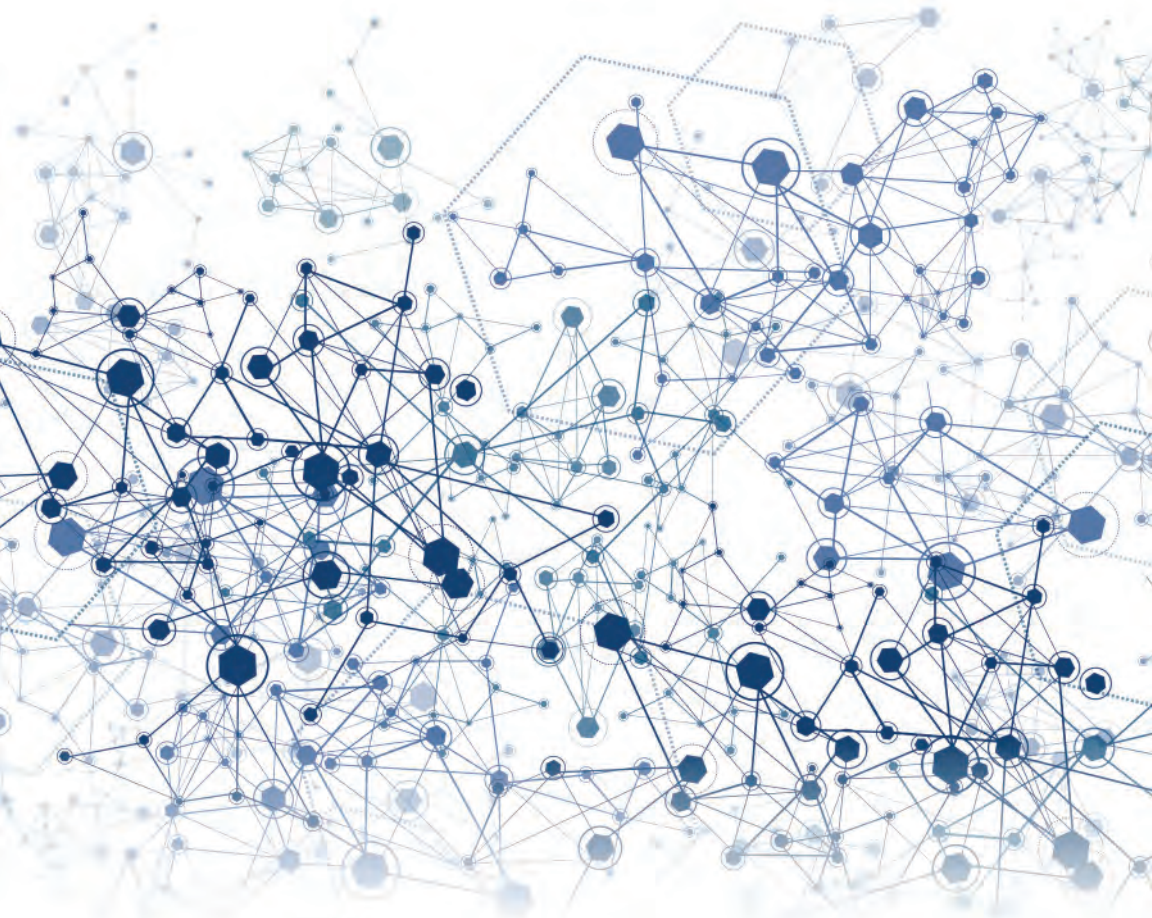


THE LEARNING HEALTH SYSTEM SERIES

**SARAH M. GREENE, MAHNOOR AHMED, PEAK SEN CHUA,  
AND CLAUDIA GROSSMAN, EDITORS**

# SHARING HEALTH DATA

**THE WHY, THE WILL, AND  
THE WAY FORWARD**



THE LEARNING HEALTH SYSTEM SERIES

# Sharing Health Data: The Why, the Will, and the Way Forward

National Academy of Medicine  
NAM Leadership Consortium: Collaboration for  
a Learning Health System

THE NATIONAL ACADEMIES PRESS  
*Washington, DC*  
**[www.nap.edu](http://www.nap.edu)**

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW WASHINGTON, DC 20001

This publication has undergone peer review according to procedures established by the National Academy of Medicine (NAM). Publication by the NAM signifies that it is the product of a carefully considered process and is a contribution worthy of public attention, but does not constitute endorsement of conclusions and recommendations by the NAM. This publication is sponsored by the Patient-Centered Outcomes Research Institute (PCORI). The 11 case studies featured in this manuscript were selected based on input from PCORI and the editors and contributors to the NAM Special Publication *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*, the progenitor of this publication. Opinions and advice presented within each case study are those of the interviewees. The case studies are solely an educational exercise and views presented in this publication are those of individual contributors and do not represent formal consensus positions of the editors' organizations; the NAM; or the National Academies of Sciences, Engineering, and Medicine.

Library of Congress Catalog Number: 2022293351

Copyright 2022 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Suggested citation: National Academy of Medicine. 2022. *Sharing Health Data: The Why, the Will, and the Way Forward*. S. M. Greene, M. Ahmed, P. S. Chua, and C. Grossmann, editors. NAM Special Publication. Washington, DC: National Academies Press.

*“Knowing is not enough; we must apply.  
Willing is not enough; we must do.”  
-GOETHE*

LEADERSHIP  
INNOVATION  
IMPACT

---

for a healthier future

## ABOUT THE NATIONAL ACADEMY OF MEDICINE

The **National Academy of Medicine** is one of three Academies constituting the National Academies of Sciences, Engineering, and Medicine (the National Academies). The National Academies provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on issues of health, health care, and biomedical science and technology. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

**Learn more about the National Academy of Medicine at  
NAM.edu.**

# SHARING HEALTH DATA: THE WHY, THE WILL, AND THE WAY FORWARD

## *Steering Committee and Working Group Participants*

**ERIN MACKAY** (*Co-Chair*), National Partnership for Women & Families

**PETER MARGOLIS** (*Co-Chair*), Cincinnati Children's Hospital Medical Center

**KRISTIN CARMAN**, Patient-Centered Outcomes Research Institute

## *NAM Staff*

Development of this publication was facilitated by contributions of the following NAM staff, under the guidance of **J. Michael McGinnis**, Leonard D. Schaeffer Executive Officer and Executive Director of the Leadership Consortium for a Value & Science-Driven Health System:

**SARAH M. GREENE**, Senior Advisor

**MAHNOOR AHMED**, Associate Program Officer

**PEAK SEN CHUA**, Consultant

**JENNA L. OGILVIE**, Deputy Director of Communications

## *Patient-Centered Outcomes Research Institute Staff*

**NAKELA COOK**, Executive Director

**CLAUDIA GROSSMANN**, Associate Director, Research Infrastructure

## REVIEWERS

This Special Publication was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with review procedures established by the National Academy of Medicine.

We wish to thank the following individuals for their contributions:

**MARY GREALY**, Healthcare Leadership Council  
**WARREN KIBBE**, Duke Cancer Institute  
**JENNIE LARKIN**, National Institute on Aging  
**JOHN ORLOFF**, Alexion Pharmaceuticals (*until July 2021*)  
**DARREN TOH**, Harvard Medical School and  
Harvard Pilgrim Health Care Institute

The reviewers listed above provided many constructive comments and suggestions, but they were not asked to endorse the content of the publication, and did not see the final draft before it was published. Review of this publication was overseen by **SARAH M. GREENE**, Senior Advisor; **MAHNOOR AHMED**, Associate Program Officer, and **J. MICHAEL MCGINNIS**, Leonard D. Schaeffer Executive Officer, NAM. Responsibility for the final content of this publication rests with the editors and the NAM.

# CONTENTS

**Foreword, xix**

**1 Introduction, 1**

**2 Case Study: OpenNotes, 9**

**3 Case Study: The University of Michigan (U-M), 17**

**4 Case Study: Luna, 25**

**5 Case Study: COVID-19 Evidence Accelerator (EA), 33**

**6 Case Study: The National COVID Cohort Collaborative (N3C), 43**

**7 Case Study: The Yale Open Data Access (YODA) Project, 51**

**8 Case Study: Vivli, 57**

**9 Case Study: The Louisiana Public Health Institute (LPHI) and the  
Greater New Orleans Health Information Exchange  
(GNOHIE), 67**

**10 Case Study: The Sanford Health System and Sanford Data  
Collaborative (SDC), 77**

**11 Case Study: Blue Cross Blue Shield of North Carolina, 83**

**12 Case Study: Mayo-Google Partnership, 89**

**13 Conclusion, 97**

**References, 107**

**Appendix A: Key Informant Interview Questions, 117**



# BOXES AND FIGURES

## Boxes

- 1 Defining Real-World Data and Real-World Evidence, 35
- 2 Common Data Models: Definition and Applied Example, 45

## Figures

- 1 Priority Barriers to Health Data Sharing, 3
- 2 Funding Sources for the Data-Sharing Initiatives, 6
- 3 Real-World Data for COVID-19 Ecosystem, 36
- 4 COVID-19 Evidence Accelerator Work Streams, 37
- 5 COVID-19 Evidence Accelerator Principles, 39
- 6 Diagram of a Federated Learning Model, 92
- 7 Mapping of Case Studies to Priority Barriers, 98

## Tables

- 1 Case Study Characteristics, 5
- 2 List of Vivli's Data Contributors by Sector, 60
- 3 Role of Vivli's Data-Sharing Platform in the Data-Sharing Ecosystem, 61

## ACRONYMS AND ABBREVIATIONS

ADT API	Admit, Discharge, and Transfer data application programming interfaces
CCBC CCPA CD2H CMS CTSA	Crescent City Beacon Community California Consumer Privacy Act National Center for Data to Health Centers for Medicare & Medicaid Services Clinical and Translational Science Awards
DUA	data use agreement
EA EHR	COVID-19 Evidence Accelerator electronic health record
FDA FHIR	U.S. Food and Drug Administration Fast Healthcare Interoperability Resources
GDPR GNOHIE	General Data Protection Regulations Greater New Orleans Health Information Exchange
HIPAA	Health Insurance Portability and Accountability Act
HHS HL7	U.S. Department of Health and Human Services Health Level 7
IRB	Institutional Review Board
LPHI	Louisiana Public Health Institute
MRCT	Multi-Regional Clinical Trials
N3C NAM NCATS	National COVID Cohort Collaborative National Academy of Medicine National Center for Advancing Translational Sciences
NHLBI NIH	National Heart, Lung, and Blood Institute National Institutes of Health

OMOP ONC	Observational Medical Outcomes Partnership Office of the National Coordinator for Health Information Technology
PATH PCORI PCORnet®	Partnership for Achieving Total Health Patient-Centered Outcomes Research Institute National Patient-Centered Clinical Research Network
PhRMA	Pharmaceuticals Research and Manufacturers of America
RWD RWE	real-world data real-world evidence
SDC SEC	Sanford Data Collaborative Securities and Exchange Commission
TEFCA	Trusted Exchange Framework and Common Agreement
U-M	University of Michigan
VA VBC	U.S. Department of Veterans Affairs Value-Based Care
YODA	Yale University Open Data Access

## FOREWORD

Health data has proven its centrality in guiding action to change the course of individual and population health, if properly stewarded and used. Consequently, we are obliged to use it to its fullest, most beneficent potential. Too often barriers, such as disagreements about data ownership and misaligned incentives, have impeded research or practice, further hindering this potential. Evolving from a mindset of data guarding to data sharing is essential to meet society's potential: it is imperative for effectiveness, efficiency, and equity in health system performance. In the context of the COVID-19 pandemic, both data and a lack of data illuminated profound shortcomings that affected health care and health equity. Yet a silver lining of the pandemic was a surge in collaboration among data holders in public health, health care, and technology firms, suggesting that an evolution in health data sharing is visible and tangible. This publication features some of these novel data-sharing collaborations born out of the pandemic.

This Special Publication by the National Academy of Medicine, featuring case studies, has been developed to provide practical context and implementation guidance important to advancing the lessons identified in its progenitor Special Publication, entitled *Health Data Sharing: Building a Foundation of Stakeholder Trust*. Developed in light of the realization that progress in data sharing and collaboration required the direct involvement of data stewards and stakeholders, the former publication was the culmination of a two-year partnership between Patient-Centered Outcomes Research Institute and the National Academy of Medicine to understand barriers and facilitators to health data sharing across a number of key stakeholder groups. With the goal of charting a collaborative path forward, that Special Publication identified several cultural, operational, financial, and ethical barriers meriting concerted attention and prioritized several action steps that could be accomplished in

a one-to-three-year time horizon. One of those important action steps is the focus of this publication: to identify and describe exemplar groups to dispel the myth that sharing health data more broadly is impossible and illuminate the innovative approaches that are being taken to make progress in the current environment. It also serves as a resource for those waiting in the wings, showing how barriers can be addressed and harvesting lessons and insights from those on the front lines.

Common to all of the case studies included in this report is the fundamental importance of strong partnerships built on common ground. The foundation of trust that enables collaborative data sharing activities is only attained through trustworthy and transparent behaviors and actions by all partners. As has been seen in numerous historic examples, misuse of health data has spawned a long legacy of mistrust. Hence, even when a data sharing compact has been established, ongoing, diligent attention to a shared commitment and shared values is imperative.

