



August 1, 2014

Senators Ron Wyden and Charles Grassley
Committee on Finance
United States Senate
219 Dirksen Senate Office Building
Washington, D.C. 20510

RE: Health Care Transparency

Dear Senators Wyden and Grassley,

On behalf of the Center for Data Innovation (www.datainnovation.org), I am pleased to submit these comments in response to the U.S. Senate Committee on Finance's request for public comment on health care data transparency.¹

The Center for Data Innovation is a non-profit, non-partisan, Washington-DC based think tank focusing on the impact of the increased use of information on the economy and society. In addition, the Center formulates and promotes pragmatic public policies designed to enable data-driven innovation in the public and private sectors, create new economic opportunities, and improve quality of life. The Center is affiliated with the Information Technology and Innovation Foundation.

Broader sharing and use of data offers a broad range of benefits to the health care sector, including more personalized and coordinated care, better hospital quality, faster treatment development, and lower costs to payers, providers, and pharmaceutical companies alike.² According to a 2013 report from McKinsey & Company, the value of big data in U.S. health care exceeds \$300 billion annually.³ Government can help maximize the benefits data offers to this diverse group of stakeholders by ensuring that taxpayer-funded research data is made accessible in a timely manner, releasing granular payments data from federal health care programs, and encouraging data sharing across the health care industry. In particular, data sharing efforts are crucial to advancing patient care, and contemporary data sharing initiatives should be conceived with an eye toward a future in which the

¹ "Wyden and Grassley Seek Next Steps on Health Care Transparency," United States Senate Committee on Finance, June 12, 2014, <http://www.finance.senate.gov/newsroom/chairman/release/?id=d31e8e26-0c94-4bcd-afb5-f2d996edd390>.

² Peter Groves et al, "The 'Big Data' Revolution in U.S. Health Care," McKinsey & Company, 2013, http://www.mckinsey.com/insights/health_systems_and_services/the_big_data_revolution_in_us_health_care.

³ Ibid.



full range of medical information, including medical histories, genomic data, behavioral data, and other sources can be coordinated in the service of better care.

WHAT DATA SOURCES SHOULD BE MADE MORE BROADLY AVAILABLE?

There are at least three main opportunities to make potentially beneficial health care data more broadly available: releasing taxpayer-funded research data in non-proprietary and machine readable formats; releasing claims data from other programs besides Medicare, including data from Medicaid, Tricare, Federal Employees Health Benefits, and Indian Health; and encouraging data sharing across the health care industry.

First, agencies must strive to maximize access to taxpayer-funded research data and ensure that it is available for reuse as quickly as possible. In February 2013, the White House Office of Science and Technology Policy (OSTP) released a memorandum directing each federal agency with over \$100 million in annual research and development expenditures to develop a plan to increase public access to the results and data of federally funded research.⁴ Agencies submitted their plans in late 2013 and early 2014 and are now revising them pursuant to OSTP's feedback.⁵ In its final review, OSTP should approve plans that maximize access to medical research data, such as the National Institute of Health's (NIH) proposal to expand PubMed Central, its research repository, to include federally-funded medical research data from agencies outside the NIH. Although the memorandum recommended agencies wait 12 months after research is first published to make it publicly available, OSTP should support efforts to shorten that timeline to six months or less for medical research.⁶ Giving researchers access to study data as rapidly as possible would help maximize the benefits of the data, allowing researchers to reuse the data for critical applications, such as lifesaving medical treatments, and ensure that the process of research commercialization proceeds as quickly as possible.⁷ Second, other federal health care programs should follow Medicare's lead in releasing granular data on physician-level payments. The Medicare data set released in 2014 promises to reduce fraud, waste, and abuse, offer greater accountability among doctors, and enable

⁴ John Holdren, "Increasing Access to the Results of Federally Funded Scientific Research," White House Office of Science and Technology Policy, February 22, 2013,

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

⁵ John Holdren, Letter on federal agencies' progress developing public research data access plans, March 24, 2014, http://www.whitehouse.gov/sites/default/files/microsites/ostp/OpenAccess_March-2014.pdf.

⁶ Confederation of Open Access Repositories, "Statement about embargo periods," May 14, 2014, <https://www.coar-repositories.org/activities/advocacy-leadership/aligning-repository-networks-across-regions/statement-about-embargo-periods/>.

⁷ Comments in response to the White House Office of Science and Technology Policy Request for Information: Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research, Mendeley, 2011, [http://www.whitehouse.gov/sites/default/files/microsites/ostp/scholarly-pubs-\(%23288\).pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/scholarly-pubs-(%23288).pdf).



entrepreneurs to create patient-facing services derived from this data to allow patients to make better decisions based on the billing habits of their doctors.⁸ The same could be true of claims data from other federal programs.⁹

Third, research agencies should continue to encourage data sharing among medical organizations. The Accelerating Medicines Partnership, an NIH-led drug discovery collaboration among 10 drug makers, which launched this year, is a model effort.¹⁰ In this effort, participating companies created a shared database that exceeded the capabilities of any individual company's data holdings. As this collaboration and others like it will accelerate drug discovery research, the NIH should strive to duplicate these efforts with additional companies in the future. In addition, government agencies should work with medical organizations on global standards and harmonization efforts, such as the Global Alliance for Genomics and Health and the Research Data Alliance.¹¹

HOW, IN WHAT FORM, AND FOR WHAT PURPOSES SHOULD THIS DATA BE CONVEYED?

