

Sharing data under the 21st Century Cures Act

Mary A. Majumder, JD, PhD¹, Christi J. Guerrini, JD, MPH¹, Juli M. Bollinger, MS¹,
Robert Cook-Deegan, MD² and Amy L. McGuire, JD, PhD¹

On 13 December 2016, President Obama signed the 21st Century Cures Act (“the Act”) into law. Many of its provisions support the creation of an “Information Commons,” an ecosystem of separate but interconnected initiatives that facilitate open and responsible sharing of genomic and other data for research and clinical purposes. For example, the Act supports the National Institutes of Health in mandating data sharing, provides funding and guidance for the large national cohort program now known as All of Us, expresses congressional support for a global pediatric study network, and strengthens patient access to health information. The Act also addresses potential barriers to data sharing. For example, it makes the issuance of certificates of confidentiality automatic for federally funded research involving

“identifiable, sensitive” information and strengthens the associated protections. At the same time, the Act exacerbates or neglects several challenges, for example, increasing complexity by adding a new definition of “identifiable” and failing to address the financial sustainability of data sharing and the scope of commercialization. In sum, the Act is a positive step, yet there is still much work to be done before the goals of broad data sharing and utilization can be achieved.

Genet Med advance online publication 25 May 2017

Key Words: 21st Century Cures Act; certificates of confidentiality; data sharing; HIPAA; privacy

INTRODUCTION

On 13 December 2016, President Obama signed the 21st Century Cures Act (“the Act”) into law.¹ As the title of the Act suggests, its overarching goal is to accelerate development of treatments through investment and changes to the policy environment surrounding discovery, drug and device development, and health-care delivery. The omnibus 996-page law works to achieve these goals through numerous, diverse provisions.

Many of those provisions are directed at promoting data sharing and, in our view, support the creation of an “Information Commons,” which the National Research Council recognized as critical to improving the health of individuals and communities.² The Information Commons can be understood as a robust ecosystem of separate but interconnected initiatives that facilitate open and responsible sharing of genomic and other data for use in research and clinical practice. The American College of Medical Genetics and Genomics recently affirmed the value of sharing data for both these uses.³

We analyze the Act through the lens of its impact on data sharing and the creation of an Information Commons. While our assessments of the Act’s data-sharing provisions are generally positive, the legislation exacerbates some existing concerns and leaves several challenges unresolved, raising important questions related to, among other things, commercialization and the identifiability of research participants

(see **Table 1**). The Act is a positive step toward the creation of an Information Commons, yet there is still much work to be done before the goals of broad data sharing and utilization can be achieved.

CREATING AN INFORMATION COMMONS

Four features of the Act are particularly notable for their potential to enhance the Information Commons: (i) supporting the National Institutes of Health (NIH) in mandating data sharing, (ii) promoting the assembly of a representative national cohort in the United States (the All of Us research program), (iii) encouraging global data sharing through a pediatric clinical study network, and (iv) strengthening patient access to information.

Pushing for open and responsible data sharing

First, Section 2014 of the Act authorizes the NIH Director to require award recipients to share data in a manner consistent with applicable federal laws and regulations. The NIH has already adopted policies mandating data sharing, for example, through its Genomic Data Sharing Policy.⁴ Thus, the Act is an endorsement of existing NIH policies and a statutory basis for expanding their scope. It is also a basis for stepped-up enforcement of these policies. However, it does not address difficult questions about funding. In an environment in which investigators are frequently asked to slash budgets, data-sharing mandates that impose additional costs (e.g., for

¹Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas, USA; ²School for the Future of Innovation in Society, Arizona State University, and FasterCures, a Center of the Milken Institute, Washington, DC, USA. Correspondence: Mary A. Majumder (majumder@bcm.edu)
The last two authors contributed equally to this work.

