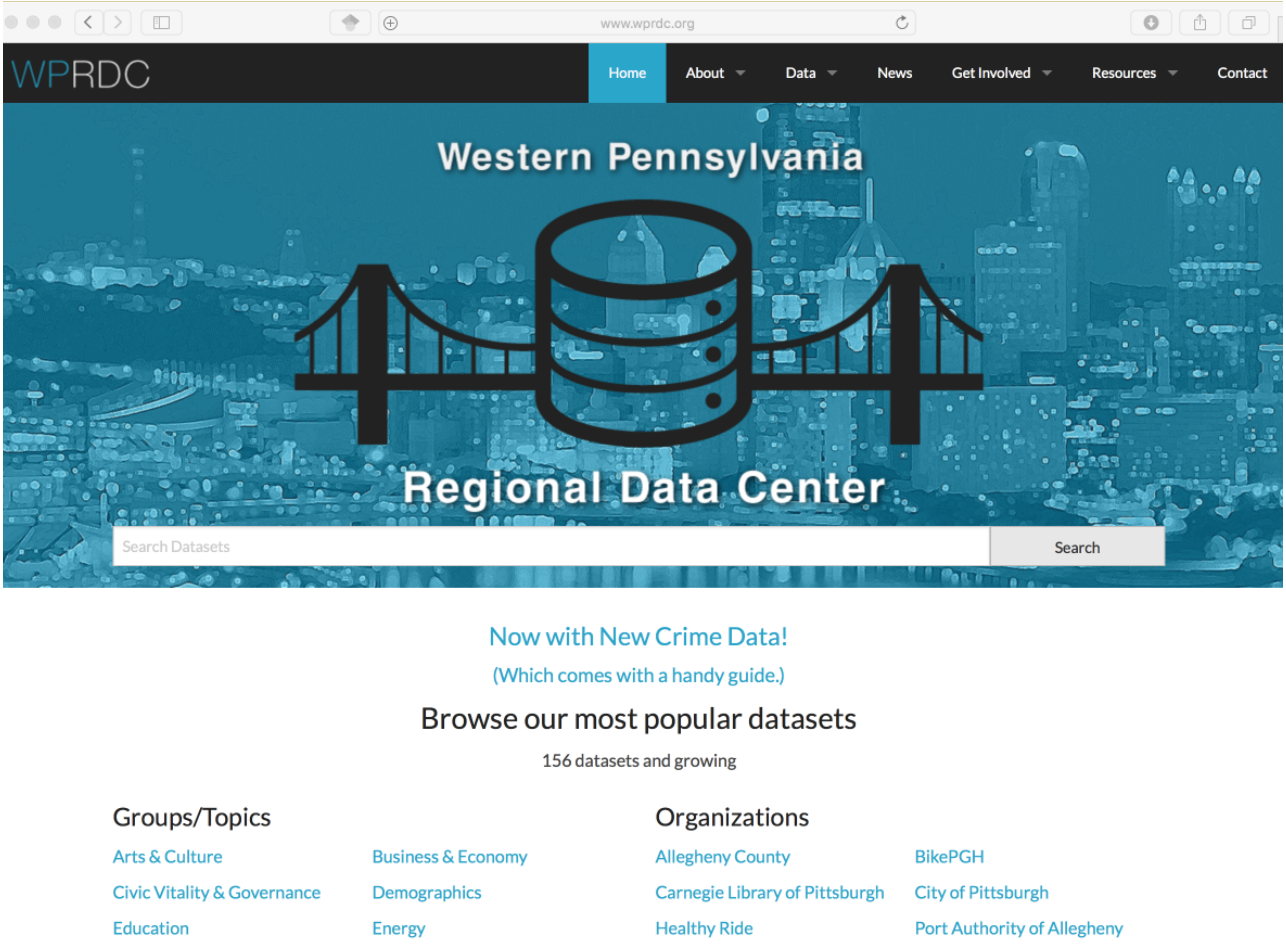
About

Exploring Data Worlds at the Public Library

Data literacy is new to the world of libraries and its meaning is still open to negotiation. Although many have advocated for the education of a data-literate population, there is little consensus on what such educational programs should look like, particularly in the context of informal learning at the public library.

The "Exploring Data Worlds at the Public Library" project will investigate youth data literacy in the context of technology-supported programs for young people at the public library. The project aims to increase awareness of the unique data literacy needs of youth as well as to develop strategies for training youth librarians so they can empower young people in our data-driven world.

By addressing gaps in the capacity of libraries to support the public's knowledge, skills, and practices surrounding data, the project will contribute to IMLS's priority of building the social and technical infrastructure of libraries nationwide, and to the development of the National Digital Platform—the "combination of software applications, social and technical infrastructure, and staff expertise that provide library content and services to all users in the USA" (IMLS). This project was made possible in part by the Institute of Museum and Library Services. The project's grant number is RE-31-16-0079-16.



Western Pennsylvania
Regional Data Center

Now with New Crime Data!

(Which comes with a handy guide.)