The profiled organizations were selected based on notable data-sharing interactions, dynamics, and barriers identified in the progenitor Special Publication. Given the dynamic state of partnerships, new exemplars are emerging with growing frequency. Interviews with representatives from these case studies not only describe how they overcame the specific barriers that drove their inclusion in this report, but also identify additional pertinent considerations that others should note as they engage in similar data-sharing efforts.

These profiles illustrate how a small but committed group of vanguard entities is willing to push for a different way. Grounded in shared values, a common purpose, and clearly defined processes, these organizations have shown how barriers, once thought to be intractable, can be surmounted. Fifteen years ago, the idea of patients having free and seamless access to the clinical notes in their medical records was progressive and exceptional. Today, legislation obligates health systems to provide patient access to records as the norm rather than the exception. Ten years ago, compensation strategies for people who contribute genomic data to a repository was inconceivable. However, numerous private companies are reevaluating the value proposition for patients who are willing to

share data to further the research enterprise.

This compilation of case studies is one of many tools to drive action and progress toward broader health data sharing. We are hopeful that examining the partnership approaches described in these case studies can motivate new thinking and new action. A potentially valuable next step articulated in the previous publication is a national conversation about the importance and benefit of sharing health data. Particularly in this circumstance, it is important to underscore the very real risks and consequences of not sharing data. Nonetheless, the approach to such a dialogue must be supported by a foundation of meaningful engagement and partnership with patients, families, and community members who often have been sidelined in these conversations, the keystone for genuine, sustained trust building.

Refining and updating the regulatory context is a logically important element on the path forward. Many regulations were developed well before the expansion of data sources, modalities for data exchange, and attendant security considerations for these new mechanisms. However, these outmoded regulations are now drawing attention to themselves and the facilitative advances needed. The enactment of information-blocking regulations by the Office of the National Coordinator for Health IT, the October 2020 release of the NIH Policy for Data Management and Sharing, and the potential changes to the Health Information Portability and Accountability Act indicate that modernization of data-sharing regulations may be realized in the decade ahead.

In the meantime, knowledge is already available to foster better health care and health outcomes, and the examples of new partnerships and new technologies described in this volume suggest how intentional and assiduous attention to health data sharing can enable unparalleled advances, securing a healthier and more equitable future for all.

Nakela L. Cook, MD, MPH  
Executive Director, PPCORI

J. Michael McGinnis, MD, MPP  
Executive Officer, National Academy of Medicine

# 1

## INTRODUCTION AND INITIATIVE BACKGROUND

Sharing health data and information<sup>1</sup> across stakeholder groups is the bedrock of a learning health system. As data and information are increasingly combined across various sources, their generative value to transform health, health care, and health equity increases significantly. Facilitating this potential is an escalating surge of digital technologies (i.e., cloud computing, broadband and wireless solutions, digital health technologies, and application programming interfaces [APIs]) that, with each successive generation, not only enhance data sharing, but also improve in their ability to preserve privacy and identify and mitigate cybersecurity risks. These technological advances, coupled with notable policy developments, new interoperability standards (particularly the Fast Healthcare Interoperability Resources [FHIR] standard), and the launch of innovative payment models within the last decade, have resulted in a greater recognition of the value of health data sharing among patients, providers, and researchers. Consequently, a number of data sharing collaborations are emerging across the health care ecosystem.

---

<sup>1</sup> As defined in the progenitor publication, *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*, health data is all the information that accumulates about a person or population that may affect health outcomes. This includes, but is not limited to: 1) health data generated during clinical encounters and stored in electronic health records or other data systems; 2) health insurance claims data; 3) data gathered from clinical and health services research; 4) genomic, proteomic, and immunomic data; 5) data related to the social and environmental determinants of health collected during clinical encounters or outside of the health care system through community, state, and federal organizations; and 6.) patient-generated health data, which has been defined as health-related data created, recorded, or gathered by or from patients (or family members or other caregivers).

Health information results from the analysis and synthesis of various pieces of data.

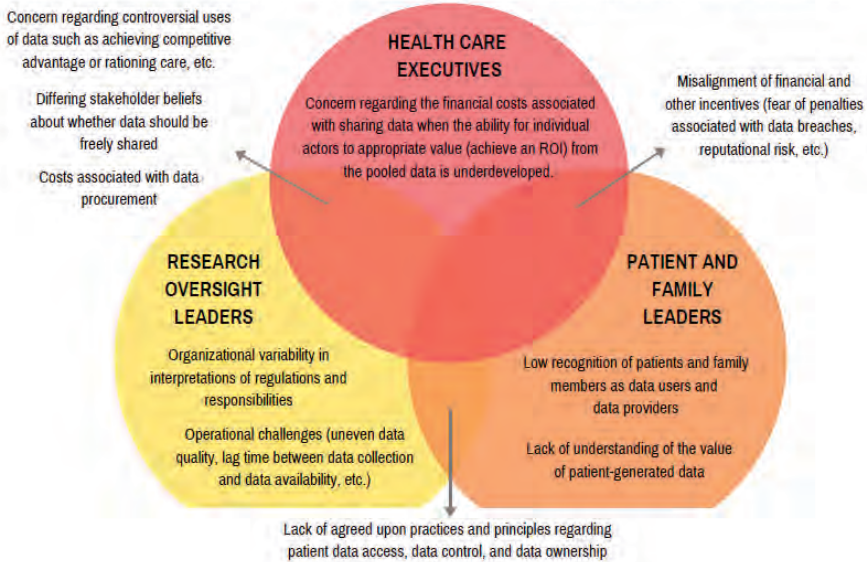
Unquestionably, the COVID-19 pandemic has had a catalytic effect on this trend. The criticality of swift data exchange became evident at the outset of the pandemic, when the scientific community sought answers about the novel SARS-CoV-2 virus and emerging disease. Then, as the crisis intensified, data sharing graduated from a research imperative to a societal one, with a clear need to urgently share and link data across multiple sectors and industries to curb the effects of the pandemic and prevent the next one.

In spite of these evolving attitudes toward data sharing and the ubiquity of data-sharing partnerships, barriers persist. The practice of health data sharing occurs unevenly, prominent in certain stakeholder communities while absent in others. A stark contrast is observed between the volume, speed, and frequency with which health data is aggregated and linked—oftentimes with non-traditional forms of health data—for marketing purposes, and the continuing challenges patients experience in contributing data to their own health records. In addition, there are varying levels of data sharing. Not all types of data are shared in the same manner and at the same level of granularity, creating a patchwork of information. As highlighted by the gaps observed in the haphazard and often inadequate sharing of race and ethnicity data during the pandemic, the consequences can be severe—impacting the allocation of much-needed resources and attention to marginalized communities. Therefore, it is important to recognize the value of data sharing in which stakeholder participation is equitable and comprehensive— not only for achieving a future ideal state in health care, but also for redressing long-standing inequities.

Prior to the COVID-19 pandemic, the National Academy of Medicine (NAM), in consultation with the Patient-Centered Outcomes Research Institute (PCORI), undertook an initiative committed to this vision. During the fall of 2018 through the summer of 2019, the NAM held a series of convenings involving three stakeholder groups—patients and family leaders, researchers and research oversight leaders, and health care executives—to identify the barriers to data sharing that each of these groups experienced or perceived. The culminating Special Publication, *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*, elucidated the list of most pressing barriers, shown in *Figure*



## CULTURAL, ETHICAL, REGULATORY, AND FINANCIAL BARRIERS TO DATA SHARING, LINKAGE, AND USE



**FIGURE 1** Prioritized Cultural, Ethical, Regulatory, and Financial Barriers to Data Sharing, Linkage, and Use

SOURCE: National Academy of Medicine. 2020. *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust*. D. Whicher, M. Ahmed, S. Siddiqi, I. Adams, M. Zirkle, C. Grossmann, and K. L. Carman, editors. NAM Special Publication. Washington, DC: National Academy of Medicine.

1 (Whicher et al., 2020).

Underpinning these concerns is the lack of trust among and within the three stakeholder groups. “The patient and family community does not trust that health care systems and researchers will make data and the conclusions based on the data available to them and will not misuse their data by rationing care and sharing with unauthorized third parties. Researchers have a similar mistrust in the intentions of third-party users. Meanwhile, health systems worry that patients will misinterpret data or use data inappropriately, such as allowing it to be combined with other elements and rendering the data identifiable. Health systems are also reluctant to share data with industry partners for fear of losing their competitive advantage” (Whicher et al., 2020).

Authors of the progenitor publication (Whicher et al., 2020) also coalesced around a set of action items that could be taken in the near-term (1-3 years) to begin addressing these issues. They include:

- Building a consortium of organizations committed to data sharing to help mobilize stakeholders around the idea of sharing.
- Identifying priority use cases and data sharing exemplars to demonstrate how barriers could be overcome.
- Reframing the risk discussion or business case related to data sharing to highlight evidence-based arguments about the risks of not sharing data.
- Engaging in a national dialogue with various stakeholder groups about the benefits of bidirectional data exchange and the current barriers to accessing and contributing data. A component of the national dialogue would be to prepare and empower various stakeholder groups to meaningfully participate in health data sharing.

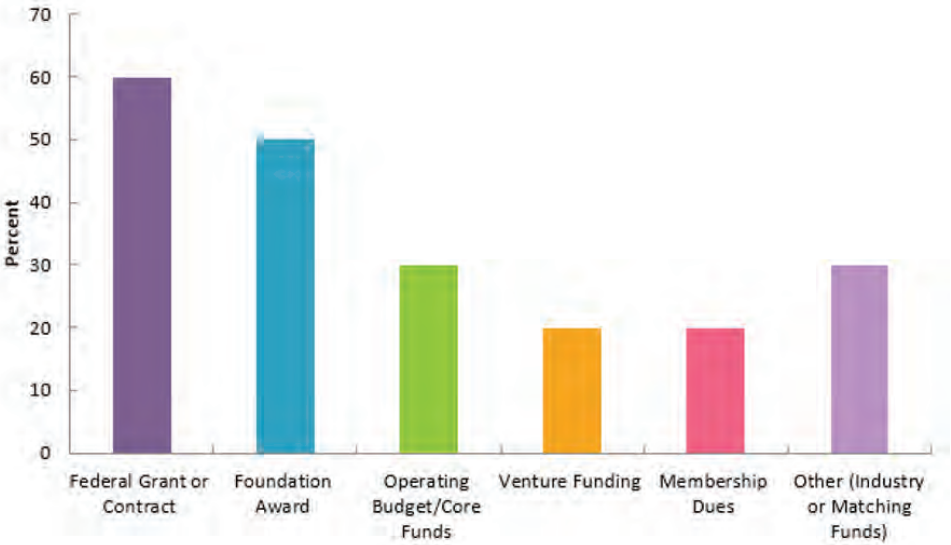
From this list, the authors concluded that the most fruitful next step would be to develop a compilation of case studies of successful health data sharing across different stakeholder groups with the intent that this resource could serve as the basis for informing and catalyzing work on the other aforementioned priority action items. In addition, the prospect of a case study compilation was independently raised in several other NAM forums, including National Academy of Medicine's Leadership Consortium: Collaboration for a Value & Science-Driven Health System's Action Collaboratives (NAM, 2020a; NAM, 2020b).

This companion Special Publication consists of a series of 11 case studies that illustrate diverse approaches to data sharing with the aim of responding to the most pressing issues detailed in the progenitor publication, *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*. In consultation with PCORI, case study candidates were selected, drawing upon specific exemplars identified in the progenitor publication and supplemented by more contemporary efforts related to the pandemic. Additional consideration was given to geographic and organizational diversity and to ensuring that the full complement of case studies reflected a spectrum of data-sharing interactions anchored to the three stakeholder groups spotlighted in the previous publication (i.e., patients and families, health care execu-

Case Study	Geographic Scope	Organization Type	Primary Stakeholders
Blue Cross Blue Shield of North Carolina	Regional	Private, for profit	Payer + Health systems
COVID-19 Evidence Accelerator	National	Public	Researchers + Health systems
Louisiana Public Health Institute / Greater New Orleans Health Information Exchange	Regional	Public-private partnership	Community health groups + Health systems
Luna	National	Private, public benefit corporation	Consumers + Researchers + Private industry
Mayo Clinic - Google	Regional	Private health system	Health system + Technology company
National COVID Cohort Collaborative	National	Public	Researchers + Health systems
OpenNotes	International	Global movement with organizational home in academic health system	Patients + Health systems
Sanford Health Data Collaborative	Regional	Private health system	Health system + Researchers
University of Michigan	Regional / National	Public university	Researchers + Patients
Vivli	National	Not-for-profit	Researchers + Biopharma + Tech companies
Yale Open Data Access Project	National	Organizational home in academic health system	Researchers + Biopharma companies

**TABLE 1** Case Study Characteristics

tives, and researchers and research oversight leaders, as detailed in *Table 1*). Nonetheless, the authors acknowledge that the health data-sharing ecosystem is vast and diverse, involving many other stakeholder groups not highlighted in this publication.



**FIGURE 2** Funding Sources for the Data-Sharing Initiatives

While the individual case studies demonstrate different data sharing use cases, collectively, they address the breadth of priority issues (see *Figure 1*) and exemplify data sharing as the linchpin to achieving each of the entities’ immediate strategic goals and consequently their ability to improve health, health care, and health equity.