Submitted 8 January 2017; accepted 3 April 2017; advance online publication 25 May 2017. doi:10.1038/gim.2017.59

Table 1 21st Century Cures Act provisions relevant to data sharing

Topic	What the Act does	Potential impact on data sharing	Complications
Large-scale NIH research initiatives	Authorizes up to a total of \$4.8 billion for these initiatives (§ 1001)	Supports rapid advance of cohort study and formation and growth of brain- and cancer-focused data commons	Funding remains contingent on annual appropriations by Congress
PMI cohort study (All of Us)	Imposes high-level requirements related to privacy and security protection, ensuring diversity, securing trust (via access policies), and transparency (§ 2011)	Possible source of innovative policies and practices in these areas and model for other initiatives	Requirements may not be effectively operationalized
Research subject privacy	Expands reach and strengthens protections under certificates of confidentiality (§ 2012); authorizes FOIA exemption for biomedical information; low threshold for determination that information is identifiable, triggering protections (§ 2013)	Reduces risks to research participants related to disclosure of information collected or compiled in the course of research	Potential to reinforce siloing of research and clinical care; increased complexity due to adoption of identifiability definition that differs from other regulatory definitions
NIH promotion of data sharing	Authorizes NIH Director to require award recipients to share data in a manner consistent with applicable laws and regulations (§ 2014)	Endorsement of existing NIH policies mandating data sharing and basis for expansion and for stepped-up enforcement	Balancing commitment to data sharing with protection of trade secrets and commercial information; does not address financial sustainability issue
HIPAA clarifications and revisions	Directs Secretary to issue clarifying guidance on remote access for reviews preparatory to research and authorizations for future research use/disclosure of PHI; sends controversies to working group (§ 2063)	Reduces burden on researchers conducting preparatory reviews and likely to advance broad consent paradigm, which should facilitate data sharing	Defers consideration of important questions related to health data-based research and markets in health data (e.g., whether to permit sale of PHI to researchers without patient authorization)
Global pediatric clinical study	Supports NIH participation in formation and operation of global pediatric clinical study network (§ 2072)	Confirms Congressional receptivity to global data-sharing trend, which has many benefits	Leaders need to be aware of and address potential public resistance to global sharing
Patient access to electronic health information	Builds on existing patient access right under HIPAA, promoting awareness and attention to barriers (§§ 4006–4008)	Should increase patient use of access right to contribute directly to commons	Need investment in infrastructure if direct contributions are to grow in line with proponents' hopes; may exacerbate diversity problem

FOIA, Freedom of Information Act; HIPAA, Health Insurance Portability and Accountability Act; NIH, National Institutes of Health; PHI, protected health information; PMI, Precision Medicine Initiative.

preparation and transmission of data, and participant recontact in cases where it is unclear that consent extends to broad data sharing) may not be financially sustainable.

The Act does speak to the tension between data sharing and trade secrecy, although perhaps not with the degree of nuance that those who study these issues might wish to see. This tension has become more salient in recent years as a result of two trends: the growing importance of large data sets and related interpretive algorithms to research and innovation, and patent law developments that have increased incentives to protect that information as trade secrets.^{5–9} The Act seems to affirm that proprietary interests trump data-sharing interests through its recognition that the Director's authority to mandate data sharing remains limited by existing policies intended to protect award recipients' trade secrets, proprietary interests, confidential commercial information, and intellectual property rights. Therefore, while supporting broad sharing of data from publicly funded research, the Act leaves

room for a free market to develop around commercially protected data.

Launching a diverse, trustworthy, transparent national cohort program

Section 1001 of the Act authorizes up to \$4.8 million in funding over 10 years for three NIH research initiatives that aim to build large data sets of health information: the Precision Medicine Initiative, including its All of Us program; Brain Research through Advancing Innovative Neurotechnologies; and the Beau Biden Cancer Moonshot. The All of Us program in particular aims to enroll at least 1 million US participants, and Section 2011 includes several implementation requirements.¹⁰ Along with privacy and security, these requirements address diversity, trust, and transparency. Creating genomic data sets that are reflective of the US population as a whole and include sufficient representation from African and Latin American ancestry groups and

indigenous peoples to support valid subgroup analyses has been challenging.¹¹ Further, a program that ignores or widens health disparities will be judged a failure from a public health perspective. Accordingly, the Act directs the Secretary of Health and Human Services (HHS) (Secretary) to “ensure inclusion of a broad range of participants” to include “consideration of biological, social, and other determinants of health that contribute to health disparities.” Trust is contingent on success in crafting policies under which sharing is wide but potential for abuse is low, and the Act directs the Secretary to ensure that only authorized individuals have access to collected data. Finally, to promote transparency, the Secretary is charged with creating a website that identifies entities with data access and summarizes their research projects. While these mandates are laudable, it remains to be seen whether they will be effectively translated into innovative policies and practices or be dismissed as merely hortatory.

