Secondary Uses of Research Data and Specimens: Legal and Compliance Considerations

CLAIRE TEMPEL
HEALTH PRIVACY OFFICER,
INDIANA UNIVERSITY

MARCIA GONZALES
VP & COMPLIANCE OFFICER,
FRED HUTCHINSON CANCER CENTER

1

What is the secondary use of health data?

World Health Organization

[t]he processing of health data for purposes other than the initial purposes for which the data were collected.

An example of this is when researchers re-process clinical and health insurance data to investigate the cost-effectiveness of a service or product. Secondary use of health data, including data on various determinants of health, provides an important resource for decision-making, health system management and improvement, and research.

With the growing role of artificial intelligence (AI) and big data in health care and the adoption of AI-powered solutions, there is an increasing demand for primary and secondary health data from traditional sources such as electronic health records and non-traditional sources like social media.

WHO Meeting on Secondary Use of Health Data 13 Dec 2022: https://www.who.int/europe/news-room/events/item/2022/12/13/default-calendar/meeting-on-secondary-use-of-health-data

2

What is the secondary use of health data?

American Medical Informatics Association

"non-direct care use of PHI [personal health information] including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities".

Can be distinguished between four types of secondary use of clinical data:

- research,
- improving quality and safety of care,
- informing financial management, and
- education.

Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment – *Journal of Medical Internet Research* https://www.jmir.org/2021/6/e26631/PDF

3

Secondary Uses - examples

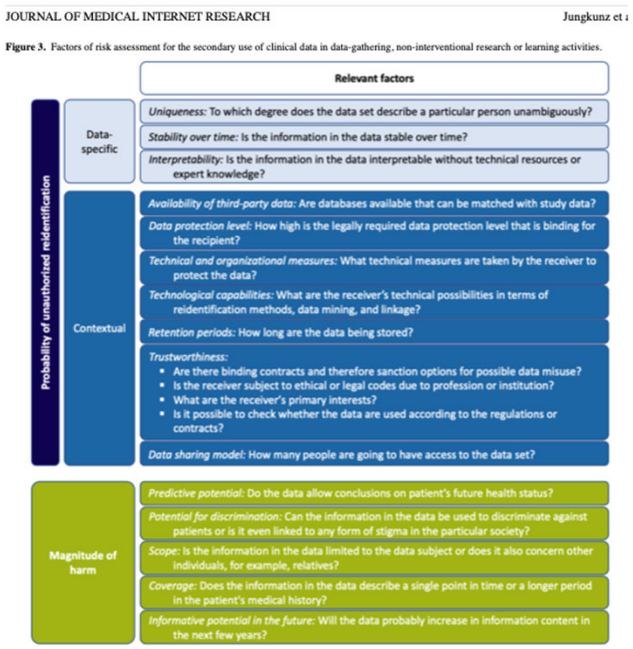
Table 1. Possible types of research or learning activities within secondary use of clinical data in data-gathering, non-interventional research or learning activities.

Type of research or learning activities	Object of investigation	Area of application
Improvement of infection control	Clinical unit	Quality control and improvement
Early detection of possible hazards from germs	Clinical unit	Quality control and improvement
Public health surveillance	General public	Public health research
Epidemiology	General public	Public health research
Outcomes research	Patients or clinical unit	Public health research or quality control and improvement
Health services research	General public	Public health research
Register studies	General public	Public health research
In-silico hypothesis testing	Patients	Non-interventional (observational) clinical research
Comparative effectiveness research	Patients	Non-interventional (observational) clinical research
Experimental therapy evaluation	Patients	Non-interventional (observational) clinical research
Drug safety and efficacy studies	Patients	Non-interventional (observational) clinical research
Studies on risk factors	Patients	Non-interventional (observational) clinical research
Medical informatics research	Patients, clinical unit, or general public	Possible in all three areas of application
Explorative use	Patients, clinical unit, or general public	Possible in all three areas of application

Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment – *Journal of Medical Internet Research* https://www.jmir.org/2021/6/e26631/PDF

4

Risk Assessment of Secondary Uses



5

Regulatory Obligations - The Usual Suspects

- NIH
- HIPAA
- GDPR
- State law
- Agreements and other documents

6

NIH requirements - examples

NIH Institute and Center Data Sharing Policies

- https://sharing.nih.gov/other-sharing-policies/nih-institute-and-center-data-sharing-policies

Consortium agreements:

- A provision addressing ownership and disposition of data produced under the consortium agreement. This includes whether cell lines, samples or other resources will be freely available to other investigators in the scientific community or will be provided to particular investigators only.

