

CGP

Centre of Genomics and Policy

THE POLICY IMPLICATIONS OF BIG DATA AND COMPUTING IN GENETICS

Professor Bartha Maria Knoppers

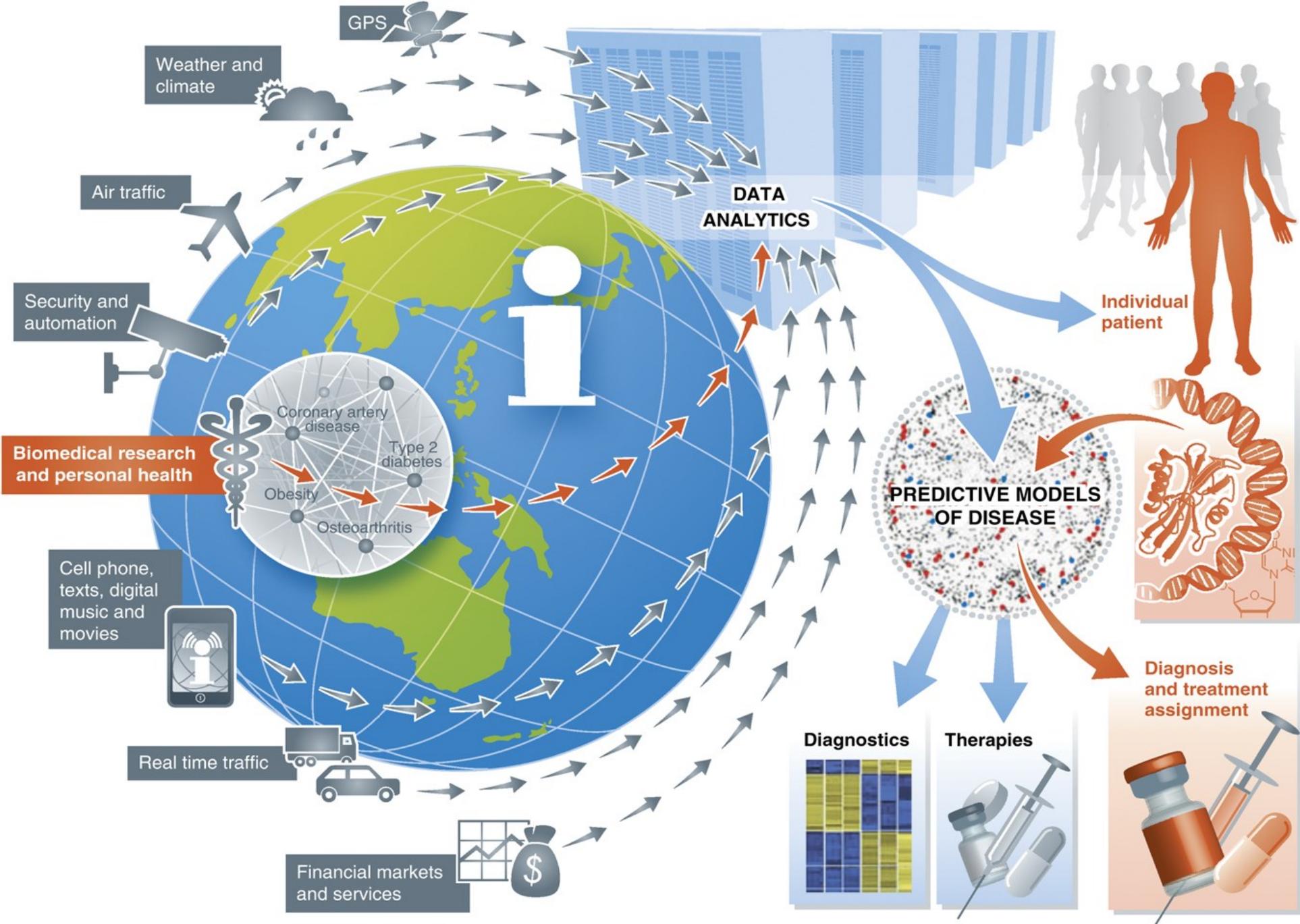
Director, Centre of Genomics and Policy
Canada Research Chair in Law and Medicine
McGill University

Mark Phillips

Academic Associate
Centre of Genomics and Policy

22 June 2016

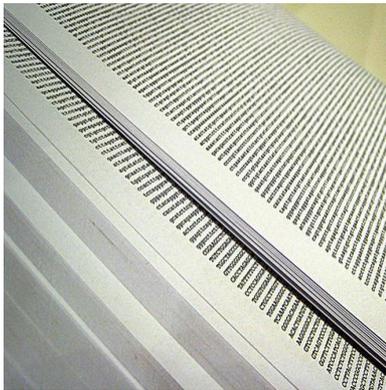




Eric E. Schadt, The Changing Privacy Landscape in the Era of Big Data, *Molecular Systems Biology* 8, 612 (2012).