Case study narratives were constructed from interviews conducted by NAM staff with leaders of these organizations in fall and winter of 2020. NAM staff used a semi-structured interview guide (see *Appendix A*) to solicit responses about the impetus for the collaboration; contextual factors giving rise to the opportunity for and success of the data sharing arrangement; barriers that were overcome and ones that endure; details about the type and level of data shared, governance model, and technical infrastructure; and advice for the field. Interviewees were forthcoming in their responses about the challenges as well as the positive outcomes of their work. In the process of developing each case study narrative, the editorial team remained attuned to the paramount importance of providing a balanced summary of each group's work. While enthusiasm about achievements may be well-placed, the insights that arise from un-

Understanding the barriers is pivotal to progress. Hence, barriers are carefully elucidated for all profiled entities.

Following the interviews, a survey was administered to collect information about each initiative's operations and funding. Each narrative includes an "At a Glance" sidebar showing the summary of barriers addressed and insights for the field.

The featured entities vary in size, longevity, and financing mechanisms (see *Figure 2*). Many of the interviewed organizations were funded by more than one source; thus, the sum of the percentages in *Figure 2* exceed 100%. They also differ in their approach to data sharing. While each organization may have taken a different tactic based on a different motivating rationale, the examples point to how organizations, with an intrepid spirit, can best collaborate to share and link data while overcoming obstacles and addressing reservations about data sharing. The editors of this Special Publication hope this compendium of case studies proves to be an accessible reference for the field and helps to cultivate the will and trust for data sharing.

## 2

# CASE STUDY: OPENNOTES

*Interviewees: Catherine M. DesRoches, DrPH, Executive Director;  
and John Santa, MD, MPH, Director of Dissemination*

### ABSTRACT

Trust and communication are the cornerstones of the patient-provider relationship. OpenNotes is a vendor-agnostic, international movement that strives to enhance this patient-provider dynamic by encouraging clinicians to provide patients electronic access to their clinical notes. The program pursues its goals through demonstration projects, research, and patient awareness and advocacy. One of its focus areas is bidirectional data sharing through the “OurNotes” project, which creates a shared space for patients to not only view their physician’s clinical notes, but contribute their personal health data to them. Since its founding in 2010, OpenNotes has attained a wide reach, with 266 participating health systems across the U.S. and Canada (OpenNotes, 2020). Despite growing evidence of the value of patient data sharing, OpenNotes faces concerns from pockets of the health care community related to issues of workflow, liability risk, and market competition. The program has tried to overcome these barriers by relying on influential voices in the health care community to champion the effort and by generating demand among patients. The OpenNotes movement is one of several campaigns aimed at enhancing patient data access that has amounted to several new federal efforts, notably Patient Access Rules with origins in the Centers for Medicare & Medicaid Services (CMS) and Office of the National Coordinator for Health Information Technology (ONC) (CMS, 2000). Implemen-

### *Case Study at-a-Glance: OpenNotes*

*Movement designed to encourage clinicians to provide patients with electronic access to their clinical notes*

#### **Key Barriers Addressed**

- Lack of dialogue between people, patients, and clinicians in developing an appreciation and understanding of consumer health data uses
- Improved safety and accuracy in health and health care through patients' familiarity with their personal health data

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Engage locally with clinician and health system leadership to identify chief concerns about sharing (e.g., liability, the perceived burden on the health care teams, technical barriers) and invite patients into those discussions
- Conduct research and engage in advocacy and to provide additional insights into the benefits of sharing

#### **Insights for the Field**

- Leverage growing awareness of the importance of transparency to encourage patients' agency over their data
- Partner with internal champions within a system to help overcome resistance and enable tailoring of change management messaging to local concerns

tation, compliance, and enforcement of these new rules are being observed closely throughout the health care industry. However, OpenNotes suggests that financial incentives for organizations and providers could help to advance patient data access.

## BACKGROUND

OpenNotes is a movement that encourages clinicians to routinely share clinical notes with their patients. While the Health Information Portability and Accountability Act (HIPAA) Privacy Rule of 1996 guarantees patients the right to obtain copies of their medi-

cal records, access by and large has been limited and cumbersome for patients, improving moderately over the last decade due to ONC Meaningful Use requirements (HHS, 2015). Nevertheless, some health care entities continue to invoke HIPAA privacy rules as rationale to constrain access to medical records, claiming that this access might overwhelm patients with information and, thus, providers with patient inquiries. There is an added concern that possible documentation errors could lead to increased liability risk for providers. However, studies have not supported these fears. In fact, opening access to medical records has shown to give patients a sense of agency in their care, with the possible benefits of bolstering patient engagement, improving health outcomes, and reducing medical errors (Bell et al., 2020; DesRoches et al., 2019; UIC, 2018).

OpenNotes began as an exploration of these benefits and an effort to correct the imbalance in data sharing. OpenNotes' co-founders were troubled by the observation that, despite the ubiquity of data sharing in the industry, patients themselves had limited access to their own records. Drawing from the co-founders' early work in patient engagement, a pilot study was launched in 2010. The study, which was funded by the Robert Wood Johnson Foundation, involved 105 providers from Beth Israel Deaconess Medical Center, Geisinger Health System, and Harborview Medical Center inviting 20,000 of their patients to view their notes electronically. The study's results validated much of the prior research and contradicted provider assumptions that the practice of sharing notes is burdensome and unduly causes worry or confusion among patients.

Since this initial demonstration, OpenNotes has diversified into the mental health space in a push for the sharing of all notes with exceptions when safety/harm or privacy are an issue. Psychotherapy notes are unique in that they often contain information about the practitioner's reactions and process during therapy sessions (e.g., countertransference) and often exist outside of the medical record. For this reason, they are afforded special protections under HIPAA and are not required to be shared under the 2000 federal information blocking rule (*45 Code of Federal Regulations §164.501 Final Rule*, 2003) (HHS, 2000). However, OpenNotes has made strides with the sharing of other mental health notes, finding that



patients who read their mental health notes report significant benefits and there are few drawbacks for clinicians. As a result, many health care networks have begun opening access to their mental health notes (Sun, 2014; OpenNotes, 2020; OpenNotes, 2021). In addition, OpenNotes has been championing the notion that holding back mental health notes may contribute to stigma and health care disparities.

These successes have propelled the OpenNotes movement into at least 266 health systems that, as of December 2020, offer 54.3 million patients across the U.S. and Canada electronic access to their clinical notes (OpenNotes, 2020). However, the percentage and demographic characteristics of patients who avail themselves of this opportunity is incompletely known as this is a dynamic and changing group whose characteristics differ from one health system to another. As an extension of its mission, OpenNotes pursues research with collaborators around the world, assessing the benefits of transparency in medical care. It also has launched several related initiatives, including one focused on fostering a culture of bidirectional information exchange between patients and providers called OurNotes, and another on an investigation of the utility of and challenges related to inclusion of social determinants of health data in the clinical notes exchange. A third principal area of work focuses on the potential of increased transparency to improve care safety.

Altogether, the initiatives are supported by 13 personnel. The Division of General Medicine at Beth Israel Deaconess Medical Center is the organizational home for OpenNotes. Over the last decade, OpenNotes has acquired additional funding from federal grants and foundations, namely the John A. Hartford Foundation, Agency for Healthcare Research and Quality, ONC, the Gordon and Betty Moore Foundation, the Peterson Center on Healthcare, and the Cambia Health Foundation.

## DESCRIPTION

The blueprint for an OpenNotes implementation consists first of understanding the organizational culture of a health system and gathering information about the system's electronic health record

(EHR) vendor. A critical step is to determine staff attitudes toward the EHR system (e.g., to what degree do providers experience EHR burnout) and the extent to which note-sharing occurs. Next, the Open Notes team meets collectively with the system's technical and strategic leadership (chief technical officers, chief medical informatics officers, and chief medical officers) to determine the best approach for garnering support across the enterprise. This consists of understanding which groups are opposed to the idea and their motivations for their opposition, and then crafting an approach that is responsive to their reservations. In some cases, a research-driven, fact-based approach is preferred, and in other cases, a campaign focused on the cultural issues is more appropriate for attaining stakeholder buy-in. Interviewees for this case study remarked that the hurdle of making the case for clinical note sharing to health system leadership diminished whenever representatives from patient and family advisory committees were present in these discussions. The internal culture of a service line is another influential factor of uptake.

## BARRIERS

Despite the demonstrated value of patient data sharing, risk management and workflow issues remain a concern among health care decision makers. The interviewees shared details from their encounters with malpractice insurers and providers who, despite being impressed with the evidence on the reduction in medical errors, expressed concerns about the vulnerabilities of poor notetakers. In one example shared with the NAM, a provider voiced concern that the need to correct documentation errors could increase his already strained workload.

Lack of awareness of the OpenNotes movement among the patient and caregiver communities poses an obstacle to gathering support given the role these groups play in elevating the demand for data sharing. Ongoing dissemination efforts are aimed at increasing awareness and have been boosted in light of the information blocking regulations taking effect.

Navigating the variability in state laws and cultural notions about the sharing of sensitive data such as adolescent and mental health

notes is another important barrier. Despite the advances OpenNotes has made in the sharing of mental health notes—much of it credited to the Veterans Administration (VA) Health System’s research—convincing organizations to share these types of notes requires significant effort (VA, 2017). Most participating health systems choose not to share behavioral and mental health notes. Most care delivery organizations begin by sharing ambulatory notes and then opt to begin sharing other types of notes, basing it either on individual or organizational preference. Interviewees suggested that decreasing the stigma of mental illness and broader awareness about the benefits of mental health data access could increase successful efforts toward note sharing.

## VALUE PROPOSITION AND FACILITATORS OF SUCCESS

Response from the health care community has been equally positive and advantageous to the growth of OpenNotes. A subset of chief medical informatics officers sees the movement as empowering, helping them regain a sense of autonomy as physicians and reconnect with patients in a shared decision making capacity. OpenNotes relies on champions who are well-respected in their communities to help penetrate a market. It should be noted that, at times, it takes just one person with sufficient political capital to counter the resistance and compel others into action.

OpenNotes interviewees acknowledge that EHRs have served as a catalytic force for the movement. Although health care decision makers often cite the technical cost of data sharing and usability issues related to EHRs as the reasons for not engaging in clinical note sharing, the surge in recent conversations about the ownership of health data helped solidify the value proposition for opening up EHRs to patients, ushering in a favorable cultural change to which OpenNotes is well-positioned to respond and support.

## FUTURE DIRECTIONS

One of the most consequential indicators of this culture change is the federal rule, effective on April 5, 2021, mandating that U.S. health care providers furnish patients with access to their EHR

notes free of charge (OpenNotes, 2021). The rule stems from the 21st Century Cures Act of 2016, which requires the sharing of most types of notes, including consultation notes, medication lists, and imaging and pathology report narratives (U.S Congress, 2016). Although the current regulation requires organizations to share a more limited set of information with patients, by the end of 2022, all information in the EHR will have to be shared with patients, and organizations will have to allow patients to access this information through any third-party application of their choosing. While optimistic about the anticipated effects of the rule, the OpenNotes team remains circumspect, given the exceptions that could create latitude for interpretation and enforcement. As health systems address the rule, they confront challenges with aggregating and presenting data in a patient-friendly way while balancing the liability concerns of the organization.

OpenNotes interviewees commented that along with policy levers, there is a crucial need for financial incentives, such as transparency performance metrics tied to payment or reimbursement to clinicians for time spent responding to patient emails as part of the global move to value-based care. However, until this shift occurs, the program endeavors to generate demand for data sharing through research and dissemination, which requires a steady infusion of capital. Their advice to others aiming to solve a similar problem is to focus on patient awareness and highlight issues through measurement. The body of research showing that OpenNotes does not disrupt or add to clinician workload is another important ingredient that can be applied to many other substantive changes related to health care data, dispelling one of the key concerns about health data sharing raised in the progenitor publication (Whicher et al., 2020). Although efforts to develop metrics around data sharing have gained minimal traction over the years, spotlighting the paucity of metrics and consequent inability to describe the extent of sharing could help to draw attention and curiosity to the issue.

# 3

## CASE STUDY: THE UNIVERSITY OF MICHIGAN

*Interviewees: Sachin Kheterpal, MD, MBA, Associate Professor of Anesthesiology and Associate Dean for Research Information Technology; Kayte Spector-Bagdady, JD, MBE, Assistant Professor of Obstetrics and Gynecology and Associate Director of Center of Bioethics and Social Sciences in Medicine; Brahmajee K. Nallamothu, MD, MPH, Professor in Division of Cardiovascular Diseases and Department of Internal Medicine and Co-Director of Precision Health*

### ABSTRACT

An interdisciplinary mindset and drive for a transparent approach to leveraging health data for research and precision medicine have guided the suite of activities of the University of Michigan (U-M). Activities described in this case study encompass a range of synergistic and related innovations, including the development of a data-sharing, decision-making framework; a concise and understandable informed consent pamphlet, and associated governance processes, overseen by the Human Data and Biospecimen Release Committee; and ongoing research, evaluation, and dissemination of best practices in data sharing. The leaders and faculty participating in these cross-institutional initiatives—which include the colleges of medicine, law, public health, public policy, and engineering—seek to apply policy and research advances in real time to the U-M health system, known as Michigan Medicine. This allows U-M to move toward its overarching goals of advancing critical research and supporting precision health (a population-based strat-

## *Case Study at-a-Glance: The University of Michigan*

*Creation of comprehensive policies, oversight mechanisms, and procedures responsive to research participant concerns about consent and transparency*

### **Key Barriers Addressed**

- Current regulations for protecting study participants have not kept pace with contemporary and emerging data-sharing trends
- Participants input into how their data are used in a research context

### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Develop transparent processes for governance and oversight of research data applying insights from research participants
- Promote policies that prioritize sustaining trust, rather than prioritizing data sharing and commercialization

### **Insights for the Field**

- Continue studying and improving the informed consent process
- Seek opportunities to align biorepository policies and practices across the research enterprise

egy targeted to discover genetic, environmental, social, behavioral, and clinical markers to improve health outcomes).