Supporting global networks

Section 2072 of the Act expresses congressional support for NIH encouragement and facilitation of a global pediatric clinical study network. Elsewhere, three of the authors (M.A.M., A.L.M., and R.C.D.) have written about the trend toward and benefits of global genomic data sharing.¹² We have also cautioned that evidence of public resistance to global sharing should prompt leaders of such initiatives to make the case for cross-border sharing directly to the public. It is plausible that public resistance may be lessened when the intended uses can benefit children, hence a pediatric clinical study network would be an excellent test case for raising public awareness regarding the value of global collaboration.

Strengthening patient access for purposes including research contribution

The Act strengthens patients’ access to their information, thereby encouraging the creation of consumer-driven initiatives as part of the Information Commons. For example, Section 4006 requires the Secretary to promote policies ensuring that electronic health information is accessible to patients and their designees in a manner that facilitates communication with others, including researchers (and, potentially, services that match patients with researchers), consistent with their consent. The Act also endorses an educational campaign to promote awareness that patients have a right to access their medical records and other “designated record sets,” which include genetic testing records, under the Health Insurance Portability and Accountability Act (HIPAA), as amended, and mandates a Government Accountability Office study of barriers to patient access. The steps specified in the Act should increase patients’ exercise of their HIPAA access right for the purpose of obtaining data to contribute to research. At the same time, the emphasis on raising awareness and on barriers suggests that significant further investment will be required if consumer-driven data commons are to have the transformative impact that some foresee.¹³ Also, some groups that are currently

underrepresented in research may face challenges in accessing and using electronic technologies that facilitate data access and transfer (e.g., low-income patients, residents of rural areas).^{14,15} Direct patient contribution could exacerbate the diversity problem.

ADDRESSING POTENTIAL BARRIERS

Arguably, the biggest barriers to the creation of an Information Commons are concerns about the privacy and confidentiality of shared data. The Act contains several privacy-related provisions (beyond those specific to All of Us) that should reduce the risks associated with participation in research-oriented data collection and sharing initiatives and so enhance participant trust and facilitate recruitment. Further, the Act calls for clarification of two HIPAA requirements in a manner that should benefit researchers. Although not covered here, we note that the Act also addresses several barriers to information flow by, for example, promoting interoperability and imposing new penalties for “information blocking.”

Protecting against disclosure of identifiable information

Several provisions in the Act are derived from legislation originally proposed by Senators Elizabeth Warren and Mike Enzi and will expand and strengthen the protections available under certificates of confidentiality (Certificates) issued by the NIH and its sister agencies. Historically, investigators who received federal funding to conduct research that was considered sensitive could choose to apply for a Certificate. A Certificate, when issued, enabled those researchers to refuse to disclose names or other identifying characteristics of research participants in legal proceedings if they did not wish to do so. Section 2012 of the Act directs the Secretary to issue Certificates to researchers who receive federal funding, doing away with the application process, and permits the Secretary to issue Certificates to non-federally funded investigators. Further, under the Act investigators covered by Certificates are prohibited from disclosing to anyone “identifiable, sensitive information” created or compiled in the course of the research “for perpetuity,” except in a few narrowly defined circumstances, including when necessary for medical treatment of the individual, with the consent of the individual, and for the purposes of other research that complies with applicable federal human subjects protections. Notably, the Act explicitly prohibits disclosure of identifiable, sensitive information gathered by NIH-funded researchers in legal proceedings without consent. These protections become effective June 11, 2017. In addition, Section 2013 protects identifiable biomedical information collected or used during biomedical research from disclosure under the Freedom of Information Act.