The Challenge

Unparalleled generation of human genetic data



How do we unlock its health potential?



In a way that allows data to be shared on a global level



Thus empowering new knowledge, new diagnostics and new therapeutics for patients and populations



Big Data Rationale

Broad scope of data sources

Genetic, medical, socio-economic, geospatial, cancer registry, primary care, air pollution, climate, contaminated soils or water, etc.

Associations with new factors, e.g. environmental

Stronger associations between phenotype, genotype

Benefits to public health, but also personalized medicine

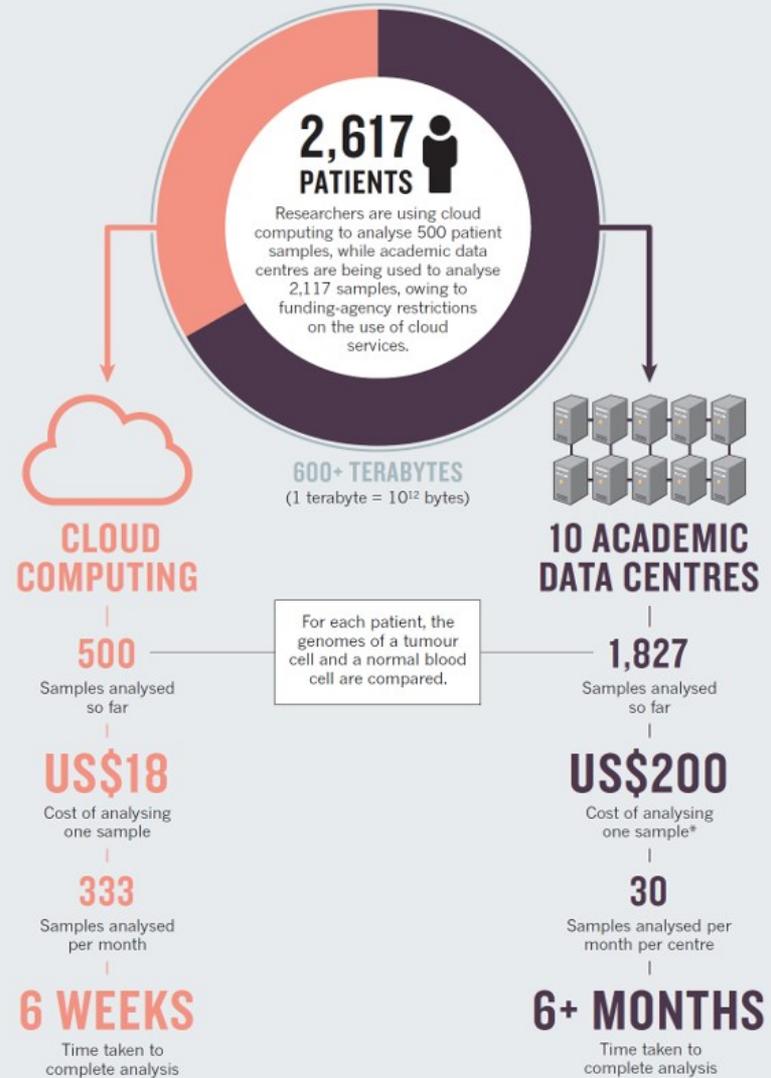
Scientific & Statistical Rationale

What accounts for the rapid rush into the clouds?

Economies of scale are making it difficult for genomic researchers to resist

EXPRESS LANE

The Pan Cancer Analysis of Whole Genomes project (in which L.D.S., P.C., G.G. and J.O.K. are involved), an effort to investigate the role of non-coding parts of the genome in cancer, demonstrates how much faster and cheaper it is to use cloud computing than to use conventional academic data centres when analysing vast biological data sets.



*If using a standard university computer system and buying the hardware.