Browse our most popular datasets

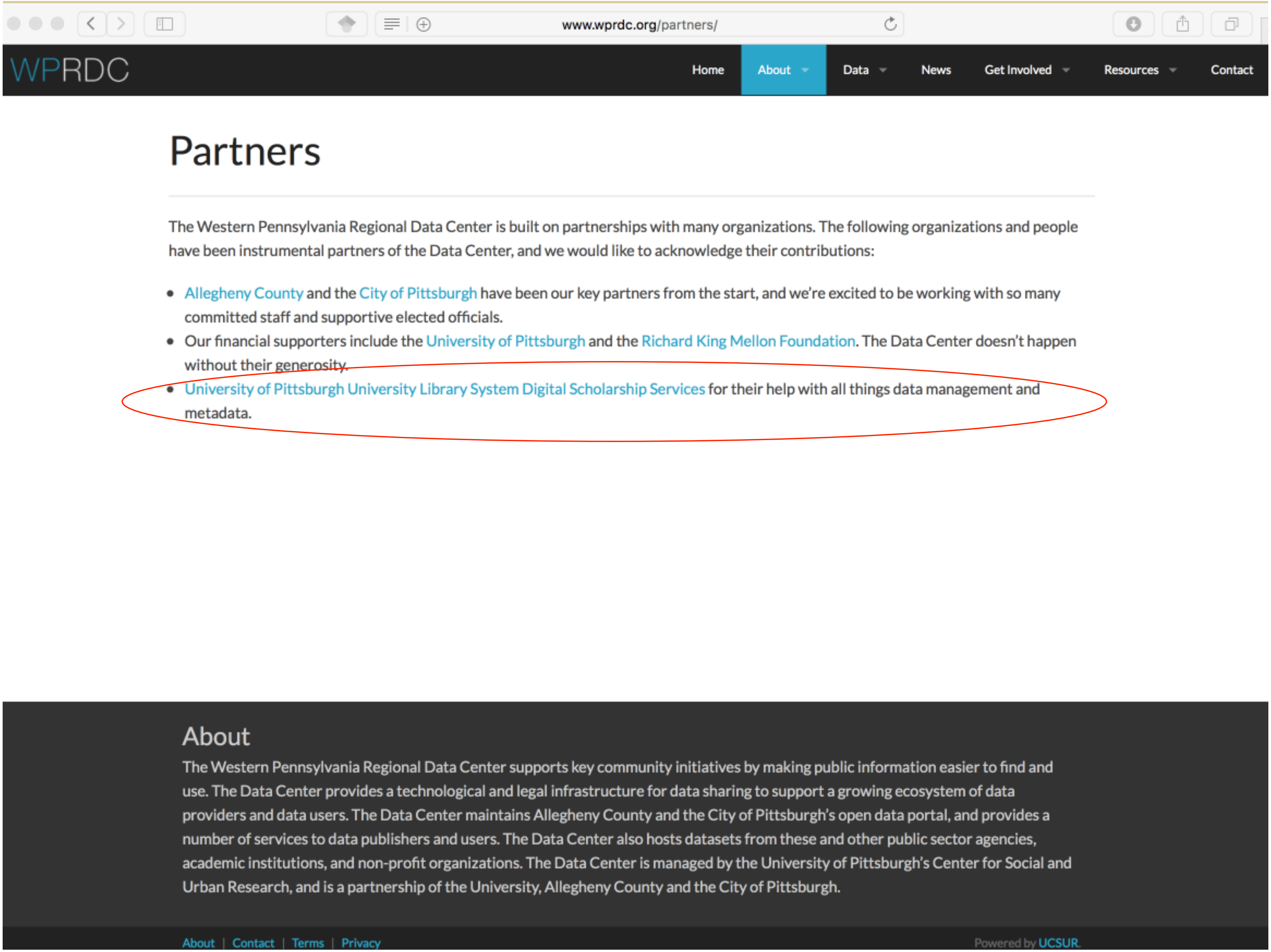
156 datasets and growing

Groups/Topics

Arts & Culture	Business & Economy
Civic Vitality & Governance	Demographics
Education	Energy

Organizations

Allegheny County	BikePGH
Carnegie Library of Pittsburgh	City of Pittsburgh
Healthy Ride	Port Authority of Allegheny