To the extent possible, data should be conveyed according to demand. Where information is already public but is not yet available in machine readable formats, data providers can prioritize releases according to the number of Freedom of Information Act (FOIA) requests, as the Food and Drug Administration (FDA) did with its OpenFDA initiative.¹² If agencies do not have access to such information, they can seek public comment or host open data workshops to learn of areas of greatest demand. Alternatively, agencies can use online tools to gain continuous feedback from users. Data.gov, for example, offers users a mechanism for requesting new datasets that could serve as a model for other agencies.¹³

⁸ "CMS Releases Trove Of Medicare Physician Billing Data," *Kaiser Health News*, April 9, 2014, <http://www.kaiserhealthnews.org/daily-reports/2014/april/09/medicare-billing-data.aspx/>

⁹ Sarah Zellner, "HDC's Open Health Data Wish List for 2014." Health Data Consortium, December 18, 2013, http://www.healthdataconsortium.org/blog_home.asp?display=70.

¹⁰ Damian Garde, "Big Pharma superteam joins NIH to share data, discover new drugs," *FierceBiotech*, February 4, 2014, <http://www.fiercebiotech.com/story/big-pharma-superteam-joins-nih-share-data-discover-new-drugs/2014-02-04>.

¹¹ David Altshuler, M.D., PhD, "Creating a Global Alliance for Sharing of Genomic and Clinical Data," White House Champions of Change, June 20, 2013, <http://www.whitehouse.gov/blog/2013/06/20/creating-global-alliance-sharing-genomic-and-clinical-data>.

Mary Martialay, "International Research Data Alliance Facilitates Data Sharing at First U.S. Meeting," Rensselaer Polytechnic Institute News, September 18, 2013, <http://news.rpi.edu/content/2013/09/18/international-research-data-alliance-facilitates-data-sharing-first-us-meeting>.

¹² Taha Kass-Hout, "OpenFDA: Innovative Initiative Opens Door to Wealth of FDA's Publicly Available Data," OpenFDA, June 2, 2014, <https://open.fda.gov/update/openfda-innovative-initiative-opens-door-to-wealth-of-fda-publicly-available-data/>.

¹³ "Request Data," <http://Data.gov>.



Public data not containing sensitive information should be released online and without restrictions in open, machine-readable formats. Data providers should strive to make the data available for bulk download, or, failing that, through an application programming interface (API) using standard representational state transfer (REST) principles. The REST API style allows users to query data using standard web protocols without requiring additional mechanisms to connect the user to the database. Data that is not intended for public use, such as research data containing sensitive information, should be available online to authenticated users through a secure portal with controlled access.

In order to support the aforementioned beneficial new applications for patients, providers, payers, pharmaceutical companies, medical researchers, and other stakeholders, data should be available for exploration and reuse beyond the purpose for which it was originally obtained. This allows data users to ask questions of the data, test hypotheses, and experiment with new applications without needing to obtain the data anew for each individual use, cutting administrative costs and ensuring that users can develop innovative applications as rapidly as possible.¹⁴

WHAT REFORMS WOULD HELP REDUCE THE UNNECESSARY FRAGMENTATION OF HEALTH CARE DATA?

The Patient-Centered Outcomes Research Institute's (PCORI) National Patient-Centered Clinical Research Network (PCORnet) is a national network through which participating health care providers and condition-specific patient groups can share clinical effectiveness research and other health information.¹⁵ Data sharing networks like PCORnet amplify the value of individual effectiveness studies by combining them with many others, facilitating meta-analysis and other kinds of derivative research. While condition-specific networks such as PCORnet are useful in the short-term, the long-term goal of Congress should be to develop a fully-integrated national health research network that integrates data from providers, payers, device manufacturers, and other data holders. As such, organizations overseeing data sharing networks should design them with an eye toward interoperability with other networks in the future.

¹⁴ Daniel Castro, "Re: Spring Privacy Series: Consumer Generated and Controlled Health Data, Project No. P145401," *The Center for Data Innovation*, <http://www2.datainnovation.org/2014-ftc-patient-generated-health-data.pdf>.

¹⁵ "PCORnet: The National Patient-Centered Clinical Research Network," Patient-Centered Outcomes Research Institute December 16, 2013, <http://www.pcori.org/funding-opportunities/pcornet-national-patient-centered-clinical-research-network/>.



WHAT BARRIERS STAND IN THE WAY OF STAKEHOLDERS USING EXISTING DATA SOURCES MORE EFFECTIVELY AND WHAT REFORMS SHOULD BE MADE TO OVERCOME THESE BARRIERS?