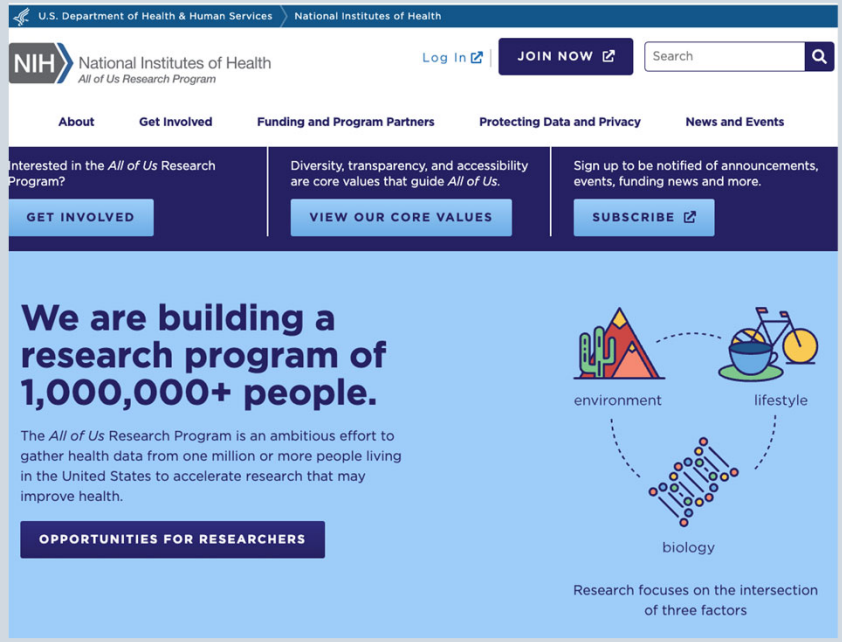
- A provision making NIH data sharing and inventions and patent policy, including a requirement to report inventions to the recipient (see Administrative Requirements-Availability of Research Results: Publications, Intellectual Property Rights, and Sharing Research Resources in IIA), applicable to each consortium participant and its employees in order to ensure that the rights of the parties to the consortium agreement are protected and that the recipient can fulfill its responsibilities to NIH. https://grants.nih.gov/grants/policy/nihgps/html5/section_15/15.2_administrative_and_other_requirements.htm#Written

7

NIH – All of Us Program

Direct volunteer pathway – individuals sign up and donate data for research

https://allofus.nih.gov/



8

ELECTRONIC HEALTH RECORDS

All participants are invited to share their electronic health records (EHR) with the program. We use the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to standardize all EHR data. Learn more about the OMOP CDM on the Data Methods page.

BIOSAMPLES AND BIOASSAYS (GENOMICS)

Most participants contribute blood, saliva, and/or urine samples that can be used for a variety of bioassays. DNA extracted from samples is sent to our genome centers for genomic analysis, including whole genome sequencing and genome-wide genotyping. In addition, a select subset of participant blood serum samples were analyzed for COVID-19 antibody testing in 2020.

SURVEYS

All participants are invited to complete surveys periodically. Current surveys include questions about participants' identities and backgrounds, overall health, lifestyles, medical histories, healthcare access, experiences with COVID-19, and more. Survey response data can be found in the Data Browser and question source information can be found in the Survey Explorer. In addition, some external survey data are available in the Researcher Workbench and can be linked with *All of Us* data.

PHYSICAL MEASUREMENTS

Trained staff measure participants during initial appointments at *All of Us* Research Program partner organizations. Data from these visits include measurements of height, weight, waist circumference, hip circumference, blood pressure, and heart rate. Program staff also note participants' pregnancy status and wheelchair use at any visits.