Guided by insights from their engagement efforts as well as a commitment to beneficence, respect, autonomy, transparency, and justice, U-M faculty and staff leaders have developed and applied best practices for governing the sharing of biomedical knowledge that goes beyond current US regulatory requirements. Michigan Medicine cares for approximately four million people in its health care delivery system and engages in a wide range of research. Therefore, its data-sharing applications vary in purpose and scope. Moreover, its dual status as both a publicly supported university and a large health delivery system underpin the motivation to continuously act in ways deserving of community trust.

Collectively, the U-M's data-sharing activities are staffed by more than 20 full-time employees and funded through a combination of federal, state, and private industry grants, university school and department operating budget funds, and foundation awards. Several interrelated activities are described in this case study to illustrate the range of operational, scientific, and bioethical considerations necessary to comprehensively guide decisions related to data use and sharing, and how insights from study participants can assist with developing organizational approaches.

## BACKGROUND

Collaborators interviewed for this case study include researchers from U-M's medical, public health, and public policy colleges who partner on many collaborative research projects and U-M Precision Health initiatives.

The foundational work supporting U-M Precision Health originated by using the participant engagement infrastructure of the NIH Clinical and Translational Sciences Award-funded Michigan Institute for Clinical and Health Research supported by the Michigan Genomics Initiative. U-M leaders invited patients and research participants to provide guidance on proposed U-M data-sharing policies. Several different themes emerged from this work, including that participants placed their trust with U-M broadly, across the various schools and colleges comprising the academic entity; they understood and supported broad data sharing for research and quality improvement within U-M, with or without explicit patient consent; and that sharing data with companies or non-academic entities raised concerns. Specifically, participants expressed discomfort with U-M sharing data with commercial companies (such as an insurance provider or a biotechnology company) without express permission. In order to address this concern, U-M ensures that data and specimens are only shared with commercial companies if the consent form explicitly discloses this. This work utilized an interdisciplinary approach to incorporate participant voices in developing a comprehensive policy approach across schools.

In addition, U-M research teams have conducted many other related qualitative, quantitative, and legal analyses of the gaps and

opportunities in this space, which have served as a foundation for their approaches and are briefly summarized in this case study. First, findings have shown that current laws and regulations do not fully protect the myriad ways that data can be generated, shared, and used in the current age, particularly from an industry perspective (Golbus et al., 2020; Price et al., 2019). Despite these regulatory limitations, focused on protections only for research participants, the vast majority of people wish to be notified if their biospecimens might be commercialized (with a minority being comfortable with such use) (Spector-Bagdady et al., 2018). This makes participant trust and institutional responsibility critical. A clear framework, context, and guidelines are necessary to ensure both timely and non-discriminatory application of machine-learning-based interventions in health care (Wiens et al., 2019). Lastly, these U-M researchers also found that participant altruism and trust in the health system and care providers were both directly associated with believing that people have an ethical obligation to allow one's health information to be used for research (Minakshi et al., 2020).

U-M collaborators used these findings to inform creation of their data-sharing policy, centering their focus beyond simply obtaining consent to ensuring that patients and providers involved reach a collective understanding of what it means to transparently, responsibly, and ethically steward data and specimens. This feedback informed an opt-in consent system for patients agreeing to share their data within its biobank, and this process has been extended for use to the broader population of patients who access health care services through Michigan Medicine.

## DESCRIPTION

The University of Michigan Medical School Office of Research also established a Human Data & Biospecimen Release Committee (Data Release Committee) to develop a decision-making framework to guide questions regarding and processes for sharing of and access to individual-level data and specimens for research (Spector-Bagdady et al., 2020). The committee consists of U-M leaders, including administrators and faculty. Collectively, committee members provide clinical, research, legal, ethical, patient-related, conflict-



of-interest, technical, and industry partnership expertise. Principles guiding the sharing and transfer of data and biospecimens are publicly viewable on the U-M Medical System website, and completion of a prescriptive checklist is required when researchers seek to share data with external for-profit organizations (UMMS, 2019; UMMS, 2018). When data are shared externally, the data use agreement stipulates that the data will only be used for research purposes and that recipients may not attempt reidentification of individuals in the dataset. The checklist differentiates among aggregated versus individual-level data as well as level of identifiability (de-identified, limited data set, fully identifiable).

To provide a sense of scale, the committee meets biweekly and uses the standardized checklist to review approximately three projects per meeting. The data-sharing proposals requiring review are identified via a variety of research process portals of entry: U-M's Institutional Review Board, Office of Sponsored Research Programs, Data Office for Clinical and Translational Research, and departmental research staff. Since its establishment in 2019, the committee has reviewed more than 250 projects. Fifteen projects were rejected due to inadequate disclosure during initial consent or unacceptable terms of use from the industry partner. The interviewees for this case study, recognizing that they were undertaking novel interdisciplinary research that not only informed institutional policy but could benefit others similar institutions, also published their approach to sharing data with industry collaborators (Spector-Bagdady et al., 2020).

To create a consistent and standardized approach to evaluating data-sharing proposals, the Data Release Committee uses a rubric that considers the rationale, scope, data elements, participant consent/authorization, data recipients, and other facets of the data-sharing request. Additionally, U-M's related data-sharing policies provide a roadmap for researchers to ensure compliance, beginning from the planning for a data-sharing request to post-approval steps by the Data Release Committee (UMMS, 2019). Notably, the rubric covers both retrospectively and prospectively collected data and does not "grandfather in" specimens collected before the new protections of the 2018 revised Common Rule which now stipulates that informed consent forms must include a statement informing

the research participant that their biospecimens (even if de-identified) may be used for commercial profit and asking whether the participant will or will not share in this commercial profit (45 CFR § 46.116[c][7]). The review is focused on transfer of individual patient data, whether identifiable or not, to for-profit or nonprofit companies, foundations, medical specialty societies, or nongovernmental agencies (not to other academic institutions).

Interviewees described the value proposition for this work as more of an underlying motivation to act in trustworthy ways through principled treatment of data and specimens, an approach they further delineated in a 2020 *New England Journal of Medicine* article, “Sharing Health Data and Biospecimens with Industry — A Principle-Driven, Practical Approach” (Spector-Bagdady et al, 2020). As researchers, they are keenly aware that, given U-M’s position as a public research institution, they are accountable to the citizens of the state of Michigan, and that transparency is imperative. They further acknowledged and seek to learn from past incidents in which patient trust was damaged as a result of data breaches or intentional data sharing without patient consent. They also seek to contribute to the evidence base on engagement and data sharing while simultaneously using that evidence to shape policies and procedures. In essence, acting in good and transparent ways worthy of trust, rather than facilitating rapid data sharing and commercialization, is their guiding imperative. U-M’s concrete policies bolster this imperative, including the Policy for the Transfer of Human Data & Biospecimens to Industry and Non-Academic and Non-Governmental Entities, data-sharing agreements, and an approval process overseen by the Medical School Human Data & Biospecimen Release Committee (UMMS, 2019).

In addition to the fact that they are working at a state university, the U-M researchers have elevated the integral importance of transparency in all of their biomedical research. This is particularly evident in their work to develop a customizable template for a brief, easy-to-understand informed consent pamphlet<sup>1</sup> for participants in studies that use U-M’s biorepository, including the Michigan Genomics Initiative. As one interviewee put it, “If you need 16 pages to explain it, you’re not explaining it well.” The

<sup>1</sup> The pamphlet template is currently available in English at [https://research.medicine.umich.edu/sites/default/files/resource-downloadres\\_irbmed\\_biorepositories\\_pamphlet\\_template\\_20181106.pdf](https://research.medicine.umich.edu/sites/default/files/resource-downloadres_irbmed_biorepositories_pamphlet_template_20181106.pdf).

pamphlet template uses consistent and unequivocal language to convey that all researchers in U-M will have access to a patient's data once they consent, and, if data or specimens may be commercialized for profit, participants will be notified during the informed consent process. Further to this, if participants are not notified of potential commercialization during the informed consent process, with a few rare exceptions, U-M will not share their data or specimens with commercial companies. In some exceptional circumstances where the data or specimens are of particular value, such as for rare or orphan diseases, and it is feasible to contact the original contributor, re-consent may be sought.

An important barrier described by interviewees is the current legal and regulatory climate. At present, regulations for the protection of human participants do not cover contemporary or emerging data sharing trends, especially as sources of health data evolve to include data collected by entities outside the purview of the Health Insurance Portability and Accountability Act (HIPAA). While HIPAA provides clear guidance for sharing identified data collected by covered entities and prescriptively defines a limited data set, the 2018 revision to the Common Rule portion of the Human Subjects Research regulations may put protections for data and biospecimens collected from research participants versus patients at odds. For example, as noted above, the new Common Rule includes regulations regarding disclosure of how biospecimens will be used (45 CFR § 46.116[c][7]). But there is no such disclosure requirement under HIPAA for clinical patients. However, the U-M interviewees asserted that current regulations are “the floor, not the ceiling of protections,” and espoused the core belief that academic medical centers should establish best practices for governing the sharing of data and biospecimens with outside entities that go above and beyond current regulatory requirements. These concerns are particularly true for digital health data collected through software applications outside the classic health care provider role.

## FUTURE DIRECTIONS

The U-M team hopes to continue its efforts promoting transparency, trust, and consent in its data sharing and research efforts,

and expand its research pursuits to include dialogue and participant engagement in discussions about the balance of individual risk and benefit relative to societal benefits of biomedical research. Future aspirations include better managing the limitations of written informed consent as the primary indicator of engagement and communication, reducing disparities created by demographic (especially race and ethnicity) biases in the recruitment and consent processes, and involving patients as partners on the Data Release Committee. Another area is to continue to engage researchers who use the services of U-M Precision Health to ensure their data collection and sharing practices align with their use of the biobank. By doing so, these policies have the ability to influence the broader researcher communities across campus.

In addition, the interviewees offered advice and encouragement to other organizations interested in following a similar approach. Chiefly, they urged that researchers should focus on participant engagement in what institutions should do instead of just disclosures of what institutions will do. While the human subject research regulations require many types of disclosure, they provide comparatively sparse guidance about the informed consent conversation itself, including optimal practices and essential considerations for the consent process. Finally, the team emphasized that transparency and engagement are meaningful and useful efforts for researchers to undertake, as these aspects support the creation of sustainable relationships with patients, participants, families, and communities. U-M's research and experience indicate that greater engagement can spur broader, deeper trust that is able to be extended from a single project to an entire research enterprise.

## 4 CASE STUDY: LUNA

*Interviewees: Dawn Barry, MBA, President and Co-founder; and Scott Kahn, PhD, Chief Information and Privacy Officer*

### ABSTRACT

Luna is a member-owned genomic and medical research platform developed by the public benefit company LunaPBC. LunaPBC seeks to redefine the relationships between people, communities, industry, and researchers. LunaPBC envisions that people and leaders can come together as a problem-solving community-of-communities and use their data as a new way to address disease and improve quality of life. Luna is structured such that anyone in the world can join the platform, share their health data for research, and U.S. residents can take ownership shares in Luna. The leadership team was motivated to establish Luna to address the challenge that most patient data collected for research does not include the voices and lived experiences of individuals. It also becomes part of institutional data silos, rendering it largely inaccessible, and the individuals who contributed their data have virtually no way to guide how it is used beyond its original purpose. Founded in 2017, Luna is funded through biotechnology companies and venture capital organizations. Approximately 35 patient advocacy organizations and affinity groups host their communities on Luna as of October 2020. Luna has designed its platform and processes to better support connections between research participants and researchers and give participants direct control over how their data are used.

### *Case Study at-a-Glance: Luna PBC and the Luna Platform*

*Vanguard organization motivated to maximize the use of data for research while enabling people to decide how their data are used and receive compensation for contributing data*

#### **Key Barriers Addressed**

- People/patients maintaining control over their health data
- Broader availability to help research and data achieve their fullest potential

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Create a business model that offers incentives to patients for sharing data
- Commit to GDPR and CCPA principles to ensure data privacy and security
- Give consumers/patients complete agency over how their data may be shared for research

#### **Insights for the Field**

- Create a virtuous cycle where increasing numbers of people understand the value and utility of their personal health data for helping others
- Use a platform-as-a-service model to reduce the need to recreate infrastructure

## BACKGROUND

Luna is a platform created by the public benefit company LunaPBC to support data sharing between individual consumers and researchers. LunaPBC serves as the management company for this platform. Ownership shares are available to Luna members based on the types and amount of health data that are shared for research; all shares in Luna are owned by members who have shared their data for research. As profits are generated from the research conducted in the platform, shareholders in Luna are entitled to dividends (cash distributions) in proportion to how many shares they

own. The creators of Luna sought to encourage privacy-protected health data sharing by consumers, demonstrate a new model of consumer data transparency and oversight, make the process of contributing data straightforward for consumers, allow consumers to maintain control of their shared data, and make the data as accessible as possible to the research community.

The underlying philosophy is data democratization by giving participants control over how their research data will be used, which is further undergirded by stringent commitment to data privacy, security, and compliance laws—chiefly the European Union’s General Data Protection Regulations (GDPR) and the California Consumer Privacy Act (CCPA). In creating the Luna platform, the founders were motivated by what they saw as fundamental flaws in the research ecosystem, particularly that many studies do not attain a sufficiently diverse population, and that research dollars are effectively wasted if data are used for one study, when they could answer additional questions. The value proposition is similar for both people sharing their data and the research community, in that it allows everyone the opportunity to “put health information to work for good.” Offering a simple, secure way to contribute to health science and participate in communities of support is an additional way that Luna is seeking to create value.