The backdrop for these provisions includes ongoing debate about what makes information sensitive and the risks of re-identification of genetic and other information from which standard identifiers have been removed, given the inherent identifiability of DNA data and the proliferation of linkable

Table 2 Comparison of identifiability definitions

	21st Century Cures	HIPAA ^a	Common Rule (HHS) ^b	FDA Protection of Human Subjects ^c
Definition of identifiable	Information “that is about an individual and that is gathered or used during the course of [covered] research...for which there is at least a very small risk, as determined by current scientific practices or statistical methods, that some combination of the information, a request for the information, and other available data sources could be used to deduce the identity of an individual” ^d	Information “[i]dentifies the individual” or “there is a reasonable basis to believe the information can be used to identify the individual.” Either of the following is sufficient for de-identification: (i) expert applying generally accepted statistical and scientific principles and methods “determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual” or (ii) 18 types of identifiers are removed and covered entity lacks “actual knowledge that the information could be used alone or in combination with other information to identify an individual” ^e	“[T]he identity of the subject is or may readily be ascertained by the investigator or associated with the information” ^f	The identity of the subject is known or may “readily be ascertained by the investigator or any other individuals associated with the investigation, including the sponsor” ^g

FDA, US Food and Drug Administration; HIPAA, HIPAA, Health Insurance Portability and Accountability Act; HHS, US Department of Health and Human Services.
^aHealth Insurance Portability and Accountability Act of 1996, PL 104-191, as amended; Privacy Rule, 45 CFR Part 164 Subpart E. ^bBasic HHS Policy for Protection of Human Subjects, 45 CFR 46. ^cFDA Policy for the Protection of Human Subjects, 21 CFR 56. ^d21st Century Cures Act, Public Law No 114-255, Sec. 2012(a) (to be codified at 42 USC 241(d)(4)). ^e45 CFR 160.103, 164.514; HHS Office for Civil Rights, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, November 26, 2012. Available at: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveridentities/Deidentification/hhs_deid_guidance.pdf. ^f45 CFR 46.102(f)(2); the basic definition would be unchanged if the revisions to the Federal Policy for the Protection of Human Subjects/Common Rule proposed in the Final Rule published 19 January 2017 are implemented (for HHS, this subsection would be codified at 45 CFR 46.102(e)(5)). The Final Rule does include new provisions directing federal agencies to periodically reexamine the meaning of “identifiable private information” and permits alteration of the interpretation of these terms “[i]f appropriate and permitted by law,” and also to “assess whether there are analytic technologies or techniques that should be considered by investigators to generate “identifiable private information” through a process that would involve an opportunity for public comment (for HHS, these provisions would be codified at 45 CFR 46.102(e)(7)). The stated effective date for the revisions is 19 January 2018. ^gFDA, Guidance on informed consent for in vitro diagnostic device studies using leftover human specimens that are not individually identifiable (2006), at pp. 6, 8.

data sets. The Act does not define “sensitive” independently of “identifiable” and sets a relatively low threshold for identifiability: if there is “at least a very small risk” that an individual’s identity could be deduced from the sum of available data using current scientific practices or statistical methods, then the information would be covered by the Certificate (see **Table 2**). Thus, it becomes possible to assure potential participants that even a very small risk of re-identification will keep their information safe from release in a variety of contexts. At the same time, owing to the research exception, identifiable, sensitive information can still circulate relatively freely for legitimate research purposes, thereby facilitating open and responsible data sharing.

Clarifying research-related HIPAA requirements

“Protected health information” (PHI) includes most identifiable information held by health-care providers and other HIPAA-covered entities. Section 2063 of the Act directs the Secretary to issue guidance on the circumstances under which

authorizations for purposes of future research use or disclosure of PHI contain a sufficient description of those purposes. An earlier version of the legislation, which passed the House but died in the Senate, more clearly signaled strong congressional support for one-time authorizations of use and disclosure of PHI for research purposes, sometimes called “broad consent.”¹⁶ Despite the shift in framing, this provision should advance the broad consent paradigm, which facilitates data sharing, especially given acceptance of that paradigm in revisions to the Common Rule and in new international ethics guidelines.^{17,18} The Act also directs the Secretary to issue guidance clarifying that reviews of PHI preparatory to research can be carried out remotely so long as security and privacy safeguards are in place and PHI is not retained by the researcher.