WEARABLE DEVICES (DIGITAL HEALTH)

All of Us participants can contribute information from their wearable health tracking devices, such as Fitbit devices, which track biometric data including heart rate, physical activity, and sleep. Data from additional wearable devices may be available in the future.

Data Roadmap

In the coming years, *All of Us* will enroll more participants and make more types of data available. Data availability and access timelines are estimates and subject to change.

	Available end of 2022	Early 2023 additions	2023 and beyond
Electronic Health Records (EHRs)	• EHRs	→	
Surveys	• The Basics • Overall Health • Lifestyle • Health Care Access and Utilization • Personal and Family Medical History • Social Determinants of Health • COVID-19 Participant Experience (COPE) • COVID-19 Vaccines Minute Surveys 1-3	• Combined Personal and Family Health History • COVID-19 Vaccines Minute Survey 4	• Life Functioning
Physical Measurements	• Physical Measurements	→	• Participant reported height and weight
Genomics	• 165K+ genotyping arrays • 98K+ whole genome sequences (WGS)	• 245K+ WGS • 312K+ Arrays • CRAM Files • 1,000 long read sequences	→
Data Linkages	• American Community Survey (ACS) 3-digit zip code	→	• ACS 5-digit zip code
Digital Health Technologies	**Fitbit:** • Heart rate by zone summary and minute-level • Activity (daily summary) • Activity intraday steps (minute-level)	• Fitbit sleep data	• Fitbit device data • Vital measurements • Apple HealthKit activity data
Assays	• COVID-19 serology data	→	
Ancillary Studies	→		• Exploring the Mind Pilot • Nutrition for Precision Health (NPH) modules

9

HIPAA and the Common Rule

https://aacrjournals.org/cancerres/article/83/8/1175/725098/Challenges-to-Using-Big-Data-in-CancerChallenges

Table 1. Comparison of HIPAA and the Common Rule.

Topic	HIPAA	Common rule	Distinctions
Overview	The HIPAA Privacy Rule requires any entity that provides or processes healthcare data to obtain consent before disclosing a person's medical history to another entity. These covered entities may disclose medical information without consent if needed for obtaining treatment, payment, or healthcare operations, or if in the public interest. [45 CFR Part 160 and Part 164 Subparts A and E] The HIPAA Security Rule protects all individually identifiable health information a covered entity creates, receives, maintains, or transmits in electronic form. [45 CFR Subpart Part 160 and Part 164 Subparts A and C]	The Common Rule provides protection for human subjects in research conducted or supported by most federal departments and agencies. It is the baseline standard of ethics by which any government-funded research in the US is held; nearly all academic institutions hold their researchers to these statements of rights regardless of funding. Consideration is given to how various aspects of research projects (including privacy, confidentiality, data collection, data maintenance, and data retention) impact physical, emotional, financial, and informational harms. The policy established IRBs to help review and ensure compliance with the policy and the requirements for informed consent. [45 CFR 46]	HIPAA provides protection for all PHI accessed by entities who provide or process healthcare. The Common Rule provides protection for human subjects in research settings.
De-identified data	De-identified data (per HIPAA standards) not subject to further HIPAA requirements. [45 CFR 164.514 (a)]	Data that are "not identifiable" are not subject to the Common Rule. [45 CFR 46.102(e) (1) & (5)]	Although the Common Rule does not explicitly reference the HIPAA de-identified data standards, IRBs have been known to rely on it in determining if research is not subject to the Common Rule.
Limited data	A limited data set can be used or disclosed for research without the need for prior consent of participants if a data use agreement is executed, setting forth research purposes and prohibiting re-identification. [45 CFR 164.514(e)]	An IRB may determine that a study is not "human subjects research" or, in the case of secondary data research, consider the study to be "exempt" because the identity of participants is not "readily ascertainable" (researchers must agree not to re-identify or contact participants). [45 CFR 46.104(d) (4)(ii)]	Common Rule exemption is limited to secondary data (a HIPAA limited data set may involve data initially collected for research purposes). The Common Rule requires limited IRB review for determination of exemption.
Secondary use	Entities have the option to broadly consent individuals to future research uses of their identifiable PHI. [Federal Register vol. 78, no. 17, page 5612 (January 25, 2013)]	Entities may create secondary research databases with broad consent, subject to limited IRB review. [45 CFR 46.104(d) (7) & (8)]	Common Rule exemption is limited to secondary data (HIPAA broad consent provisions could involve data initially collected for research purposes). The Common Rule requires limited IRB review for determination of exemption. Broad consent under the Common Rule is subject to specific requirements.
Use of PHI	PHI used in research; when broad consent is not applicable, full HIPAA authorization is required (unless altered or waived; see below). [45 CFR 164.512(i)]	If governed by HIPAA, Common Rule exemption is available [45 CFR 46.104 (d) (4)(iii)]; otherwise, identifiable information used or disclosed in research—when not subject to any of the above exemptions—requires full IRB review and full informed consent (unless waived; see below). [45 CFR 46.106 & 46.116]	The Common Rule allows entities to rely on HIPAA where it applies; otherwise, the Common Rule requires IRB review.
Waiver of consent	Waiver of HIPAA authorization by IRB or Privacy Board. [45 CFR 164.512(i) (2)(ii)]	Waiver of informed consent by IRB. [45 CFR 46.116(f)]	Waiver criteria similar; the Common Rule also requires consideration of welfare and rights of participants and whether research is minimal risk.