The main barriers to effective data use are not technical or legal use restrictions, but rather unwillingness among companies in certain industries to share data, an inability for the private sector to access this data for research purposes, and an inability by providers to integrate novel data sources, such as patient-generated health data, into electronic health record systems.¹⁶

Although there is a net benefit from data sharing to patients and researchers, there is not always a short-term benefit to companies for making their data available to their competitors. In some areas, such as pharmaceuticals, companies that hold data have recognized that sharing is in their collective interest and have participated in sharing programs, including the Accelerating Medicines Partnership and a GlaxoSmithKline-led initiative to share patient-level clinical trial data with other participating pharmaceutical companies.¹⁷ In other areas, such as medical device manufacturing, where data sharing has been a more difficult prospect, it could be helpful for the FDA to conduct a review of the industry to identify specific factors preventing companies from sharing data with one another and how any market failures might be most effectively addressed through policy.¹⁸

The private sector's inability to access certain kinds of medical data for research purposes also presents a barrier to use. Some companies have resorted to collecting their own data even though comparable information may already be available to other medical researchers. For example, Google's Baseline Initiative, in which the company will collect genomic data from 175 individuals and may expand to thousands more in order to establish a set of baseline genomic characteristics exhibited by healthy bodies, is volunteer-based.¹⁹ Private sector researchers should have access to this data under the same kinds of data sharing agreements afforded to other researchers.

¹⁶ Chris Murphy, "Sharing Electronic Medical Records Still Too Hard," *InformationWeek Health Care*, June 14, 2013, <http://www.informationweek.com/interoperability/sharing-electronic-medical-records-still-too-hard/d/d-id/1110386> and

Israel Green-Hopkins, MD, "Incorporate patient-generated health data into the EMR," *MedPage Today*, March 2, 2014, <http://www.kevinmd.com/blog/2014/03/incorporate-patientgenerated-health-data-emr.html>.

¹⁷ Harlan Krumholz, "Time To Assess Pharma Progress In Data Sharing" *Forbes*, June 2, 2014, <http://www.forbes.com/sites/harlankrumholz/2014/06/02/time-to-assess-pharma-progress-in-data-sharing/>.

¹⁸ Michael Causey, "Medical Device Makers Urged to Play Nicier by Sharing Data," *Quality Digest*, February 19, 2014, <http://www.qualitydigest.com/inside/fda-compliance-column/medical-device-makers-urged-play-nicer-sharing-data.html-0>.

¹⁹ Alistair Barr, "Google's New Moonshot Project: the Human Body," *The Wall Street Journal*, July 27, 2014, <http://online.wsj.com/articles/google-to-collect-data-to-define-healthy-human-1406246214>.



A final barrier in the way of stakeholders using existing data more effectively is that some emerging data sources are not yet widely interoperable with existing electronic health record systems. One example of this problem is patient-generated health data, information logged by patients themselves or collected in a connected device outside a clinical setting. The volume of patient-generated health data has increased rapidly in recent years due in part to the proliferation of wearable health devices such as fitness trackers.²⁰ Patient-generated data has considerable value, as it is typically much more granular than laboratory data, allows for real-time monitoring, and can capture small fluctuations that laboratory tests may miss, but it has not been widely integrated into electronic health records.²¹ One way to ensure that this data is available for use is to include patient-generated health data interoperability in the Center for Medicare and Medicaid Services' Meaningful Use incentives.²² This would accelerated the process of patient-generated data integration and ultimately allow doctors and researchers to more effectively use this unique data. The U.S. Department of Health and Human Services should also secure commitments from device manufacturers and other companies that collect patient-generated data to grant individuals access to their own data. Such a program, modeled after the Blue Button initiative for general health data, would give individuals insights into their own health and allow them to authorize third parties to create applications using that data.²³

CONCLUSION

The future of health care in the United States will be determined in part by the quality of the data systems, standards, and sharing infrastructure being established today. As such, today's health data projects should be designed with an eye toward what we want that future to look like, in which the full range of information about a patient can be brought to bear quickly and usefully to help determine the best treatment for that individual. This means building systems that encourage data flows and breaking down information silos between different parts of the medical establishment. Only by integrating data from clinical trials, genomic databases, patient-generated sources, hospital

²⁰ Michael Shapiro, et al., "Patient-Generated Health Data," RTI International, April 2012, http://www.healthit.gov/sites/default/files/rti_pghd_whitepaper_april_2012.pdf.

"Patient-Generated Health Data," Health IT.gov, <http://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data>.

²¹ "Opening the Doors to Clinical Trial Data," *Alzforum*, November 15, 2013, <http://www.alzforum.org/news/community-news/opening-doors-clinical-trial-data>.

²² Dave Moore, "House Hearing Explores Intersection of Health Care and Technology" Association of American Medical Colleges, July 18, 2014, <https://www.aamc.org/advocacy/washhigh/highlights2014/385572/071814househearingexploresintersectionofhealthcareandtechnology.html>.

²³ "About Blue Button," HealthIT.gov, <http://www.healthit.gov/patients-families/blue-button/about-blue-button>.



quality data, and many other sources can we paint a full picture of health, and it is with this goal in mind that current data efforts should proceed.

Sincerely,

Daniel Castro

Director, Center for Data Innovation
1101 K Street NW, Suite 610
Washington, DC 20005

dcastro@datainnovation.org