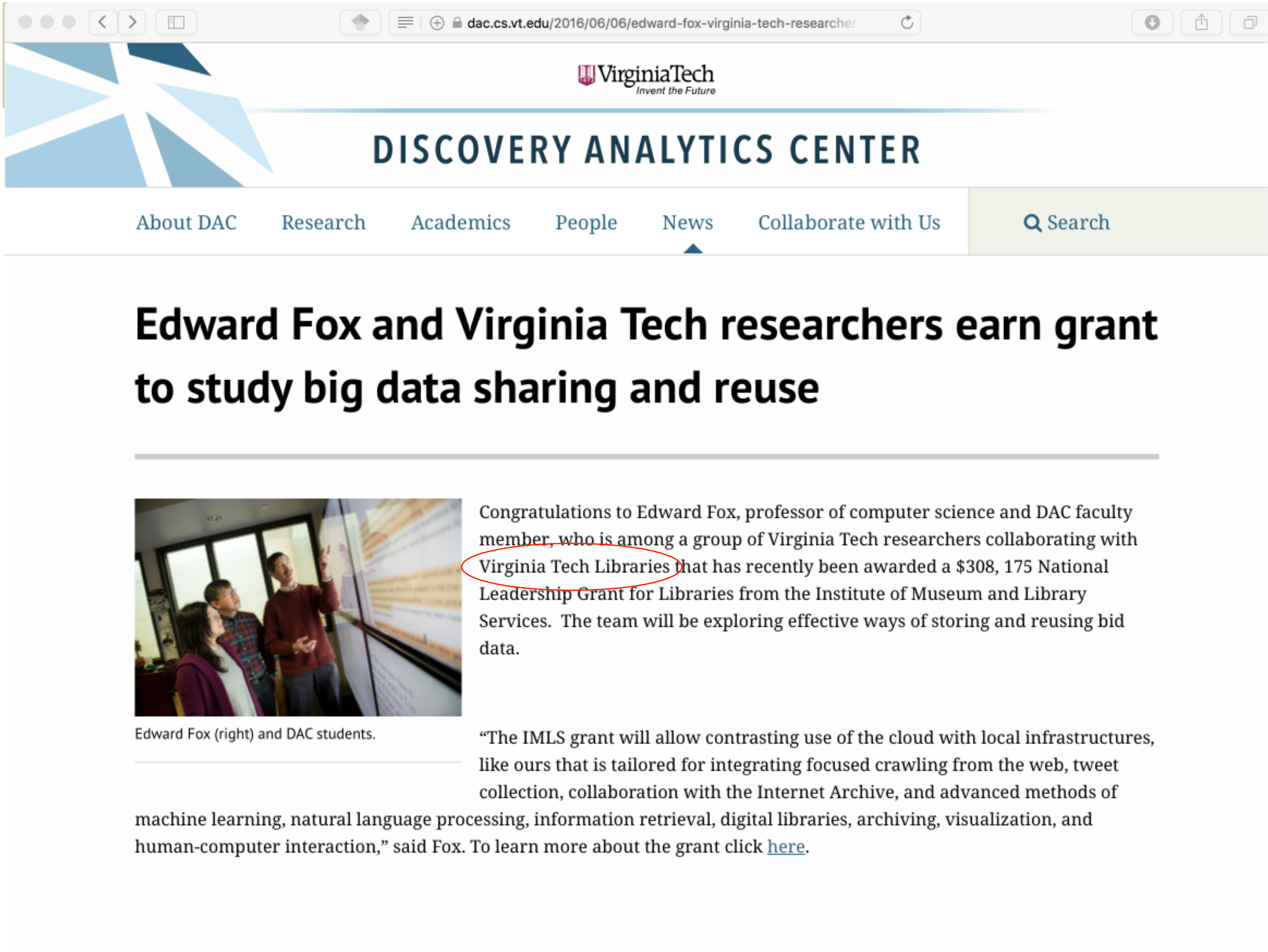
Partners

The Western Pennsylvania Regional Data Center is built on partnerships with many organizations. The following organizations and people have been instrumental partners of the Data Center, and we would like to acknowledge their contributions:

- Allegheny County and the City of Pittsburgh have been our key partners from the start, and we're excited to be working with so many committed staff and supportive elected officials.
- Our financial supporters include the University of Pittsburgh and the Richard King Mellon Foundation. The Data Center doesn't happen without their generosity.
- University of Pittsburgh University Library System Digital Scholarship Services for their help with all things data management and metadata.

About

The Western Pennsylvania Regional Data Center supports key community initiatives by making public information easier to find and use. The Data Center provides a technological and legal infrastructure for data sharing to support a growing ecosystem of data providers and data users. The Data Center maintains Allegheny County and the City of Pittsburgh's open data portal, and provides a number of services to data publishers and users. The Data Center also hosts datasets from these and other public sector agencies, academic institutions, and non-profit organizations. The Data Center is managed by the University of Pittsburgh's Center for Social and Urban Research, and is a partnership of the University, Allegheny County and the City of Pittsburgh.



Edward Fox and Virginia Tech researchers earn grant to study big data sharing and reuse

Edward Fox (right) and DAC students.

Congratulations to Edward Fox, professor of computer science and DAC faculty member, who is among a group of Virginia Tech researchers collaborating with Virginia Tech Libraries that has recently been awarded a $308, 175 National Leadership Grant for Libraries from the Institute of Museum and Library Services. The team will be exploring effective ways of storing and reusing bid data.

"The IMLS grant will allow contrasting use of the cloud with local infrastructures, like ours that is tailored for integrating focused crawling from the web, tweet collection, collaboration with the Internet Archive, and advanced methods of machine learning, natural language processing, information retrieval, digital libraries, archiving, visualization, and human-computer interaction," said Fox. To learn more about the grant click here.

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)

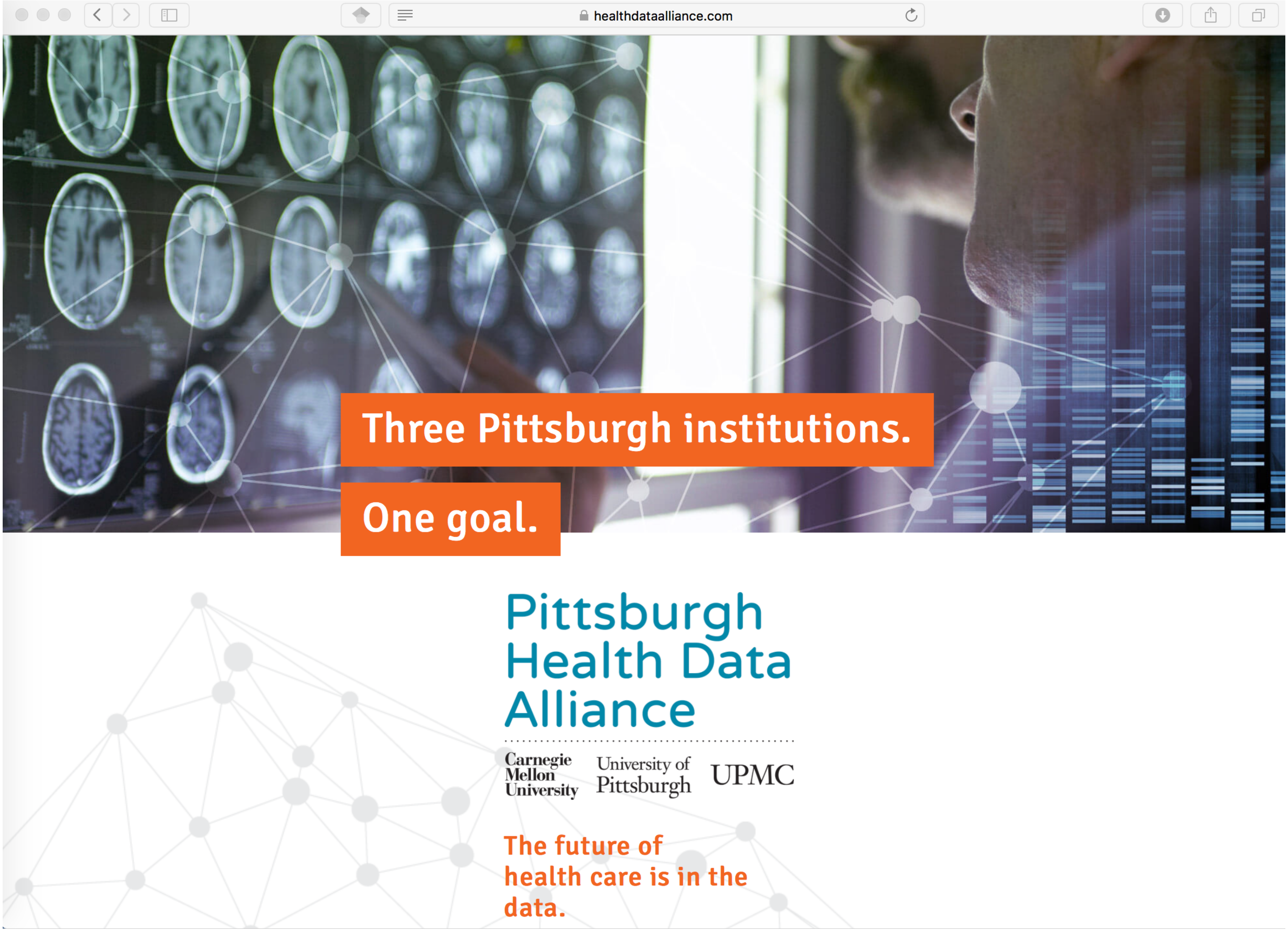
MOODS – Big Data 2.0 – The next frontier