Abbreviations: CFR, Code of Federal Regulations; PHI, protected health information.

10

International Considerations - GDPR

Does not mention "secondary use" bur refers to "further processing" of personal data. Based on Recital 50 of the Regulation, further processing refers to "processing of personal data for purposes other than those for which the personal data were initially collected" (Recital 50).

Requires data processors to carry out an "assessment of the impact of the envisaged processing operations on the protection of personal data" where there is a high risk to the "rights and freedoms of natural persons" (Article 35, 1).

"personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". Exceptions: where this general prohibition of incompatible further processing does not apply are when the further processing is based on: i) the data subject's consent; or ii) a "Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard, in particular, important objectives of general public interest" (Article 6(4); Recital 50).

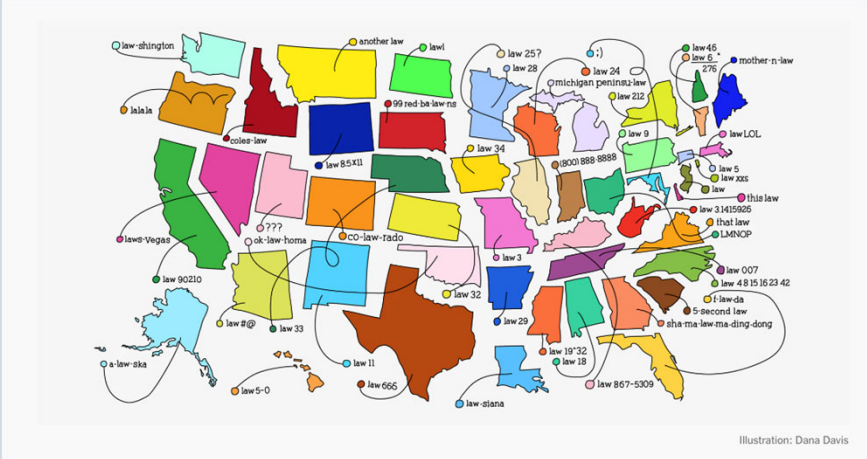
See - Secondary Use of Personal Health Data: When Is It "Further Processing" Under the GDPR, and What Are the Implications for Data Controllers? European Journal of Health Law https://brill.com/view/journals/ejhl/30/2/article-p129_1.xml