## DESCRIPTION

Luna as a platform is designed to ingest and manage numerous types and formats of data, enabled by application programming interfaces (APIs), which make it possible for different applications to talk to each other. The types of data that can be uploaded to the platform are dynamic and evolving. At this writing, the capabilities included DNA files from 23andMe, Ancestry.com, and other direct-to-consumer DNA companies, as well as patient-reported outcomes, such as Luna-generated health surveys and validated survey instruments. Information from electronic health records (EHRs) can also be supplied to the Luna platform via patient portal integrations for continuous data flow from the EHR to the platform. At a researcher’s or community’s request, Luna can bring in additional data types, such as whole genome and exome DNA files,

RNA, microbiome, fitness/activity trackers, smart devices, and medical devices. Automated quality assurance protocols are built into the process of data ingestion and submission, based on the type of data (i.e., genomic data, survey data, registry data). For example, genomic data markers are well described and transparently reported in public databases. Thus, genotypic data uploaded by an individual may be compared against the expected type and structure of data from each direct-to-consumer DNA vendor.

Once health data are uploaded by a consumer and reviewed for quality and sufficiency, they are de-identified, encrypted, and aggregated by the Luna platform. From that point forward, the person is a *member* of the platform. Researchers from academic institutions and pharmaceutical companies are the typical “customers” for the data—they pay for services associated with and access to the platform and are only permitted to use the Luna platform for approved research studies. The Luna platform itself is Institutional Review Board (IRB) approved, and proof of approval for a given study by an IRB or comparable research oversight body is required as a prerequisite to data access. For instance, an academic researcher would gain approval from their home institution, including approval to use the platform, before the research study could proceed. All data analysis is performed within a secure computational “workbench” within the Luna platform to maintain a member’s control over their shared data. Comprehensive terms of use and an accompanying privacy policy function as a contract, and explain how data contributors and those who access data are expected to comply (LunaDNA, 2020a). The community is dynamic with respect to numbers of active studies and collaborators. At the time of this writing, Luna includes approximately 50 member communities active on the platform. Member communities range from 50 to more than 20,000 active participants.

Researchers initiate collaboration with Luna via direct outreach to LunaPBC or through an online inquiry link. Collaborations can advance pending internal review by LunaPBC staff, the researchers’ provision of IRB approval or exemption, and acceptance of the terms of use and associated policies. Review typically takes one-to-three business days, depending on the nature of the study. Not-for-profit research is made available at discounted rates in line



with the LunaPBC corporate charter to advance improvements in human health and quality of life. Research in pursuit of commercial activities and the creation of intellectual property are priced accordingly.

Luna's transparency and oversight architecture centers on recognizing individuals' data as currency and compensating them for sharing personal genomic and health data for research. This is achieved by issuing shares based on the type and volume of data being shared within Luna. Per Luna's SEC filing circular, 714,528,714 shares are available and are valued at \$.07 each, for a total of \$50 million (LunaDNA, 2020b). Approximately 285,000 shares had been issued as of November 2020. The ability to attract interest on the part of both consumers (data contributors) and the research community is an integral pillar and key dependency for this value proposition to succeed. The choice to establish LunaPBC as a public benefit corporation to operate the Luna platform was fundamental in the eyes of Luna's founders, in that it creates a social contract with the shareholders.

For researchers to use the Luna platform, they must have IRB approval (or the equivalent outside of North America). Study design guidance is available through a partnership with Genetic Alliance, using the Alliance's 30 years of experience advocating for and conducting research with disease and community-led patient groups. Luna's ethos is "people-centered governance," which is consistent with its adherence to and embrace of the European Union's GDPR regulations: people have oversight and agency over their data and how they are used.

The Luna model addresses a key barrier noted in *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*, namely that there is a lack of widespread understanding of the value of patient-generated data. Luna's leaders cited a growing body of bioethical literature on personal agency and control over how one's data is used, and the desire among consumers for reciprocity and transparency. They built the Luna platform using privacy-by-design in direct response. Development and launch of Luna followed a timeline similar to GDPR (passed in 2016, implemented in May 2018) and the 2018 CCPA; privacy-by-design was "very influential" in Luna leadership's decision to create Luna as

a public benefit corporation (PBC). Data from more than 180 countries are currently shared via the Luna platform. The founders note that their Securities and Exchange Commission (SEC) filing marks the first time that data are viewed as currency, as is the implementation of a research platform that treats personal data privacy as a fundamental right of an individual (Kain et al, 2019).

While discussing sociotechnical considerations of this model, the interviewees referenced the philosophical differences between surveillance and participation, and the positive connotations of providing data for participation in research—contrasted with companies owning data on an individual that the individual does not control. Luna uses a technology that creates pseudonymous records for each member that knit different data types together. Personal identifiers are separated from the shared research data while the platform maintains the ability to crosswalk different data sources as they are dynamically shared by each member. Two-factor identification is used to enhance data security to support data privacy control, recognizing that control of an individuals' lived experience over time is paramount and will affect one's inclination to participate in a given research study. Approval of one's data use is granular to the level of study: the members themselves choose studies in which they will enroll. If a member is not enrolled in a study, the researcher cannot access their data for the purposes of executing that study. In short, trust is central to Luna's social contract.

With respect to uptake of the platform, outreach for research participation occurs via email to registered participants. The majority of members have participated in more than one study. Contributing members can withdraw their data at any time. If a member joins a registry/community or study, the community leader and study PI can use a recontact agent embedded in the platform to contact the member while preserving the member's anonymity. In this way, the member can be invited to new studies. The member also has the option to allow any researcher to contact them (and thus invite them to other studies) even if they did not join their community or study already. Inversely, members can restrict engagement to only a specific registry.

## FUTURE DIRECTIONS

Luna is one of several companies developing a model of remunerating individuals for sharing access to their health data for research. As such, its leaders urged that the research community—both academia and technology companies—should keep evolving with consideration of “modern” data privacy gestalt. It is common to want to hold onto a historically entrenched perspective with respect to data ownership, but new models and innovations related to data control and data valuation are needed. They observed that data exchange platforms and partnerships on Amazon and other technology behemoths are proliferating, while universities are deliberating approaches to monetizing and licensing as it relates to health research data (Pew Charitable Trust, 2021). Luna’s leaders urged that academic systems leverage preexisting infrastructure rather than creating their own, and that decision-makers should “lean in” and learn from historic examples of health data being misused or used without consent. A 2021 survey of perceptions of sharing health data showed growing support for sharing of data with clinicians, portending that the terrain will change for research as well.

When asked about one thing that could be changed at will (e.g., with a “magic wand”), the Luna President noted, “We know digital, remote, longitudinal trials are the way to go, but we still have natural history studies in pediatrics where patients have to fly to a clinical site to be studied, which is extremely burdensome and, in some case, dangerous.” She noted that considering how to collect robust data while keeping people safe at home can be accomplished with the right infrastructure and participation platform, including deployment of in-home technology and assurance of proper oversight and participant protections. This type of platform, then, is of the essence, particularly as stakeholders in health and health care seek to accelerate patient-centered, real-world effectiveness research.

# 5

## CASE STUDY: COVID-19 EVIDENCE ACCELERATOR

*Interviewees: Carla Rodriguez-Watson, PhD, MPH, Director of Research, Reagan-Udall Foundation; and Jeff Allen, PhD, Executive Director, Friends of Cancer Research*

### ABSTRACT

The COVID-19 Evidence Accelerator (EA) is a research collaborative comprised of 230 public and private entities spanning the health care ecosystem. The project was launched in April 2020 at the behest of the U.S. Food and Drug Administration (FDA) to understand the characteristics of the novel SARS CoV-2 virus and clinical progression of COVID-19, and to provide rigorously developed evidence regarding the diagnosis, treatment, and prevention of the disease. To that end, the EA provides a unique venue for participants to crowdsource real-world evidence (RWE), instead of data, which is the focus of many case studies in this publication; investigate a prioritized list of scientific questions; and compare results without researchers having to expend capital to procure and maintain external datasets. The added benefit of this approach is that researchers can exchange knowledge and ideas corroboratively without exposing themselves to privacy breaches or dealing with consent or control issues with data—concerns raised by the research community in the progenitor publication to this one. Given the urgency to attend to the rapidly evolving COVID-19 crisis, EA used existing information-sharing infrastructure, research protocols, and the prestige of its host organizations, the Reagan-Udall Foundation and Friends of Cancer Research, to quickly attain

## Case Study at-a-Glance: COVID-19 Evidence Accelerator

*Voluntary collaboration to share data between researchers catalyzed by the urgent needs of the COVID-19 pandemic*

### Key Barriers Addressed

- Operational and cost-related challenges of procuring, maintaining, and safeguarding data
- Privacy concerns that arise from sharing individual-level data
- Lag time from when data are collected to when they're available and usable

### Specific Solutions for Data Sharing between Patients and Clinicians

- Use Master Protocols and adopt an extant infrastructure to enable collaborators to undertake rapid analytics
- Create a code of conduct and collaboration principles to underpin a “coalition of the willing”
- Leverage an urgent crisis to create shared volition to address population health needs

### Insights for the Field

- Embrace the importance of a sustainable infrastructure that removes friction from data and knowledge sharing
- Avoid “reinventing the wheel” if existing tools, protocols, or resources can be adopted rather than developed *de novo*

critical mass and jumpstart its work. Work within the EA is guided by a code of conduct that emphasizes a commitment to scientific integrity, ruthless transparency, and respect for individual privacy. While the EA began with workstreams in COVID-19 therapeutics and diagnostics, it has recently expanded to COVID-19 vaccines and post-acute sequelae of SARS-CoV-2. It is also being leveraged to explore possibilities for the development of novel treatments in substance use disorder. Nonetheless, irrespective of use case, the EA's approach to information sharing offers a model for how concerns about privacy risks and loss of competitive advantage that often stymie data sharing efforts can be addressed.

**BOX 1***Defining Real-World Data and Real-World Evidence*

**Real World Data (RWD)** are data relating to patient health status and/or the delivery of health care that are routinely collected from a variety of sources, including electronic health records, claims and billing, patient-generated data, and mHealth data.

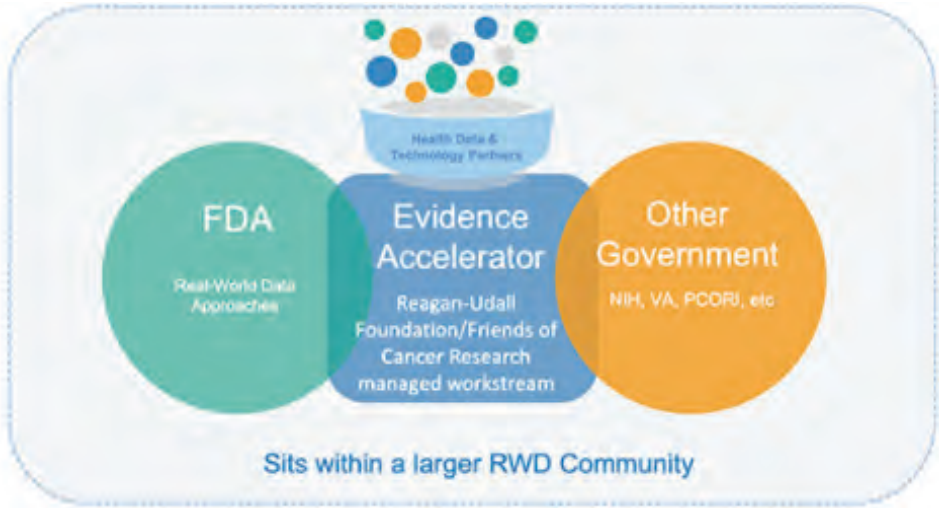
**Real World Evidence (RWE)** are results extracted from the synthesis of RWD that are used to inform the potential benefits or risks of a medical product or efficacy of a course of clinical action.

SOURCE: FDA (Food & Drug Administration). 2020. Real-World Evidence. Available at: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (accessed May 28, 2021).

**BACKGROUND**

The COVID-19 Evidence Accelerator is a multi-stakeholder collaborative created at the request of the U.S. FDA in April 2020 to understand and address the rapidly developing COVID-19 pandemic through the sharing and leveraging of RWE (FDA, 2020). The EA is managed through a partnership between the Reagan-Udall Foundation (Foundation) and Friends of Cancer Research (Friends). Established by Congress as an independent 501(c)(3), the Foundation uses its neutral position to facilitate dialogue between the FDA and other public and private entities. Comparably, Friends is a non-profit think tank that seeks to accelerate the discovery and development of new cancer treatments through public-private convenings and research partnerships.