- 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…



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.

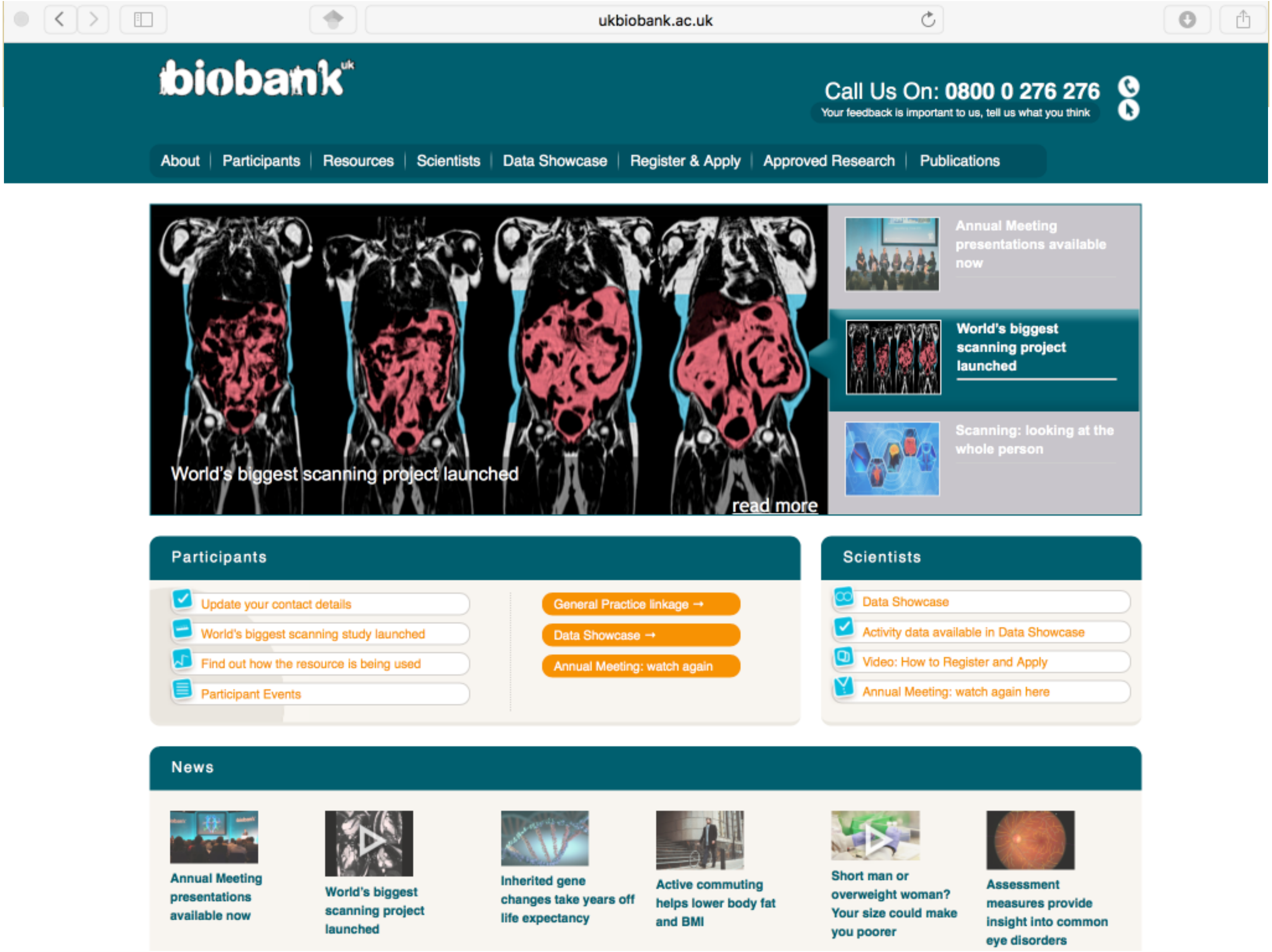
Pittsburgh Health Data Alliance

Carnegie Mellon University | University of Pittsburgh | UPMC

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

What roles can 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."



