ACTIVATING THE HUMAN RIGHT TO SCIENCE

Big Data and the responsible sharing of genomic and health-related data

Professor Bartha Maria Knoppers
Director, Centre of Genomics and Policy
Canada Research Chair in Law and Medicine
McGill University
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
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
• Data from **millions of samples** may be needed to achieve results and progress, identifying patterns that would otherwise remain obscure.

• That will take new methods and organizational models.

• Right now:
  • Data is typically in silos: by type, by disease, by country, by institution
  • Analysis methods are non-standardized, few at scale
  • Approaches to regulation, consent and data sharing limit interoperability

• If we don’t act: risk an overwhelming mass of fragmented data, as is the case with electronic medical records in many countries
REACHING FOR THE CLOUD

Internet cloud services, which provide large amounts of data storage and computing power, are becoming increasingly popular with geneticists grappling with vast data sets.

*Data from DNAnexus, a cloud-based genome informatics and data-management platform.
Scientific & Statistical Rationale

What accounts for the rapid rush into the clouds?

Economies of scale are making it difficult for genomic researchers to resist
BEING A CLOUD WAS SUCH A CAREFREE EXISTENCE - NOW PEOPLE EXPECT SO MUCH FROM US!
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
It's not boring up here - you get to look through everyone's data!
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
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 samples\(^2\).

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
• Collaboration between OICR, University of Chicago, McGill University, BC Cancer Agency, MIT, UC San Diego, U Louisville, U Texas
• Development of cloud-based techniques to allow analysis of large cancer genomic datasets by vast array of research groups
• Designed to address two problems
  1. Lack of capacity to download and store the data
  2. Lack of compute power to analyze the data
• Cloud technology can address these computational difficulties
• But privacy and security concerns remain
• The CGC aims to address these through use of
  – Secure remote computing techniques
  – Robust but not burdensome access restrictions
  – Data localization controls
Thanks to the cloud all my devices are in sync... it's only my brain that doesn't know what's going on.
What is the Global Alliance for Genomics and Health?
To accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing.
Role

Convene stakeholders

Catalyze sharing of data

Create harmonized approaches

Act as a clearinghouse

Foster innovation

Commit to responsible data sharing
The Global Alliance does not directly ...

generate or store data
perform research or care for patients
interpret genomes

Membership is open to entities even if they do not hold or share data
Work together **internationally** to ensure **interoperability** of data and of methods, to **harmonize** approaches to ethics and regulation, and to promote **participant** autonomy.

Support pilots, reference implementations and data sharing **projects** that responsibly and effectively harmonize, analyze and share genomic and clinical data.

**Engage** professional communities and the public; build **trust** and encourage appropriate sharing and learning.
Global Alliance members include:

1. Universities and research institutes (31%)
2. Academic medical centers and health systems (11%)
3. Disease advocacy organizations and patient groups (5%)
4. Consortia and professional societies (7%)
5. Funders and agencies (6%)
6. Life science and information technology companies (40%)
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.
Regulatory and Ethics Working Group

Leadership

Bartha Maria Knoppers, McGill University
Kazuto Kato, Osaka University

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.
GA4GH 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
“The Right to Science”

“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”

“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.”
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.
### Framework for Responsible Sharing of Genomic and Health-Related Data

Available in **11** languages on the GA4GH website. Thanks to all who contributed!

<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>إطار لتبادل مسؤول للمعلومات الجينومية والمتصلة بالصحة</td>
</tr>
<tr>
<td>Chinese</td>
<td>基因组学与健康相关数据负责任的共享框架</td>
</tr>
<tr>
<td>French</td>
<td>Cadre pour un partage responsable des données génomiques et des données de santé</td>
</tr>
<tr>
<td>Greek</td>
<td>Πλαίσιο για την Υπεύθυνη Κοινοχρησία Γονιδιωματικών και άλλων Ιατρικών Δεδομένων</td>
</tr>
<tr>
<td>Japanese</td>
<td>ゲノム及び健康関連データの責任ある共有に関する枠組み</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Framework para Compartilhamento Responsável de Dados Genômicos e Relacionados à Saúde</td>
</tr>
<tr>
<td>Spanish</td>
<td>Marco de actuación para el uso compartido responsable de datos genómicos y relativos a la salud</td>
</tr>
<tr>
<td>German</td>
<td>Rahmenkonzept für die den verantwortungsvollen Datenaustausch genomischer und gesundheitsbezogener Daten</td>
</tr>
<tr>
<td>Hindi</td>
<td>जीनोमिकी और स्वास्थ्य संबंधी डेटा को उत्तरदायित्वपूर्ण रूप से साझा करने के लिए रूपरेखा</td>
</tr>
<tr>
<td>Italian</td>
<td>Framework per la condivisione responsabile di dati genomici e relativi alla salute.</td>
</tr>
<tr>
<td>Russian</td>
<td>Рамки для ответственных передаче геномных и связанных со здоровьем данных</td>
</tr>
</tbody>
</table>
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
Aims
• Foster responsible data sharing;
• Protect and promote the welfare, rights, and interests of groups and individuals who donate their data;
• Provide benchmarks for accountability;
• Establish a framework for greater international data sharing cooperation, collaboration, and good governance; and,
• Serve as a dynamic instrument.