At the time of writing, the EA consists of 230 participating organizations engaged in any or all of its three workstreams. The Diagnostic Evidence Accelerator workstream focuses on addressing diagnostic and serological questions, such as those related to the real-world performance of COVID-19 diagnostic tests. The Thera-

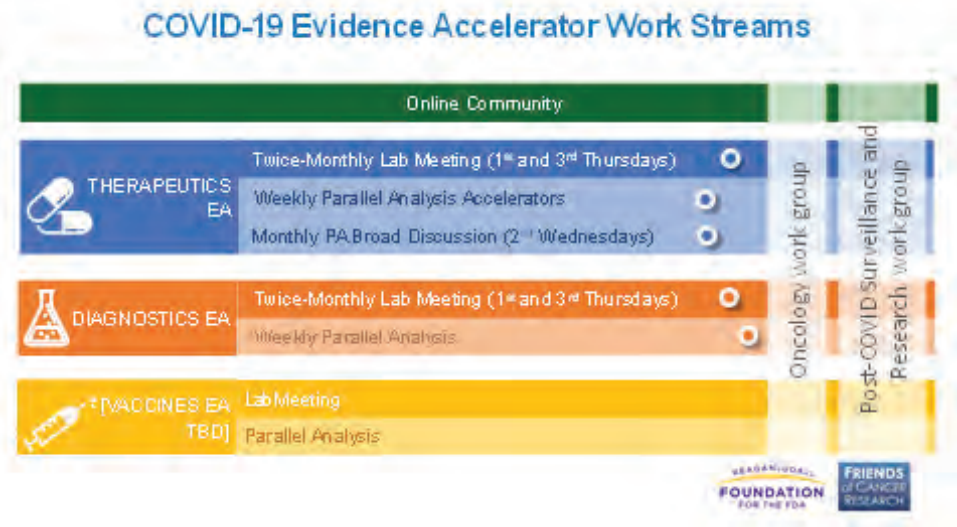


**FIGURE 3** Real-World Data for COVID-19 Ecosystem

SOURCE: Roe, L., A. Abernethy, J. Franklin, E. Sigal, J. Allen, A. Bhat, C. Rodriguez-Watson, and S. Winckler. 2021. Accelerating Evidence Generation by Convening Diverse Stakeholders Across the Real-World Data Ecosystem. Available at: <https://evidenceaccelerator.org/sites/default/files/2021-08/Ev%20Acc%20Lab%20Focus%20Science%20Poster%202021%20Final%20with%20OCC%20review.pdf> (accessed December 6, 2021).

peutic Evidence Accelerator workstream is devoted to expediting the identification of effective therapies for mitigating COVID-19 symptoms. The Vaccine Evidence Accelerator focuses on questions of vaccine performance. Representation in the collaborative spans the health care ecosystem: health care systems, national insurers, health care technology companies, pharma and biopharmaceuticals, laboratories, academics, and various branches of the federal government engage and share their expertise within the EA. Participating organizations range in size, have access to diverse sources of data, and serve diverse populations.

In addition, the EA operates within a larger national data community by interfacing with FDA-adjacent initiatives, such as the Sentinel Program, Biologics Safety and Effectiveness System, and the National Evaluation System for Health Technology; domestic activities, such as the Patient-Centered Outcomes Research Institute (PCORI); and interacts with international forums, such as the



**FIGURE 4** COVID-19 Evidence Accelerator Work Streams

SOURCE: Reagan-Udall Foundation and Friends of Cancer Research. 2021. COVID-19 Evidence Accelerator. Available at: <https://evidenceaccelerator.org> (accessed November 26, 2021).

Observational Health Data Science and Informatics program (see *Figure 3*).

At its inception, the EA was not externally financed, but in July 2020, the project secured support from a federal grant and a private foundation. As activities of the EA have expanded, the EA is seeking additional funding that will support additional staff to advance new projects.

## DESCRIPTION

The EA convenes experts via Lab Meetings to rapidly share information related to real-world studies of COVID-19. During the Lab Meetings, participants share preliminary findings on research, data analytics, and methods relevant to addressing the COVID-19 pandemic in three discrete areas (see *Figure 4*). In addition, findings from parallel analyses (described below) are conveyed to this broad community. EA Lab Meetings provide a “safe collaborative



space” for key players (nearly 300 organizations, at the time of writing) across the health data ecosystem to assimilate and evaluate data generated from across the country. Given the importance of information sharing in the wake of the COVID-19 pandemic, meeting summaries are posted on the EA website (Reagan-Udall Foundation and Friends of Cancer Research, 2021).

The second main EA activity is gathering various data analytics experts to explore discrete research questions across a variety of real-world data sources (e.g., insurance claims, EHRs, registries). In the parallel analysis approach, analytic partners (Accelerators) contribute to the rapid development of master protocols to illustrate the use of various treatments or diagnostics, as well as characterize the natural history of certain components of COVID-19. Some of these protocols, particularly those around several potential COVID-19 therapeutics, have been reported on EA’s website and are hosted on a cloud-based file storage site (Reagan-Udall Foundation and Friends of Cancer Research, 2020). Master protocols enable parallel analyses of the same question across different data systems to quickly test reproducibility of results. Unlike most of the cases highlighted in this Special Publication, data exchange is not a feature of the EA. Hence, there is no need for a data coordinating center or data use agreements (DUAs) with research partners, although research partners who engage in data sharing external to the EA may employ DUAs. Each analytic partner applies the master protocol to their own data and performs their own analysis. Therefore, only aggregated results are shared among collaborators for knowledge generation. The benefits of this method are that it mitigates privacy risks and addresses concerns about the loss of competitive advantage that can arise from sharing identifiable data points.

To render results in an easily comparable format and reduce the burden of data handling, each master protocol consists of detailed analysis plans and uniform data tables, ensuring results are synthesized and presented uniformly. A key undertaking common to all three workstreams is to identify appropriate common data elements upfront, which contributes to the acceleration capability.

Outside of COVID-19, this approach was originally piloted to explore population characteristics, outcomes, and novel endpoints



**FIGURE 5** COVID-19 Evidence Accelerator Principles

SOURCE: Reagan-Udall Foundation and Friends of Cancer Research. 2021. COVID-19 Evidence Accelerator. Available at: <https://evidenceaccelerator.org> (accessed November 26, 2021).

for cancer treatments across multiple data sets. The EA has recently expanded to include post-acute sequelae of SARS-CoV-2 and an oncology working group, which are exploring the feasibility to evaluate potential differences in response to cancer treatments for patients who have previously had COVID-19. The EA is also being leveraged to explore possibilities for the development of novel treatments in substance use disorder.

An outcome of this two-fold work (convening experts and pursuing discrete regulatory questions) has been the development of principles (see *Figure 5*) to guide the EA's work, which underscore the importance of producing results that are reliable, credible, and usable by regulators and the health care community.

EA's ethos can be characterized as a “coalition of the willing.” There is no minimum time commitment or fee to participate in the collaborative, and organizations can enter and exit at will. While, on the periphery, health systems may have pre-existing DUAs with third-party analytics companies, collaborations within the EA are not governed by legal documents or a codified decision-making

process. Projects presented for IRB review under the EA have been determined to be exempt. In the absence of an enforcement mechanism, EA abides by the aforementioned set of principles that recognize the sense of urgency “to act fast” without sacrificing data privacy and scientific integrity. Another key operational tenet is embracing convergence and discordance to facilitate understanding about the nuances of the underlying datasets, such as how are the data gathered, the context in which they are gathered, and methods and perspectives of interpreting the data to the extent the data is harmonized.

While Friends and the Foundation guide and provide overall support for the project’s activities, leadership for identifying research questions and resolving issues come about organically. The EA invites collaborators’ input into decisions about which questions should be of priority and how and where resources should be allocated. Final decisions about priority scientific questions rest with the FDA.

The EA is unique in part because of the mutually reinforcing relationship between the value it presents to collaborators and the factors contributing to its success. Born out of the critical need to understand the rapidly evolving natural history of COVID-19, the EA offers a frictionless knowledge sharing environment in which collaborators can share results quickly without being encumbered by bureaucracy. The benefit is derived from leveraging the tools and cachet of the EA’s host organizations (Friends, Foundation, and the FDA). Given the urgency to rapidly construct a data sharing apparatus to keep pace with the quickly evolving nature of the pandemic, the project builds upon the data-sharing and analysis efforts of the oncology-related pilot studies conducted by Friends and numerous data partners. For example, the experience of Friends offered a model in how to formulate common research protocols and apply a parallel learning model. The EA’s data-sharing infrastructure borrows from the groundwork laid by the Foundation’s post-market safety surveillance work as well. EA leaders attribute these elements as key to the project’s success.

The prefabricated infrastructure of the EA coupled with the urgency of COVID-19 endows the effort with momentum by attracting a multitude of research partners who bring to bear their ca-

pabilities and expertise. It is noteworthy to acknowledge that the credibility of the FDA, Friends, and the Foundation created an added stimulus for organizations to join. EA organizers stated how the project quickly became a favorable venue for COVID-19 researchers, especially early in the pandemic (April–May 2020), which spurred others to join for fear of regret of missing an advantageous opportunity.

In turn, the cumulative strengths of the research partners have helped the collaborative develop stable research practices and rigorously generated evidence at a time when the quality and credibility of notable scientific studies of COVID-19 were being challenged. These aspects enhance the attractiveness of the EA, setting it apart as a low-cost, low-risk information sharing solution for entities concerned with the steep financial investment to obtain and maintain high-quality data, which has been noted as a deterrent to data sharing (Whicher et al., 2020).

## FUTURE DIRECTIONS

In addition to capitalizing on the resources and efforts of others, EA leaders underscore the significance of a “just do it” mentality, which at times can serve as the best antidote to the inertia of the health care industry. While early attempts to launch the project included missteps, through an iterative process driven by a willingness to learn from others and a commitment to transparency, EA leaders were able to reach a workable steady state within five months. For example, project leaders learned that a step-wise approach to analysis allowed the EA to review data and push out preliminary results to a wider audience. One of the interviewees for this case study aptly observed, “There are decisions being made every day based on limited evidence, so if we can do something, anything, to help we will do it.” While the statement was made in reference to COVID-19 decision making, the kernel of this statement is at the heart of accelerating continuous learning in health and health care and guides EA's operations to this day as it catalogues lessons learned and tackles new research questions.

## 6

# CASE STUDY: THE NATIONAL COVID COHORT COLLABORATIVE (N3C)

*Interviewees: Melissa Haendel, PhD, Co-director; Chris Chute, DrPH, MD, MPH, Co-director; and Andrea Volz, Communications Manager*

### ABSTRACT

The National COVID Cohort Collaborative (N3C) was rapidly established in spring 2020 as an open science partnership between the 60 Clinical and Translational Science Awards (CTSA) Program hub sites, the National Center for Data to Health (CD2H), multiple distributed clinical data research networks, and other partner organizations (Haendel et al., 2021; NCATS, 2021; CNDH, 2021). Born from the urgency to understand COVID-19, the N3C endeavors to improve the accessibility and efficiency of a large COVID-19 clinical data set while demonstrating a novel approach to sharing patient-level data and enabling individual researchers to use the data for approved projects. When COVID-19 emerged, CD2H leaders had been working on the harmonization of different common data models (see Box 2) already in use by the research community (Weeks and Pardee, 2019). The pandemic galvanized the CTSA community, and the N3C leadership accelerated progress toward the launch of a cloud-based consortium for aggregating institution-level data in a research enclave. Leaders' deep familiarity with collaboration challenges—including expedient human subjects review, aligned incentives for sharing, governance, and data privacy/security requirements—enabled them to quickly gain cooperation from research collaborators, the cloud computing host, funding

### *Case Study at-a-Glance: National COVID Cohort Collaborative*

*An open science and team science partnership between researchers and the National Institutes of Health to aggregate data to support COVID-19 research*

#### **Key Barriers Addressed**

- Traditional competitor mindset that typifies the grant-funded research environment
- Data latency and interoperability

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Ensure that typical incentives in academia, i.e., funding and publications, are in place to support cooperation
- Use existing common data models to improve data quality and comparability, and unify processes by creating single data use/data transfer agreements

#### **Insights for the Field**

- Use urgency of the crisis (in this case, COVID-19) to dislodge entrenched barriers to collaboration and fuel adoption of unified solutions, which should persist regardless of the pandemic

agency, and journal editors. The result is an active partnership and dynamic data enclave containing data for more than 3.2 million COVID-positive cases, and approximately 9.3 million patients as of November 2021. The N3C publications and data insights also continue to grow rapidly.

## BACKGROUND

The N3C is anchored by its data enclave, a secure, cloud-based platform housing individual-level clinical data from its contributing partners (all based in the U.S.). Partner organizations, such as the National Patient-Centered Clinical Research Network (PCORnet®) and TriNetX, provide access to structured data from elec-

**BOX 2***Common Data Models: Definition and Applied Example*

Common Data Models support the extraction and transformation of local data (e.g., from an electronic health record or wearable device) to a common format. This facilitates creation of high-quality data that is checked for accuracy and logic, mapped to agreed-upon standards, and suited for rapid and rigorous research.

A simple example is mapping of the birthdate “March 27, 1969,” which could be recorded in different formats by System A (1969/03/27), System B (03/27/69) and System C (03271969). Each of these can be mapped to the common format, MM/DD/YYYY to ensure comparability across systems’ data. When data are mapped to a common data model, the aggregated data can then be queried rapidly to answer an array of questions, such as the age, gender, and racial distribution of COVID-19 hospitalizations.