Big Data and the cloud each pose new potential risks

- Broad linkage can jeopardize social rights, health care, welfare, housing, employment and education and equal treatment
- Loss of data control in the cloud takes various forms, from data integrity and availability issues, legal enforceability of contractual terms, to broad state surveillance practices

These risks must be addressed

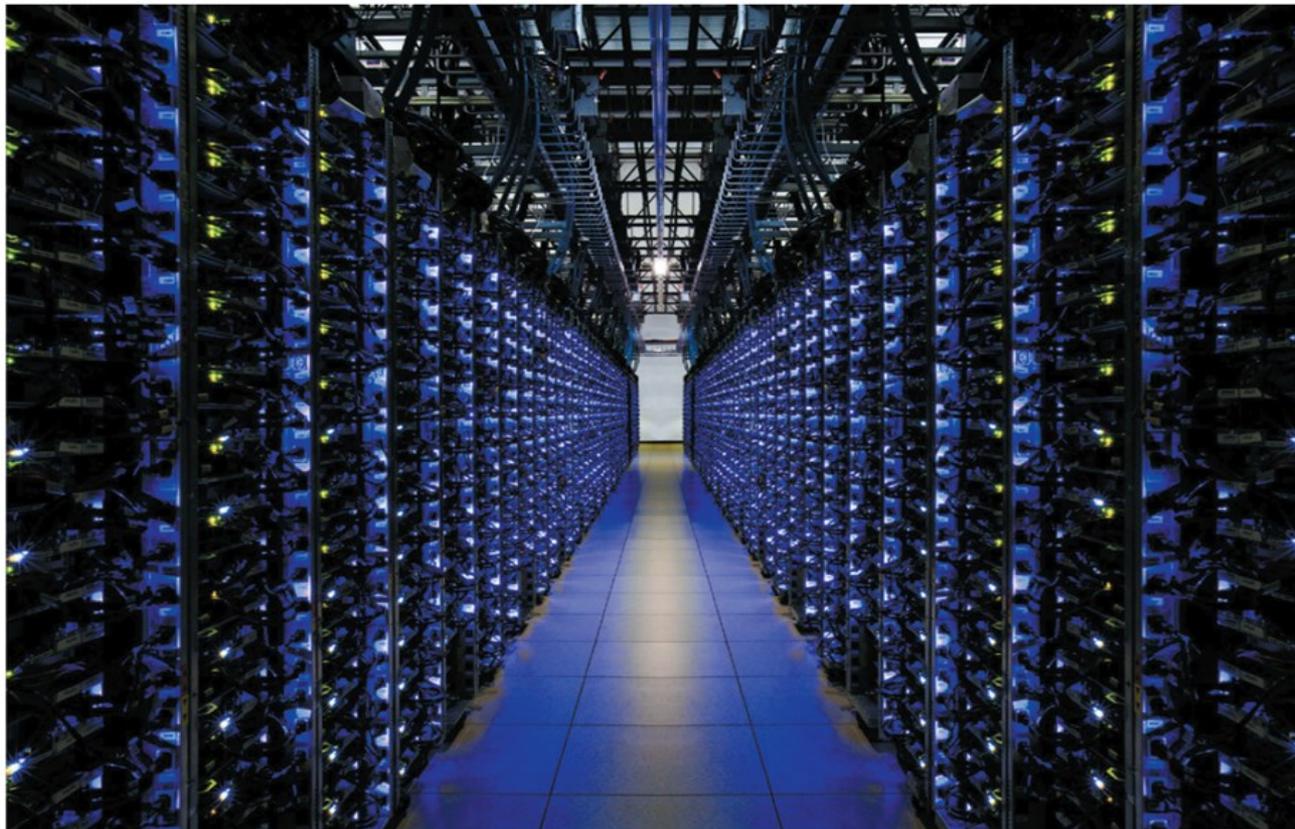
Existing data privacy and other regulation has not been drafted with big data and the cloud in mind

The EU Article 29 Working Party notes:

- a tension between Big Data and privacy limitations on collection, purpose, and processing of data
- Big Data “might require innovative thinking” but there is “no reason to believe that the EU data protection principles ... are no longer valid and appropriate for the development of big data”
- Quid: Safe Harbour / Privacy Shield



EUROPEAN
COMMISSION



Google's cloud services are among those increasingly being used by researchers who want to analyse large genomics data sets.

Create a cloud commons

Major funding agencies should ensure that large biological data sets are stored in cloud services to enable easy access and fast analysis, say **Lincoln D. Stein** and colleagues.

There was a collective cheer in the human genomics community earlier this year, as researchers — ever more stymied by the challenges of accessing vast data sets — saw a major roadblock disappear. In March, the US National Institutes of Health (NIH) lifted its 2007 restriction on the use of cloud computing to store and analyse the tens of thousands of genomes and other genetic information held in its

repository, the database of Genotypes and Phenotypes (dbGaP)¹.

Cloud services offer customers large amounts of storage and computing power on a pay-as-you-go basis. Because these services are available through the Internet, and multiple users share hardware, numerous funding agencies have been concerned that their use in genomics could threaten the privacy of people who supply samples².