11

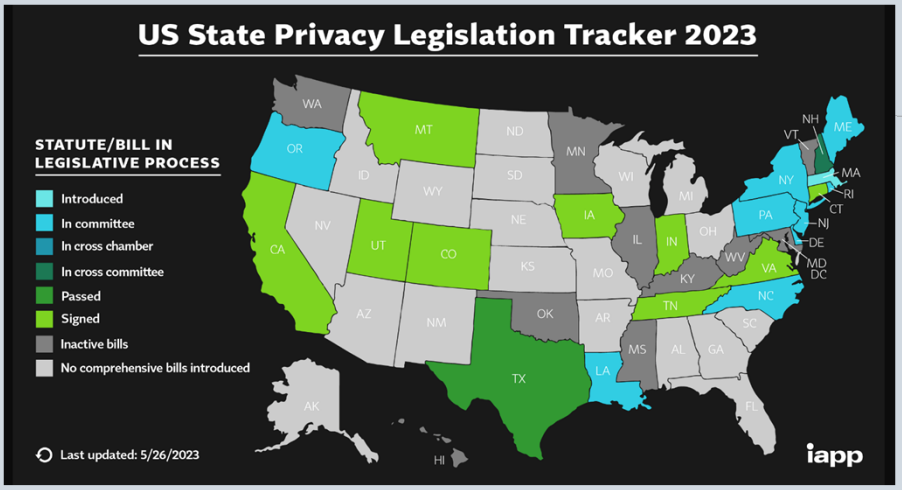
Other considerations . . .

12

Emerging State Statutes



13



State Law

Washington – My Health My Data Act

https://www.ncsl.org/technology-and-communication/2023-consumer-data-privacy-legislation

IAPP State Privacy Legislation Tracker 2023 Chart - https://iapp.org/media/pdf/resource_center/State_Comp_Privacy_Law_Chart.pdf

14

Regulatory Obligations - manifested

Mechanisms to Govern Responsible Sharing of Open Data: A Progress Report

https://sage-bionetworks.github.io/governanceGreenPaper/manuscript.pdf

Table 3: Data collection mechanisms

Design Pattern	Description	Example
Clinical protocol	A written plan for how a health condition, drug, or device is to be studied and the procedure to be followed by the study, including technology. Connected to international and national laws in bioethics.	CDISC Clinical Trial Protocol Representation Model,[36] NIH Protocol Templates for Clinical Trials[37]
Informed consent	Standard language and user interfaces that have been vetted through legal and ethical review for specific types of data, e.g. health records or DNA	Sage Bionetworks Elements of Informed Consent,[38] GA4GH Model Consent Clauses[24]
Privacy policy	Standard language and user interfaces that have been vetted through legal and ethical review for apps and websites that collect user data	Sage Bionetworks Privacy Toolkit[39]
Terms of service	Standard language and user interfaces that have been vetted through legal review for apps and websites that collect user data	Various auto-contract systems e.g. Formswift,[40] LegalZoom[41]

15

Authority and Oversight

- Who decides if the data and biospecimens are identifiable?
- Who decides where and with who data will be shared?
- Who decides if the data should be shared?
- Who reviews and approves contract language for data sharing and use?

16

Internal Stakeholders

- IRB / HRPP
- Research Contracting
- Legal
- Privacy Officer
- Data Administration / Data Stewards
- Procurement

17

Data Governance

- Executive Management
- Data Administrative Council / Data Commons
- Data Steward
- Data Manager
- IT Professionals / Management
- Researcher / User
- Data Classification

18

Data Governance

Board of Trustees

The Board of Trustees is Indiana University's governing board, legal owner, and final authority. For the purposes of information security and privacy governance, the Board is the owner of all information, except information excluded from university ownership as set forth in IU Policy UA-05.

Executive Management

IU's Executive Management includes those individuals assigned executive management responsibilities, typically with the titles of President, Vice President, and Chancellor. This group also includes the Academic Deans.

University Data Management Council

The purpose of the University Data Management Council is to provide university-wide strategic planning, governance, and oversight for Indiana University's institutional data. The UDMC consists of eleven permanent members, including Data Stewards, the Chief Privacy Officer, and the UIPO.

Data Stewards

Individually, each Data Steward has management and policy-making responsibilities for their specific data subject areas. For the purposes of information security and privacy governance, the title of Data Steward describes an individual with a role title related to university-wide information governance purposes.

Data Managers

Data Managers are assigned responsibilities to receive, evaluate, and authorize or deny requests for access to systems, applications, and/or databases containing institutional information. They are assigned by their respective Data Steward to govern the daily operations of their data domain.

IT Management

Includes those individuals assigned technology management/director responsibilities for a unit or service.