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

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

World's biggest scanning project launched

read more

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

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"

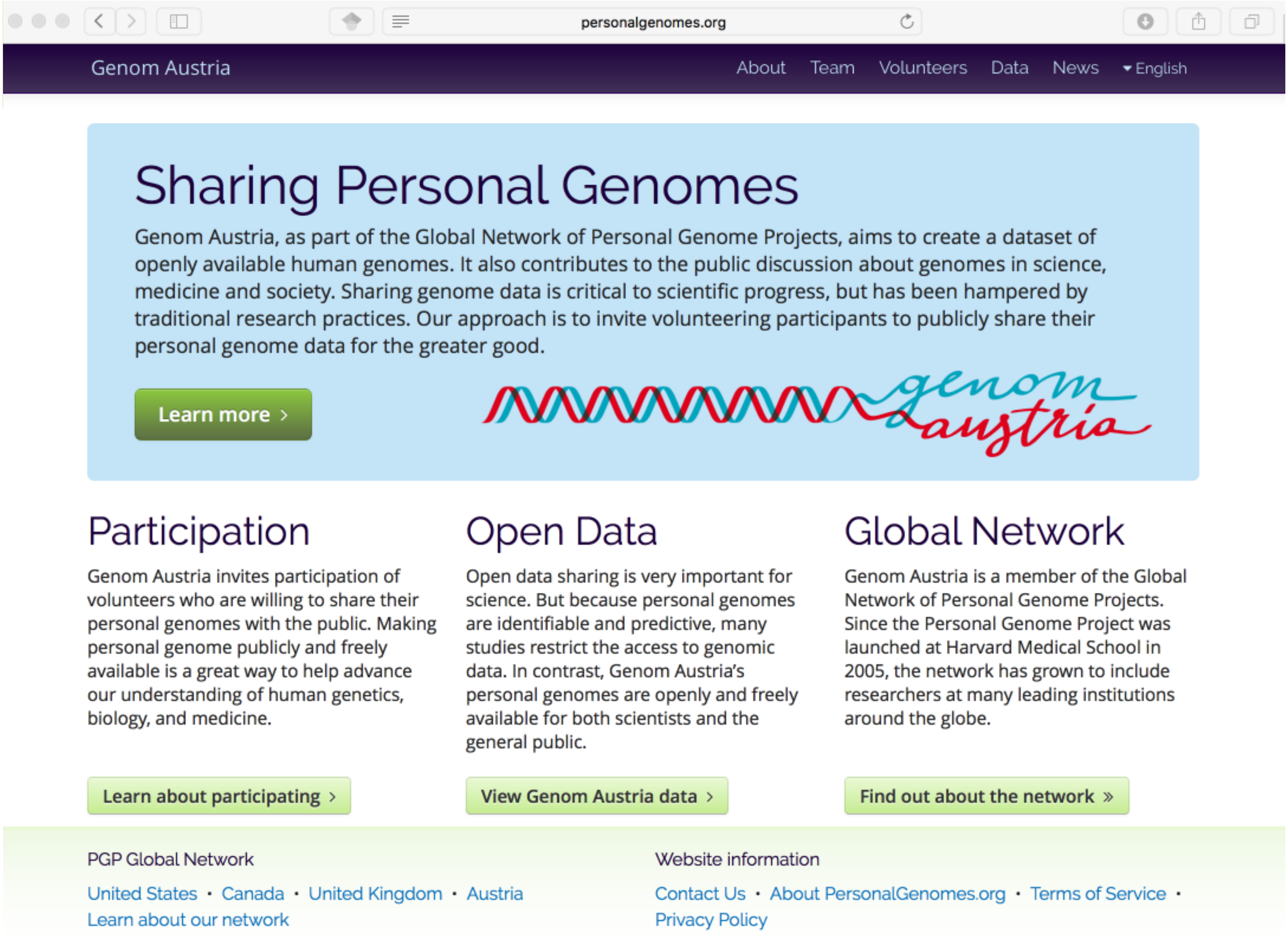
Best **Ethical** Practice?

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?



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.

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.