CHALLENGES EXACERBATED OR UNRESOLVED

The Act goes a long way to advance data sharing and the creation of an Information Commons that includes initiatives

in the public and private sectors while addressing privacy and other barriers. However, provisions covering privacy and identifiability could inadvertently raise some new barriers to realizing the vision of a robust commons serving multiple purposes. The Act also postpones action on two major ethical and policy questions related to data sharing pending further study.

Siloing of research and clinical care

If the goal is a commons ecosystem that can be used for clinical as well as research purposes, then an unintended consequence of the Act may be reinforcement of the siloing of research and clinical care. Commentators have argued that the distinction between research and clinical care is a significant barrier to creation of learning health-care systems that benefit patients and society, and that a bifurcated policy approach, with one set of practices for research and another for clinical care, does a poor job of matching regulatory protections (and associated regulatory burdens) to risks.¹⁹ As noted above, the Act's new research subject privacy protections permit broad sharing of information for research purposes even when that information remains identifiable. Yet, there is no parallel pathway permitting sharing of information created or compiled in the course of research—even if stripped of all direct identifiers, so long as there is at least a very small risk of re-identification—for clinical purposes without the consent of the individual. Hence, the utilization of consent language that encompasses appropriate clinical uses will take on additional importance, as will clarification that legal representatives may provide consent on behalf of individuals who lack capacity to consent.

Reconciling definitions of identifiability

Further, from a complexity perspective, institutions that engage in research and clinical care will now have to navigate, potentially, four different federal definitions of identifiability set forth in: (i) the Act, (ii) HIPAA, (iii) human research subject regulations that apply to federally funded research known as the Common Rule, and (iv) human research subject regulations that apply to activities regulated by the Food and Drug Administration (FDA) (see **Table 2**). The HIPAA definition and related standards are sufficiently complex that a 32-page guidance document is required to aid interpretation.²⁰ Most relevant here, an expert determination that re-identification risk is “very small” is a condition for one HIPAA-sanctioned approach to finding that information is *not* individually identifiable. Yet, a finding that there is a “very small” risk of re-identification would be sufficient to establish identifiability for purposes of the Act. Clearly, then, the thresholds for identifiability as a trigger for compliance with HIPAA and the Act are not the same (see **Table 2**). Under the Common Rule and FDA regulations, identifiable means that the identity of the individual is known or may readily be ascertained, yet another approach.

Unleashing health data-based research and data markets?

Congress ultimately ducked two important ethical and policy issues that were addressed in an earlier version of the legislation, turning them into questions for a working group to address via recommendations. The first question is whether to remove HIPAA's patient authorization requirements for use and disclosure of PHI for at least some categories of research. If so, these researchers will be relieved of the burden of securing waivers of patient authorization or removing identifiers before use or disclosure. The second question is whether to remove HIPAA's requirement that health-care providers and other covered entities obtain specific patient authorization before selling PHI to researchers for profit. At least one commentator has argued that removing restrictions on data sales by laboratories and other health-care providers, thus accelerating development of markets in health data, would help reduce the financial burden of data sharing.²¹ An earlier version of the Act answered both questions in the affirmative, exempting several categories of research from patient authorization requirements and permitting sale of PHI for research, generating controversy.²² If the timetable specified in the Act is followed, the recommendations of the working group on these and other matters will be published within two years.