Foundational Principles
• Promote Health, Wellbeing and the Fair Distribution of Benefits;
• Respect Individuals, Families and Communities;
• Advance Research and Scientific Knowledge; and,
• Foster Trust, Integrity and Reciprocity.
Framework for Responsible Sharing of Genomic and Health-Related Data

Core Elements for Responsible Data Sharing

- Transparency
- Accountability
- Engagement
- Data Quality and Security
- Privacy, Data Protection and Confidentiality
- Risk-Benefit Analysis
- Recognition and Attribution
- Sustainability
- Education and Training
- Accessibility and Dissemination
Consent Policy

• **Purpose**: To guide international data sharing in a way that respects autonomous decision making while promoting the common good.

• Gives principled and practical guidance on consent issues:
  - Transparency
  - Privacy Safeguards
  - Withdrawal
  - Consent to access policies and terms
  - Sharing of Legacy Data

[https://genomicsandhealth.org/consent-policy-read-online](https://genomicsandhealth.org/consent-policy-read-online)
• How to manage privacy and security risks and related expectations?
• **Proportionate Safeguards:** Data privacy safeguards should be proportionate to the sensitivity, nature, and possible benefits, risks, and uses of the Data
• Security: **organizational, technical and physical measures** to manage risks to privacy and data integrity.

[https://genomicsandhealth.org/privacy-and-security-policy-read-online](https://genomicsandhealth.org/privacy-and-security-policy-read-online)
Accountability Policy

• Sets forth how Members of the Global Alliance can be governed and held accountable for the trust they engender in the responsible sharing of genomic and clinical data.

• Explores mechanisms to promote responsible data sharing, as well as to prevent and appropriately sanction data misuse.

• Possible Inclusion: Why are you NOT sharing, if the participant consented to sharing?
• Explores the regulatory and ethical issues of data sharing in the paediatric context and develops policies, tools and guidelines to accelerate the sharing of such data.

• Initial focus: “Genomic Newborn Screening: Public Health Policy Considerations and Recommendations” [in preparation].
REWG – Current Initiatives

Accountability

Ageing and Dementia

BRCA Challenge Ethico-Legal and Engagement

Data Protection Regulation

Data Sharing Lexicon

Ethics Review Equivalency

Individual Access

Machine Readable Consent

Paediatric

Registered Access
• Myriad’s BRCA patents lost their exclusivity
• Curation of the BRCA variants represents a critically needed resource for the community
• BRCA Challenge: An international non-commercial effort at federated database of variants for BRCA1/BRCA2
NCBI ClinVar database
  – 7536 variants

LOVD Databases
  – 3362 variants

UMD (France)
  – 3913 variants

Goal: Expand to include ENIGMA, CIMBA & many other DBs, as well as new data, to encompass all variant data
Roadmap

• Consensus on terminology & classification
• Broad inclusion of data sources
• Address potential liability concerns
• Single portal for access to several systems available for sharing
• Stable oversight of curation of deposited data
International Policy interoperability and data Access Clearinghouse (IPAC)
International Policy interoperability and data Access Clearinghouse (IPAC)

A one-stop service for researchers

- **Data/Sample Collection – ELSI Interoperability**
  - Validate whether studies can work together (e.g. consent; confidentiality; etc.);
  - If not, create tools.

- **DACO – Review of Data and Samples Access Request/Authorization and Compliance**
  - Authorize studies to access controlled databases (e.g. ethics approval/waiver; institutional sign off; etc.).

- **ELSI Clauses/Agreements Database**
  - Open access resource allowing users to search and select models of clauses that best suit their needs.

- **DataTrust**
  - Support process of re-contacting participants and returning individual-level results in translational research projects.