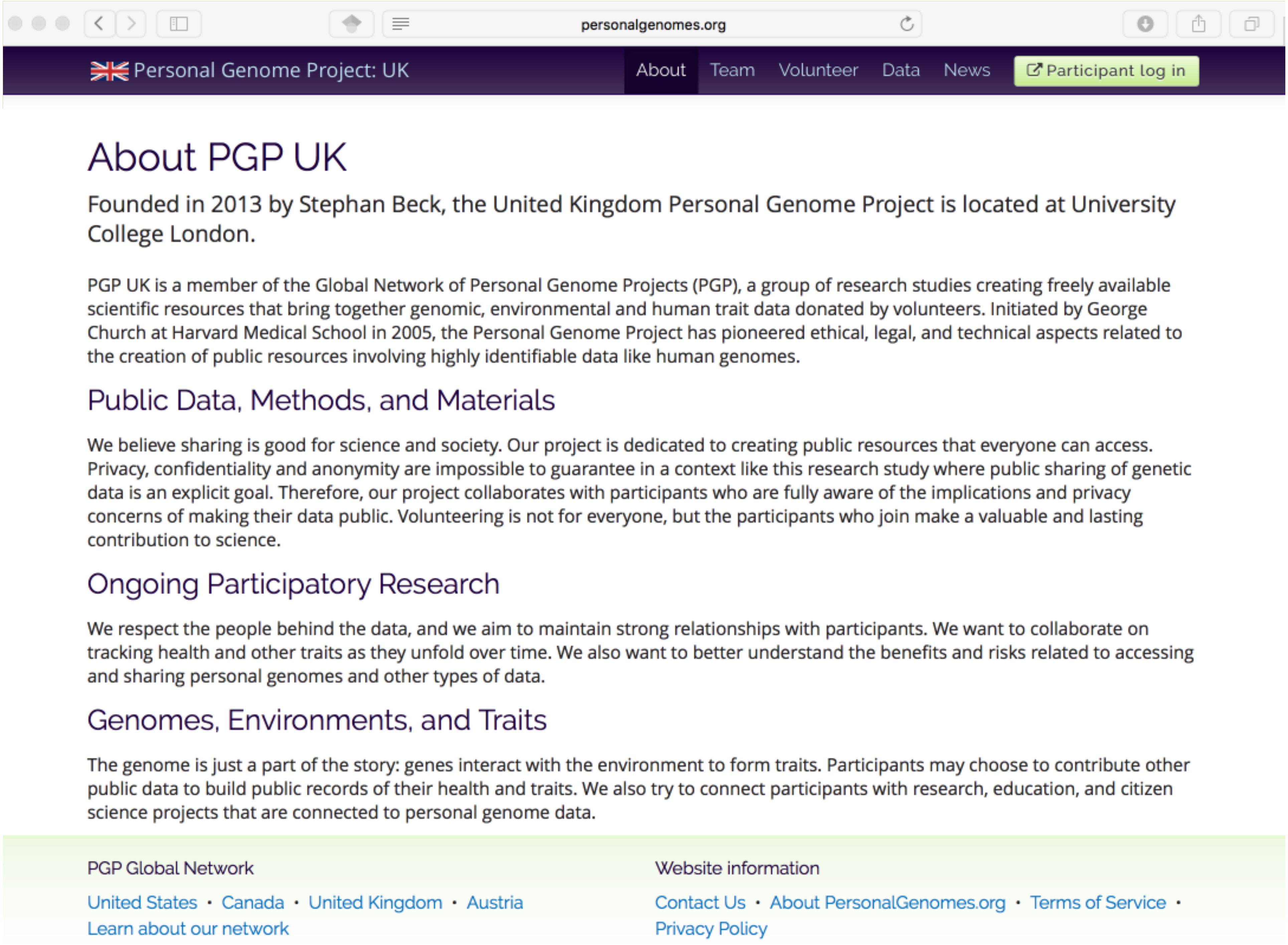
Open Data

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.

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 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

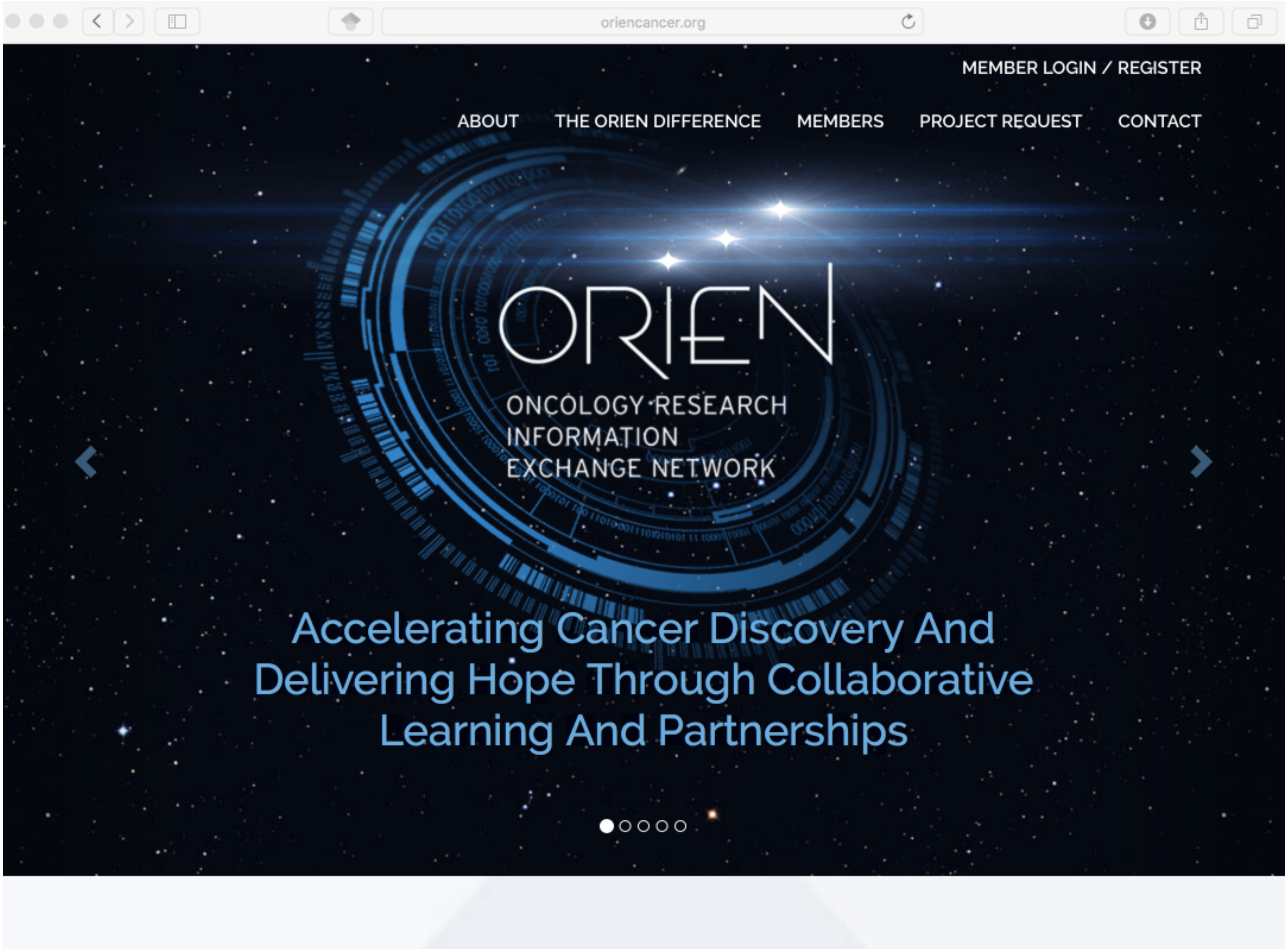
i.e., raw data and not just summary results "where feasible"

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**"