CONCLUSION

The 21st Century Cures Act promotes an environment favorable to a flourishing Information Commons. Among other things, we applaud its emphasis on direct engagement of patients as active participants in the management of their health data and believe that its provisions will increase patients' access to their data. Yet, if this engagement is to translate into sharing with researchers and clinicians, the ability to access and transmit data in interoperable format must be built into the infrastructure of health systems, exchanges, and repositories. Further, it will be important to involve stakeholders in future deliberations about whether to remove HIPAA requirements that position patients as important (if largely passive) gatekeepers for health data research and markets. Finally, it is past time for the definition of identifiability, which is the trigger for most legal protections of data, to be harmonized across all federal requirements to the extent feasible. By creating yet another standard for the kinds of data that merit protection, the Act adds, rather than reduces, complexity for the very individuals responsible for realizing the law's vision of a data-sharing future.

ACKNOWLEDGMENTS

This work was funded by National Institutes of Health National Human Genome Research Institute grant R01 HG008918.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. 21st Century Cures Act, Pub. L. No. 114-255 (2016). <https://www.congress.gov/bill/114th-congress/house-bill/34/text>. Accessed 10 January 2017.

2. National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. National Academies Press: Washington, DC, 2011.
3. ACMG Board of Directors. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet Med*; e-pub ahead of print 5 January 2017.
4. National Institutes of Health. Genomic Data Sharing Policy (28 August 2014). <https://gds.nih.gov/03policy2.html>. Accessed 10 January 2017.
5. Simon BM, Sichelman TM. Data-generating patents. *Northwestern Univ Law Rev* 2017;111:377–437.
6. Ray T. Supreme Court patent cases haven't hindered diagnostics innovation, preliminary data suggest. 5 December 2016. <https://www.genomeweb.com/molecular-diagnostics/supreme-court-patent-cases-havent-hindered-diagnostics-innovation-preliminary>. Accessed 10 January 2017.
7. Cook-Deegan R, Conley JM, Evans JP, Vorhaus D. The next controversy in genetic testing: clinical data as trade secrets? *Eur J Hum Genet* 2013;21:585–588.
8. Conley JM, Cook-Deegan R, Lázaro-Muñoz G. Myriad after *myriad*: the proprietary data dilemma. *N C J Law Technol* 2014;15:597–637.
9. Laakmann AB. The new genomic semicommons. *UC Irvine Law Rev* 2015;5:1001–1039.
10. National Institutes of Health. All of Us Research Program. <https://www.nih.gov/research-training/all-of-us-research-program>. Accessed 10 January 2017.
11. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538:161–164.
12. Majumder MA, Cook-Deegan R, McGuire AL. Beyond our borders? Public resistance to global genomic data sharing. *PLoS Biol* 2016;14:e2000206.
13. Evans BJ. Barbarians at the gate: consumer-driven health data commons and the transformation of citizen science. *Am J Law Med* 2016;42:651–686.
14. Ancker JS, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. *Am J Manag Care* 2016;22:539–540.
15. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. *J Rural Health* 2016;00:1–9.
16. 21st Century Cures Act, H.R. 6, as engrossed in House July 10, 2015. <https://www.congress.gov/bills/114/congress/house-bill/6/text/eh>. Accessed 10 January 2017.
17. Federal Policy for the Protection of Human Subjects: Final Rule. 82 Fed. Reg. 7149 (19 January 2017).
18. Council for International Organizations of Medical Sciences. International Ethical Guidelines for Health-Related Research Involving Humans. Geneva, Switzerland: Council for International Organizations of Medical Sciences, 2016. <http://www.cioms.ch>. Accessed 10 January 2017.
19. Faden RR, Beauchamp TL, Kass NE. Informed consent, comparative effectiveness, and learning health care. *N Engl J Med* 2014;370:766–768.
20. U.S. Department of Health and Human Services. Office for Civil Rights. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 26 November 2012. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/covered_entities/De-identification/hhs_deid_guidance.pdf. Accessed 10 January 2017.
21. Evans BJ. Mining the human genome after association for molecular pathology v. *myriad* genetics. *Genet Med* 2014;16:504–509.
22. Monegain B. Privacy advocates blast 21st Century Cures bill. 14 May 2015. <http://www.healthcareitnews.com/news/privacy-advocates-blast-cures-bill>. Accessed 10 January 2017.