SOURCE: Ross, T. R., D. Ng, J. S. Brown, R. Pardee, M. C. Hornbrook, G. Hart, and J. F. Steiner. 2014. The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. *eGEMs: The Journal for Electronic Health Data and Methods* 2(1):1049. <https://doi.org/10.13063/2327-9214.1049>.

tronic health records (EHRs) across the country that can be queried to answer questions related to COVID-19. The N3C developed a comprehensive list of demographic and clinical data elements to create a research registry of patients who have been tested for or diagnosed with COVID-19, augmented with data on treatment and outcomes. Data are mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (Haendel et al., 2021). Researchers identify questions of mutual interest via meetings and during weekly presentation forums and can form dedicated workstreams based on collective expertise. The N3C website provides detailed information about various domain teams that have organized to address particular topics, including how to participate via Slack channels and other forums (N3C, 2021a). Potential collabora-

tors have access to onboarding documents, descriptions of domain teams, and other resources. This team science approach leverages complementary capabilities and domain expertise in disciplines such as informatics, epidemiology, biostatistics, data science, and a range of clinical specialties (e.g., cardiology, pulmonology, nephrology, neurology), which is particularly important given the progress in understanding COVID-19 as an illness with varying short- and long-term effects on different organ systems. The enclave is notable for its dynamic nature; data partners contribute new patient records an average of twice per week. Updates on the refreshed data are made available on N3C's website and the National Institutes of Health (NIH) website. Nonetheless, the enclave is dependent on the availability of "local" data and documentation in EHRs and/or claims data, which may be incomplete or inconsistent from one health system to the next.

## DESCRIPTION

Access to the N3C data enclave leverages best practices in collaborative data stewardship, privacy, and security, including institutional- and user-level permissions, two-factor authentication, compliance with all federal provisions for protecting data (e.g., Federal Information Security Management Act, HIPAA), and review of the nature and appropriateness of a given data request (Haendel et al., 2021). This is balanced by measures that enhance efficiency, namely the creation of a single institutional review board (IRB) to review all requests to query the N3C research data. Details regarding the protocol approved by the Johns Hopkins University IRB are publicly available, as are the extensive data specifications and software tools for data visualization and constructing efficient queries. In addition to creating a limited dataset and de-identified dataset, a subgroup of N3C researchers have developed a unique synthetic dataset, comprised of data that are computationally derived from the limited dataset and resemble patient information statistically, but are not actual patient data.

The data use request process for the N3C data enclave (as of April 2021) is briefly outlined here (N3C, 2021b). The requirements sup-



port data security, consistency, and continual reinforcement of the trust fabric across N3C, and are consistent across the limited, de-identified, and synthetic datasets, except as noted below.

- An institution-level data use agreement (DUA) must be executed as a prerequisite.
- Users must have completed required training in NIH information security and protection of human participants.
- Users register for an account to access the enclave, verifying that they have a tool in place to complete two-factor authentication in order to access data.
- Users then complete a data use request form, specifying the nature and scope of the research question and justifying the level of detail needed in the data.
- Users supply documentation of IRB approval as part of the data use request process for limited and de-identified data (not applicable for the synthetic dataset).
- A Data Access Committee reviews and approves or declines each request. Approved requests are valid for one year, and training/support is provided as needed.

Tools to support cohort exploration, including data views and analyses of de-identified data, are offered to CTSA researchers as well as citizen scientists, upholding the N3C's intent to maximize transparency and inclusivity, while preserving privacy and security.

Overall stewardship is provided by NIH's National Center for Advancing Translational Sciences (NCATS), which funds the N3C (Bennett et al., 2021). A Steering Committee approves activities and assures alignment of working groups, committees, and cores with the overall N3C goals. Much of the governance is focused on the data enclave, as described above. Primary governance documents include the Attribution and Publication Guidelines, the Community Guiding Principles, and the User Code of Conduct. This approach balances scientific autonomy with the creation of an open, respectful environment that encourages collaboration on this extraordinary health challenge. For instance, the Community Guiding Principles are partnership, inclusivity, transparency, reciprocity, accountability, and security.

While the N3C was an offshoot of the CTSA, a longstanding initiative that emphasizes collaboration, building the will for the N3C's creation was not automatic. Recognizing the need to garner rapid cooperation in a team science environment, leaders of N3C sought to identify champions from the CTSA hubs, the funding agency, and other community affiliates who could serve as ambassadors for the importance of this work. In addition, the N3C leaders did not stipulate that data users had to also be data contributors. This was an important aspect of gaining buy-in, as was the early and explicit plan to recognize all contributors as manuscript authors. In many academic institutions, participation on a publication is a key type of “currency” that supports promotion and tenure and may be particularly significant for early career investigators. Thus, N3C balances the goals of a very large team science consortium and the needs and values of individual investigators in academia. The leadership recognized the recurring tension in academia regarding research productivity and aims to produce high impact papers that recognizes all of the contributors. To this end, a consortium authorship model was developed to address the objective of recognizing the vast number of contributors to any given N3C paper.

The CTSA program is a prominent and prestigious feature for many academic institutions. The ability to foster both intra- and inter-institutional collaboration has been a hallmark of the CSAs for more than two decades and has helped launch careers in clinical and translational science for hundreds of scientists. Nonetheless, in the interviews that informed this case study, N3C leaders describe this as a “social engineering experiment,” in that it engenders a new level of openness and data sharing. The complexity of COVID-19 has helped collaborators recognize the importance of a diverse team with specialized expertise that could range from acute kidney injury to the Python programming language to pharmacokinetics. This team science approach is also intended to foster higher caliber research outputs, in that strong multidisciplinary teams and high-quality data can yield higher impact papers in leading journals. Participants in the N3C are encouraged to get involved via a prominent link on the N3C home page, either by joining existing collaborative groups listed on the N3C website, or self-organizing around topics of interest to create new “domain

teams,” which could range from a specific clinical topic (COVID-19 outcomes among people with diabetes) to increasingly broad and cross-cutting issues (impact of the pandemic in rural communities, or genomics and COVID-19). This rapid growth presented early challenges, especially in resource management and communication.

A unique challenge early in the N3C’s formation was developing the DUA between the NIH and the contributing institutions. The research data reside on a platform funded by the NIH, and the N3C itself is not a formal legal entity—simply a funded project. As such, NCATS is the fiduciary agent, holding the data and operating in accord with pertinent federal rules. Consequently, the Data Access Committee is composed exclusively of federal officials; N3C community members cannot participate. Progress and successful execution of the DUA was facilitated by the urgent need for this research platform, and a strong partnership between the NIH and NCATS leadership and the principal investigators of the CD2H initiative (which incubated the N3C). A related challenge was establishing a single IRB for the N3C. Johns Hopkins assumed that responsibility, and the logistics of linking other IRBs and applications was greatly eased by the SmartIRB infrastructure ([smartirb.org](http://smartirb.org)). This obviated the need for each data-contributing organization to write, submit, and review its own IRB application, instead ceding this regulatory requirement to a central IRB.

In the progenitor publication, *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust*, barriers cited by researchers and research oversight leaders centered on pace, process, and price of accessing data; data latency; and variability of IRB requirements. The urgency of the pandemic spurred partners to organize quickly and address issues related to rapid availability of high-quality, curated data as well as research oversight needs. In the time since the N3C leaders were interviewed, N3C has grown to 31 domain teams and contains data from more than 9.3 million patients, including more than 3.2 million COVID-19 cases. Nearly 200 institutions have signed a Data Transfer Agreement, signaling their willingness to contribute data to the enclave once it is harmonized to the Common Data Model. While the data enclave is the centerpiece, the N3C architects also describe

it as “a collaborative research community committed to the rapid generation and dissemination of knowledge for the public good, and to the advancement of COVID-19 science.”

## FUTURE DIRECTIONS

Team science can offer an uneven value proposition, insofar as large, complex consortia can become unwieldy or bureaucratic and can present political or communication-related challenges. The N3C leaders noted that one of their goals was to show that building something of this magnitude can be done rapidly and without significant friction. Though it took a pandemic to attenuate many of the traditional “pain points” in research (variable interpretations of the same protocol by multiple IRBs, lag time to attain research-ready data), it has also shown that science can move much faster and have a more immediate impact on health care. It will be critical to hold the gains in this regard, preserving both efficiency and data quality in collaborative research without reverting to pre-pandemic “business as usual” practices that could slow overall progress. Results from N3C studies will inform how COVID-19 is treated in both the short- and long-term (Bennett et al., 2021). The progress of the N3C to date demonstrates that both the philosophical and technical milestones of this initiative can serve as a blueprint for accelerating research, as well as implementation of findings in clinical practice.

# 7

## CASE STUDY: THE YALE OPEN DATA ACCESS (YODA) PROJECT

*Interviewees: Joseph Ross, MD, MHS, Co-director*

### ABSTRACT

The Yale University Open Data Access (YODA) Project acts as a data intermediary to facilitate the sharing of clinical research data between members of academia, government, and private industry for the purposes of conducting meta-analyses, replicating trial results, building upon prior findings, and conducting secondary analyses. Notable data partners include Johnson & Johnson, Medtronic, Queen Mary University of London, and SI-Bone, Inc (YODA, 2014). The YODA Project emerged to address issues in the clinical trial research community with regard to data transparency and integrity. A linchpin of its work is clearly defined, stakeholder-driven policies and procedures for data access and use, which have helped inform other data-sharing efforts and propelled the field toward embracing open access principles. Guided by a commitment to transparency, YODA has demonstrated the benefits of open access, promoted the responsible conduct of research, and eased a key barrier to sharing data between researchers and industry collaborators.

### BACKGROUND

The YODA Project was conceived in 2011 as an effort to address issues within the clinical trial research enterprise. At the time, the

### *Case Study at-a-Glance: Yale Open Data Access (YODA) Project*

*Facilitating data sharing between academic, government, and industry partners by accelerating use and dissemination of clinical trial data*

#### **Key Barriers Addressed**

- Costs associated with data procurement
- Selective sharing of clinical trial data by researchers
- Varied incentives to share data, augmented by heterogeneous organizational and individual viewpoints on data ownership

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Raw, de-identified clinical trial datasets and protocols are standardized and made available at no cost for approved requests
- To encourage transparency, all approved data use proposals are published on the YODA website, along with abstracts once results are available

#### **Insights for the Field**

- Engage a broad array of stakeholders in governance and creation of initial policies and procedures to garner widespread buy-in
- Espouse the essential importance of reporting all trial results and maximizing data use as foundational elements of the responsible conduct of research
- Provide widely accessible forums for meaningful engagement and discourse about data stewardship and sharing

program's co-directors had identified, through the course of their research, a number of challenges with regard to research integrity, transparency, and dissemination. For instance, their research demonstrated widespread problems with selective publication, as it took up to five years for two-thirds of completed clinical studies to be disseminated, and the remaining one-third were not disseminated at all. Simultaneously, Medtronic approached the co-directors to solicit an independent study replicating the results of the company's proprietary recombinant human bone morphogenetic protein-2 marketed for back pain (Ross et al., 2018). The oppor-

tunity helped position the YODA Project as a coordinating body for independent study reviews, offering private entities a mechanism to validate their research externally.

An important emerging dimension was a focus on data sharing. Following the conclusion of the Medtronic studies, the YODA Project convened a multi-stakeholder advisory group, consisting of regulators, industry leaders, clinical trial researchers, and patients, to discuss the merits of sharing data and what the infrastructure needs would be. The group contemplated a range of considerations, including the concern that the published data could be misused for commercial or litigious purposes or data dredging. However, the consensus among patients was that as long as their identities were protected, they supported the practice. In this spirit, the group decided to adopt an open data policy, and a significant outgrowth of the advisory group's engagement is the development of the YODA Project's Policies and Procedures.

The YODA Project requires data partners to make available raw, de-identified datasets, as they would with the FDA. In addition, the data are not limited to Phase 3 randomized control trials, but instead include all clinical trial data, regardless of whether they have been published or disseminated (YODA, 2021). The YODA Project's data-sharing pathways emulate National Heart, Lung, and Blood Institute's Biologic Specimen and Data Repository, or BioLINCC (NHLBI, 2021). In the same vein, requesters are granted access to the data files along with the study protocols and reporting analysis plans through a secure enclave (YODA, 2019). Early on, the YODA Project addressed the burden of data preparation by deciding against accommodating custom data requests, such as imaging data, which are difficult to transform into a tabular format. When asked if the decision foreclosed any opportunities, Joseph Ross, Co-director of the YODA Project, who participated in this interview, responded by saying that it strives to be a one-stop, nimble shop for data use. Entertaining specialized data requests would be too cumbersome for the initiative and detract from its appeal. He pointed out, though, that, if there is a need for specialized datatypes, the YODA Project could facilitate contact between the requester and data provider.

## DESCRIPTION

Requesters initiate access by submitting an online application. Along with submitting a research proposal, they are required to review the YODA Project's policies and procedures, complete a conflict of interest disclosure, and view a training video on the data use agreement (DUA). Data inquiries are thoroughly reviewed by a team of clinical investigators associated with the YODA Project. A key criterion of the evaluation process is whether the requested datasets match the intent of the project. Proposals also undergo blinded reviews by the data partners to ensure that the provided data are rendered appropriately for the request.

- Receipt of the data is contingent upon signing a DUA. DUAs are notoriously complex, nuanced, and can undergo multiple iterations; the more bespoke these agreements are, the more time they take for approval. To solve this issue, the YODA Project created three templates: one designed for U.S.-based researchers, another for researchers affiliated with the U.S. government, and the last one for foreign research entities. YODA's DUA circumscribes the following elements (YODA, 2019).
- The data must be used in service of the project goals as described in the research proposal.
- Data users agree to allow YODA to publicly post project proposals on its website in recognition of open science.
- Requesters are forbidden from redistributing or publicly posting the data online on their own channels.
- Beyond the completion of a project, investigators can access the data for up to five years.
- The YODA Project permits investigators to disseminate their findings in a peer-reviewed journal, as a preprint, or at scientific conferences.
- In pursuit of transparency, the YODA Project calls for data users to share a copy of their abstract with the project's data partners.
- Data requesters also are required to report their results if they choose not to publish their findings. Access to the data is pro bono. The YODA Project is supported by grants provided by



participating pharmaceutical and biomedical device companies.