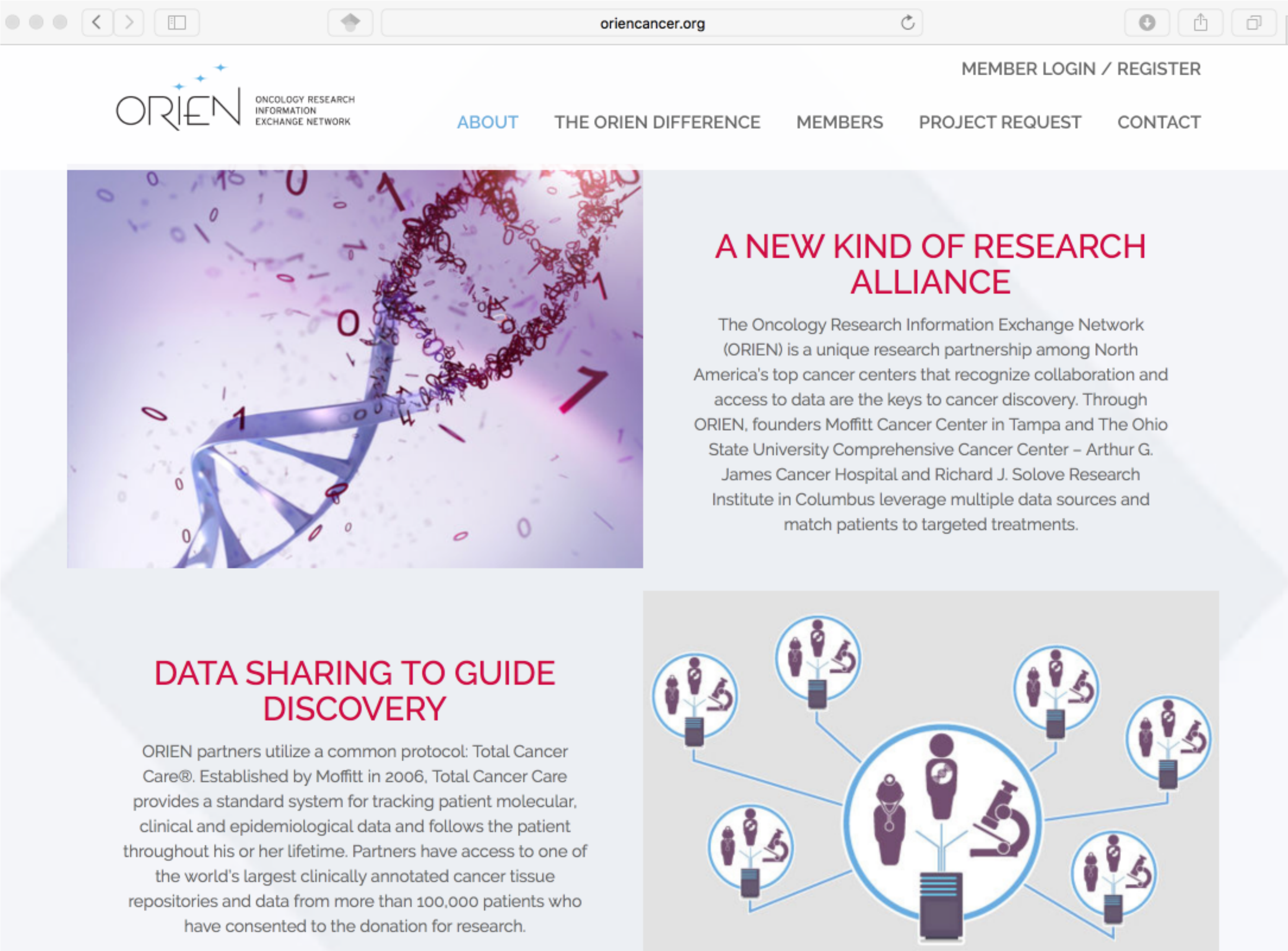
Pretty Good Privacy?



ORIEN

ONCOLOGY RESEARCH
INFORMATION
EXCHANGE NETWORK

Accelerating Cancer Discovery And Delivering Hope Through Collaborative Learning And Partnerships



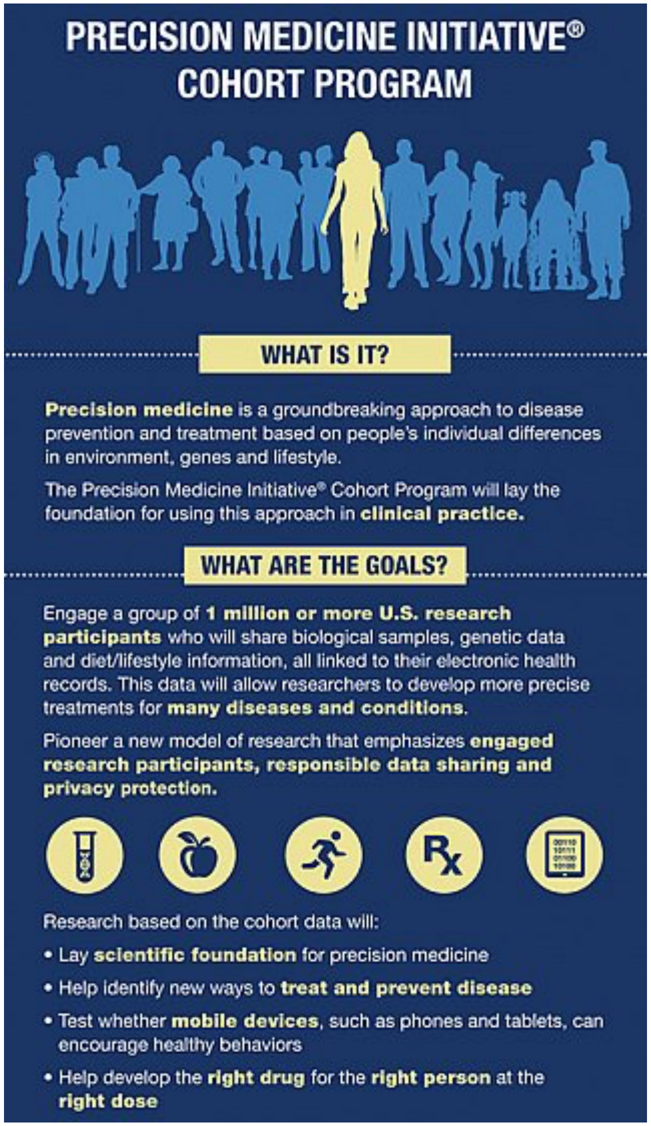
A NEW KIND OF RESEARCH ALLIANCE

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.