IT Professionals

Includes those individuals assigned responsibilities for developing and implementing technology infrastructure, systems, and applications to support functional needs upon direction from IT Management. For a directory of the IT People for each department, see the IT People Database (IU Login required).

User

Refers to any individual who interacts with institutional information.

19

Data Classification

Indiana University

There are four classification levels of institutional data at Indiana University.

Institutional Data is categorized into data classifications as defined in IT Policy DM-01: Management of Institutional Data to ensure proper handling and sharing of data based on sensitivity and criticality of the information. Data classifications are listed below from most sensitive to least sensitive:

Critical

Inappropriate handling of this data could result in criminal or civil penalties, identity theft, personal financial loss, invasion of privacy, and/or unauthorized access to this type of information by an individual or many individuals.

Restricted

Because of legal, ethical, or other constraints, this data may not be accessed without specific authorization. Only selective access may be granted.

University-Internal

This data may be accessed by eligible employees and designated appointees of the university for purposes of university business. Access restrictions should be applied accordingly.

Public

Few restrictions are placed on this data, as it is generally releasable to a member of the public upon request. Upon receipt of a request, seek advice from the appropriate data steward. If the request is made in regard to the Indiana open records statute, seek advice from the Office of the VP and General Counsel, as well as the appropriate Data Steward.

20

Governance Structures

Mechanisms to Govern Responsible Sharing of Open Data: A Progress Report

https://sage-bionetworks.github.io/governanceGreenPaper/manuscript.pdf

Table 1: Governance structures and their attributes

Governance structure	Number and linkage of parties	Degree of data Availability	Degree of freedom to use data	Challenges common to the governance success	Primary governance design pattern
Pairwise	One-to-one	Medium/High	Medium/High	Uneven status of parties, value of data	Informal or closed contract
Open Source	One/some-to-many	High	High	Rights permanently granted to user	License
Federated Query	Many-to-many, via platform	High	Medium/Low	Defection of creators	Contract and club rules
Trusted Research Environment	One/some-to-many	Medium/Low	Medium/Low	Users agree to be known, surveilled	Data transfer and use agreements
Model-to-Data	One-to-many	High	Low	Not all who apply can use data	Restricted analyses, data curation
Open Citizen Science	Many-to-many	High	High	Capacity for analysis is uneven	Contract or license
Clubs, Trusts	Some-to-some	Medium/Low	High	Easy to create things governed more liberally. Trusteeship can be revoked.	Club / Trust rules
Closed	Many (to none)	Low	High	Fundamental limits to collaboration	Public laws, security protocols
Closed and Restricted	Some (to none)	Low	Low	Fundamental limits to collaboration	Public laws, security protocols

21

Governance Structures

Mechanisms to Govern Responsible Sharing of Open Data: A Progress Report

https://sage-bionetworks.github.io/governanceGreenPaper/manuscript.pdf

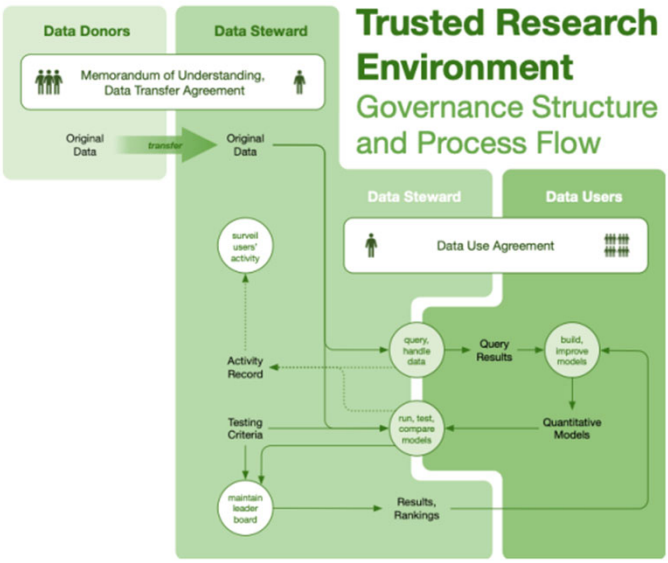


Figure 6: Governance structure and process flow for trusted researcer environment

22

Questions?

23