The NIH turnaround is part of a growing suite of efforts aimed at addressing the fact that in the human genomics research community, the challenges of accessing big data sets are now blocking scientists' ability to do research, and especially to replicate and build on previous work (see go.nature.com/h9jgs1).

To take full advantage of the possibilities that cloud computing offers, we ►

Enabling a Cloud Commons

- Networked computing forms part of the “intellectual commons”
- Has led to emerging pushes toward a “genomic commons” directed and controlled by researchers themselves
- Allows architectures designed to maximize research benefits and privacy protection

New clouds driven by genomic researchers

- European Open Science Cloud



- Aims to “seamlessly integrate existing networks” run by researchers and universities by the end of 2016
- Aims to provide a unified “virtual environment to store, share and re-use data across disciplines and borders”

- Cancer Genome Collaboratory



- Research cloud allowing analysis of large cancer-genomic datasets by a vast array of research groups
- OICR, MIT, McGill, U Chicago, U Texas, etc., collaboration

- Genomic Data Commons



- US/Chicago project

Addressing genomic cloud privacy and security

- Inter-jurisdictional legal challenges
 - Data localization controls
 - Model contracts and business associate agreements
- Identifiability of participants
 - Robust but not burdensome access restrictions
 - Secure remote computing techniques
- Economies of scale
 - Cloud providers comprehensively address security
 - Reduces burden on academic centers and researchers

IT'S NOT BORING
UP HERE - YOU GET TO
LOOK THROUGH EVERYONE'S
DATA!



Sharing data from the European Union

- Governed by the EU Directive, now the new Regulation
- “Adequacy” remains central authorization mechanism
- Current legal precariousness of sharing to the US:
 - Adequacy of the Safe Harbour scheme retracted in 2015
 - The successor “Privacy Shield” may or may not pass muster
- Alternatives to adequacy decisions
 - Consent — has been actively discouraged by EU regulators
 - Binding Corporate Rules (BCRs) — burdensome except for large institutions
 - Model contracts — often the most compelling alternative
 - Codes of conduct & certifications — a newly opened possibility



Framework for Responsible Sharing of Genomic and Health-Related Data

- Current frameworks are founded on the principle of protection from harm. In contrast,
- GA4GH Framework aims to **activate** the right to science and the right to recognition for scientific production by promoting responsible data sharing.

<http://genomicsandhealth.org/framework>

Universal Declaration of Human Rights, (1948)

“The Right to Science”

27(1)

“Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.”

“The Right to Recognition”

27(2)

“Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.”

Legal Force

- Dual rights rendered legally binding by the International Covenant on Economic, Social and Cultural Rights (ICESCR) (1966) – Article 15.
- 164 States have ratified the ICESCR.
- States are bound to implement the treaty in their national laws.

The Right to Science → Actionable

Why Human Rights?

- Universalizing force
- Political and legal dimensions that reach beyond the moral appeals of bioethics
- International legal force
- Belong to groups as well as individuals (reciprocity)
- Impose positive duties on governments and private actors

Strong Protection in Three Key Areas

Contours of responsible access, sharing, and attribution delineated by other human rights:

1. privacy;
2. anti-discrimination and fair access; and
3. procedural fairness.



The **Clinical Working Group** aims to enable compatible, readily accessible, and scalable approaches for sharing clinical data and linking genomic data. Clinical Working Group strives to address both research and clinical use scenarios and be physician-oriented, researcher-focused, and patient-centered.



The **Data Working Group** concentrates on data representation, storage, and analysis of genomic data, including working with academic and industry leaders to develop approaches that facilitate interoperability. physician-oriented, researcher-focused, and patient-centered.



The **Regulatory and Ethics Working Group** focuses on ethics and the legal and social implications of the Global Alliance, including harmonizing policies and standards, and developing forward-looking consent, privacy procedures, and best-practices in data governance and transparency.



The **Security Working Group** leads the thinking on the technology aspects of data security, user access control, and audit functions, working to develop or adopt standards for data security, privacy protection, and user/owner access control.

Framework for Responsible Sharing of Genomic and Health-Related Data

- Consent Policy
- Consent Clauses and Template for International Data Sharing



- Privacy & Security Policy
- Data Safe Havens
- Accountability Policy



GA4GH Regulatory and Ethics Initiatives



Accountability



Ageing and Dementia



BRCA Challenge
Ethico-Legal



Mobile Health Consent



Data Protection
Regulation



Data Sharing
Lexicon

Ethics Review
Equivalency



Participant Values



Machine Readable Consent



Participant Unique Identifiers

Paediatric



Registered Access