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.**

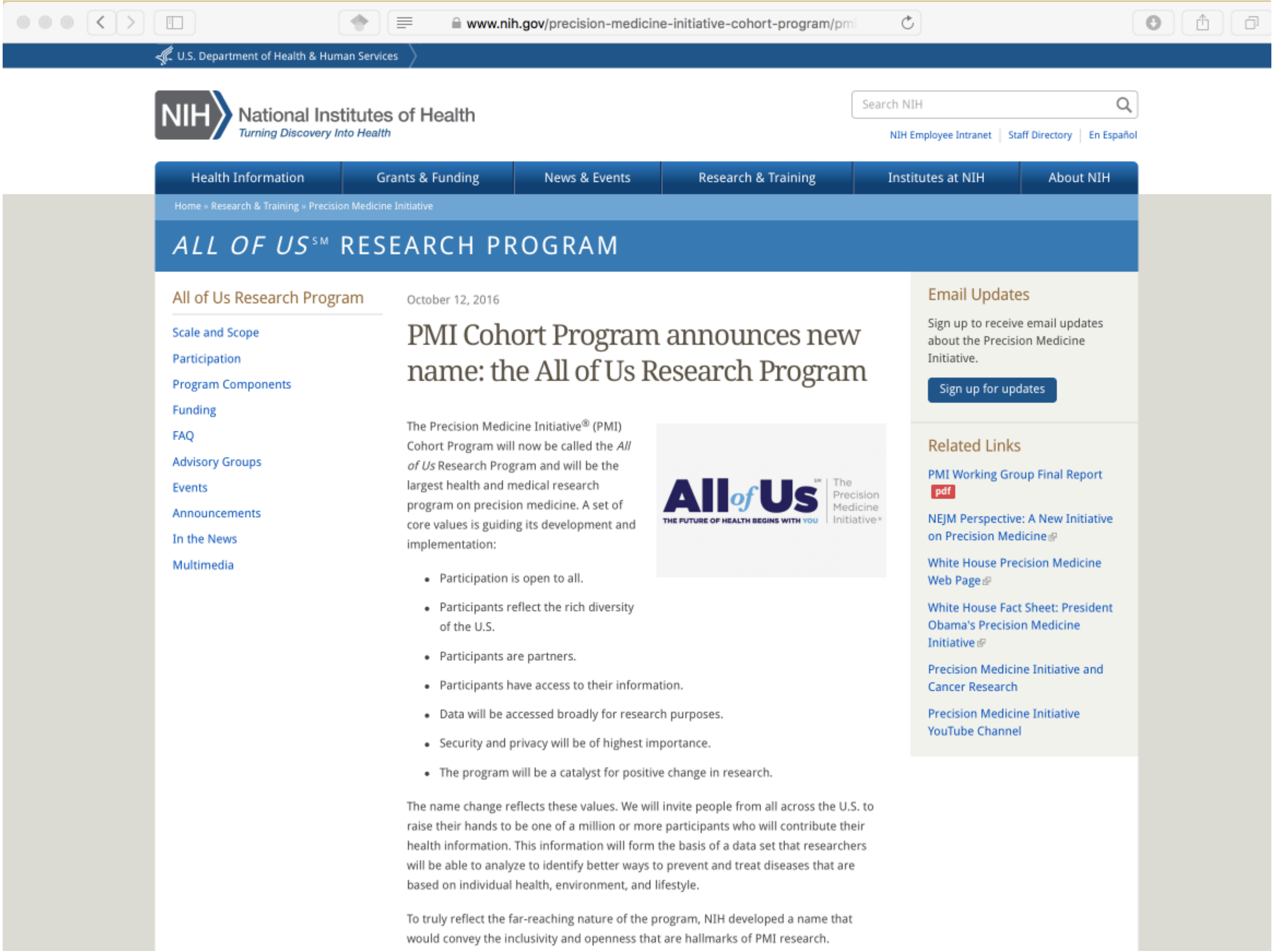
Research based on the cohort data will:

- 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"

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

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®

- 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.

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

Issues Arising from **Big Data 2.0** Projects

Legal compliance

- Privacy laws
- Data protection legislation
- Right to be forgotten
- Genetic information laws
- Freedom of information
- Intellectual property

  e.g., patenting human genes (cf. EU and US case ruling)
- Licensing/contractual issues
- Publishing

Ethical challenges

- Privacy
- Anonymity

  protection from bad actors e.g., cybercriminals, hactivists
- Monetization

  selling of health data
- Conflicts of interest
- Informed consent
- Solicitation of donors for participation in other studies

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

Policy Issues Arising From **Big Data 2.0**

- 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?

Potential Roles for Libraries and Librarians

- Extending **data literacy** programs to include dealing with data in education, in the workplace, and in our personal lives
- Adapting/applying **library expertise** in scholarly communication and digital curation to the evolving Big Data/MOODS arena
- Encouraging stakeholders to pay attention to the **human issues** as well as the technology infrastructure
- Adding **data ethicists** to the "family of data scientist roles"
  - data analyst, data archivist, data engineer, data journalist, data librarian, and data curator/steward

    (Lyon & Brenner, 2015, p. 114)

The Conscience of the Big Data World?

Acknowledgment

The study presented here forms part of an ongoing program of research with Dr. James D. (Kip) Currier, University of Pittsburgh, School of Information Science, Dept. of Information Culture & Data Stewardship, whose expert contribution to the Big Data 2.0 case studies is gratefully acknowledged.

References

Eggers, W. D., Hamill, R., & Ali, A. (2013). Data as the new currency: Government's role in facilitating the exchange. *Deloitte Review*, *13*, 18-31. Retrieved from http://mobile.deloitte.wsj.com/riskandcompliance/files/2013/11/DataCurrency_report.pdf.

Fister, B. (2015, March). Big data or big brother? Data, ethics, and academic libraries. Library Issues, 35(4),

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. Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/349.

Pryor, G. (2012, February 6). *Re-skilling for research – observations on an RLUK report.* Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/news/re-skilling-research-observations-rluk-report.

Rundle, H. (2015, February 1). A measured approach [Blog post]. Retrieved from https://www.hughrundle.net/2015/02/01/measured-approach/.