The YODA Project leaders noted that the data-sharing mandates issued by the European Medicine Agency and the 2007 FDA Amendments Act, collectively, have put “wind in their sails.” Additional impetus came from the Pharmaceuticals Research and Manufacturers of America and the International Council of Medical Journal Editors, which issued consensus statements in support of data sharing, signaling a shift in attitudes toward research integrity and transparency.

Despite initial hesitancy from the health care community to open access to the data, the YODA Project’s first successful demonstrations were vital to engendering stakeholder confidence. As YODA’s codirector noted, once it was shown that the “sky did not fall when others had access to the data,” the community gained a sense of appreciation that compelled other companies to join or feel more comfortable in their own data-sharing efforts. Notably, Johnson & Johnson’s involvement, given its standing in the industry, helped to elevate the organization’s profile.

The YODA Project’s transparent and public processes are paramount to upholding relationships with data partners. The intent is to sustain engagement with data partners and foster an ethos of open science collaboration in the health care ecosystem that extends beyond the interactions occurring within the YODA Project. The YODA Project has started to witness signs of further adoption of open access data policies through changes mandating the sharing of data from research funded by the National Institutes of Health and the Patient-Centered Outcomes Research Institute. These changes contribute to the sentiment that data sharing, including patient access to data and cooperation between researchers, are becoming more of a common expectation (Dey et al., 2017).

## FUTURE DIRECTIONS

The YODA project recognizes that the sharing of clinical research data is a cultural issue in that there are variable norms, incentives, and expectations across the research community with respect to

the inherent value of sharing, and that stakeholders hold differing views on its relative benefits. Data has served as an important scientific currency in academia; consequently, there may be persisting reluctance to share it without an explicit mandate to do so. Acknowledging the potency of policy levers, the YODA Project believes that a mandate and portal akin to ClinicalTrials.gov for the sharing of privately sponsored research could compel the field to move toward this goal. While current data-sharing conditions may not be ideal, the YODA Project's leadership offers optimism for the field, imparting the advice that having a vision and unwaveringly advocating for that vision is essential. Equally significant is the imperative to cultivate trust from the outset by responding to concerns from stakeholders and being transparent about not only the approach to a policy or process, but the rationale behind the approach. Understanding the "why" can accrue trust about the "what."

# 8

## CASE STUDY: VIVLI

*Interviewees: Rebecca Li, PhD, Executive Director*

### ABSTRACT

Researchers often work in silos based on their institutional affiliation, reducing the likelihood of advancing research on scientific discoveries and hampering collaboration between researchers. With significant backing from academic, foundation, and industry stakeholders, Vivli created an open, low cost, cloud-based platform to facilitate the increased scale and speed of clinical trial data sharing and realize its mission of advancing science and improving public health. Through Vivli's cloud-based platform, researchers can utilize clinical trial data for research requiring data aggregation, reuse, and novel analysis. Crediting its success to its partners, Vivli's flexibility, accessibility, and commitment to privacy have also enabled its growth. Despite these achievements, challenges remain for efforts to incentivize and encourage broader and deeper participation in the platform, standardize the data shared, and increase the pace of uploading new data into the platform.

### BACKGROUND

Vivli was founded in July 2018 with a mission to "promote, coordinate, and facilitate scientific sharing and reuse of clinical research data through the creation and implementation of a sustainable global data-sharing enterprise" (Vivli, 2020). Vivli's vision is to "advance human health through clinical research data sharing"

### *Case Study at-a-Glance: Vivli*

*A trusted intermediary that developed and hosts a cloud-based platform of data from clinical trials to facilitate aggregation, data sharing, and analysis on a broad scale*

#### **Key Barriers Addressed**

- Clinical trial data are not routinely used to their fullest potential after initial analyses are complete
- Operational, financial, and logistical challenges to reusing data, including institution's varied interpretations of data use policies

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Absorb costs of de-identifying/anonymizing data for external sharing
- Remove cost as a barrier to facilitate researchers' use of data from the Vivli platform
- Require use of a single Data Use Agreement as a condition of acquiring data from Vivli platform

#### **Insights for the Field**

- Increase the use of data standards to facilitate broader clinical trial data sharing
- Accelerate data sharing by broadening and harmonizing mandates and policies across federal and philanthropic funders

while honoring and respecting clinical trial research participants' contributions.

Vivli initially began in 2013 as a collaborative project from the Multi-Regional Clinical Trials (MRCT) Center of the Brigham and Women's Hospital and Harvard University. As the idea began to gain traction, Vivli entered a technology partnership with Microsoft and BlueMetal in 2017 to develop an Azure-based platform for cloud-based access to research data and refine its technological capabilities. Vivli's work addresses key barriers related to ease and cost of data access.

After developing its technological capabilities, Vivli sought to expand its network and clout in the data-sharing ecosystem. Vivli obtained strong initial partners that pledged their clinical trial data as institutional contributors. Initial backers included GlaxoSmithKline, Pfizer, and Takeda Pharmaceuticals; foundations, such as the Doris Duke Charitable Foundation and Helmsley Charitable Trust; and academic institutions, such as the University of California San Francisco, Duke University, and Harvard University.

Today, Vivli is an independent, nonprofit organization with 11 full-time employees. As of the beginning of 2021, Vivli has over 31 different institutional data contributors and data from close to 6,000 clinical trials listed for sharing. These clinical trial data represent over 3.6 million participants (Li et al., 2020). The members, also listed on Vivli's website, are included in *Table 2*, categorized by organization type (Vivli, 2021).

## DESCRIPTION

Vivli is led by professionals with knowledge of clinical trial data and technical operations and strategy and is governed by a board of directors and three committees. Internally, the Vivli steering committee comprises Vivli institutional representatives, who meet regularly to discuss operational, governance, and policy issues.

Externally, Vivli is guided by an external advisory committee, which provides strategic advice related to governance, implementation, and institutional growth. Vivli convenes a community roundtable group drawn from the broader data-sharing community to ensure community engagement and partnership. The roundtable includes academics, patient advocacy representatives, and other stakeholders in data-sharing issues.

Finally, Vivli partners with the Wellcome Trust, which provides a secretariat function in overseeing an independent review panel. The independent review panel allows data contributors to funnel data requests from researchers to the platform to be externally reviewed for scientific merit and approved before the data contributor shares the data. This independent review process is maintained at arm's length from the data contributor. It enables stakeholder

Private-Sector Corporations	Education and Research Institutions and Databases	Nonprofit Organizations
AbbVie AstraZeneca Biogen Inc. Boehringer Ingelheim Celgene Corporation CureDuchenne Daiichi Sankyo Company GlaxoSmithKline Johnson & Johnson Kyowa Kirin Group Eli Lilly and Company Lundbeck Mitsubishi Tanabe Pharma Corporation Pfizer Inc. Regeneron Pharmaceu- ticals Roche Mallinckrodt Pharma- ceuticals Taiho Pharmaceutical Takeda Pharmaceutical Company Tempus	BioLINCC Harvard University Duke University Imimport Johns Hopkins University Project Data Sphere University of California San Francisco	Critical Path Institute Doris Duke Charitable Foundation The Leona M. and Harry B. Helmsley Charitable Trust

**TABLE 2** List of Vivli's Data Contributions by Sector

SOURCE: Vivli. 2021. Our Members. Available at: <https://vivli.org/members/our-members/> (accessed May 28, 2021).

trust in the data-sharing process by allowing transparent data sharing to advance scientific research.

Vivli began as a project from the MRCT Institute based at Harvard University and credited its success to its sponsors and institutions' backing, all of whom are leaders in their respective fields. Funding and advice from Vivli's leadership and sponsors enabled the creation of the Azure-powered platform.

Vivli is disruptive to the data-sharing community because of its role as a trusted intermediary with augmented technological capabilities and unprecedented access to de-identified participant-level clinical data. Prior to Vivli's creation, researchers typically would

**Table 2** Vivli data sharing platform's role in the data sharing ecosystem

Stakeholder	Stakeholder role	Vivli's complementary role
Non-profit funder	Funds trials and supports grantees to make data available	Offers non-profit funders a mechanism for their grantees to share data beyond the grant period
Academic researcher	Sharing and accessing data	Provision of metrics and digital object identifiers to measure open science behaviours in future grant-making decisions
For-profit funder	Sponsor trials and makes data accessible to researchers	Offers for-profit sponsors a managed-access mechanism for sharing their trials
Research participant	Participates in trials	Safeguards participant privacy while maximising the benefit of their contribution
Clinical trialists and researchers	Conducts trials and/or conducts secondary research on clinical trial data	Provides a mechanism for data access, long-term archiving and data reuse
Journals and publishers	Publishes findings from clinical trials and secondary analyses of clinical trials	Provides a controlled-access mechanism for journals and publishers that require clinical trial data to be shared

**TABLE 3** Role of Vivli's Data-Sharing Platform in the Data-Sharing Ecosystem

SOURCE: Li, R., J. Wood, A. Baskaran, S. Neumann, E. Graham, M. LEvenstein, and I. Sim. 2020. Timely access to trial data in the context of a pandemic: the time is now. *BMJ Open* 10:e039326. <https://doi.org/10.1136/bmjopen-2020-039326>.

have had to cooperate with individual organizations, researchers, and teams to overcome regulatory, trust, and institutional barriers before obtaining granular, participant-level data. Vivli also reviews the study protocol and the clinical study report to ensure quality control.

Vivli facilitates effective data sharing, aggregation, reuse, and analysis on an application built within the Microsoft Azure cloud computing platform to achieve its mission and vision. Using this application, Vivli serves as an intermediary to obtain rich datasets from a search of listed studies from renowned companies and academic institutions. These datasets can then be analyzed via Vivli's tools, aggregated with data from other studies and institutions, and shared.

The platform outlines five steps in its data request and sharing process. First, a researcher searches for data within Vivli's database. Second, a researcher must complete a data request form, and then an approving entity will review it. Third, if approved, the data package consisting of the individual participant data and supporting materials is made available after the requester signs the Vivli data use agreement (DUA). The data is then released and available within a secure research environment or through the Vivli platform. Fourth, the researcher will be able to use analytical tools to combine and analyze multiple datasets. Finally, researchers can

disseminate their research results using the platform to meet publication requirements.

As shown in *Table 3*, Vivli has characterized their value to stakeholders across nonprofit and for-profit funders, academic and clinical trialist researchers, journals and publishers, and research participants, with attention to participant privacy, data access, storage, and reuse, and the use of their platform as a mechanism to disseminate data beyond traditional mediums and beyond particular grant periods.

A facilitator for researchers using Vivli is the ease with which its platform can gather and reuse participant-level data to obtain insights and answer scientific questions. However, broad participation in the platform is driven by its flexibility and willingness to collaborate with stakeholders. Vivli members choose the kind of data they want to share while maintaining transparency.

Datasets contributed to Vivli can be provided by participating institutions to different levels of breadth and depth. For example, a private company may share Phase I-IV data of approved and unapproved therapeutics while accepting requests for additional data from competitors. Other private companies or entities might choose to share Phase II-III data from approved drugs while rejecting requests for additional data from competitors. To lower participation barriers, Dr. Li mentioned that Vivli doesn't require data contributors follow any particular guidelines beyond the imperative to anonymize data. Vivli also suggests participants follow the guidelines and timings suggested in the 2015 Institute of Medicine report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Dr. Li also added that industry members of the Pharmaceuticals Research and Manufacturers of America (PhRMA) and the European Federation of Pharmaceutical Industries and Associations typically adhere to their own principles. However, all members must transparently outline their sharing criteria for trials, which data they are willing to share, and the process of sharing their data on their Vivli member page. Adding to Vivli's flexibility is members' ability to remove data once they've shared it on the platform.

In addition to the flexibility to share specific pieces of data with other entities, Vivli allows stakeholders to choose the governance required to deliver and transmit various data. Data contributors



may select which datasets they wish to share and the mechanism through which they will share. Data can be downloaded from the platform after signing a DUA; alternatively, data contributors may choose to make their data available via a secure access environment. They also decide whether their data requests will be reviewed by an independent review panel or reviewed by a panel they manage.

Within Vivli's platform, only anonymized data is provided. Vivli has partnerships with Privacy Analytics and D-Wise for those that require anonymization services (they have both agreed for COVID-19 data under an agreement to waive anonymization fees if those data are shared in Vivli). Despite these efforts, Vivli's enclave mostly includes clinical trial data and, to a far lesser degree, shares registry data. These data were collected before most researchers started using broad consent forms. Unless stated otherwise, data given to Vivli is used with the understanding that participants have consented to their data being used for research and development purposes.

Under its DUA, Vivli does not charge users for obtaining data or submitting a request for data. Instead, Vivli is funded by foundation awards from the Doris Duke Charitable Foundation, the Leona M. and Harry B. Helmsley Charitable Trust, the Laura and John Arnold Foundation, the Lydia Hill Foundation, and PhRMA. Additionally, they are funded to a lesser extent by membership dues from their data contributors. Other platform features, such as R, Stata and SAS, are provided as part of a remote desktop-based statistical analytics environment. Access to these analysis resources is free in the first year of enrollment, but incur a modest daily computing fee after the first year of data use.

Vivli is both a nonprofit organization and a platform, distinguishing it from data sharing repositories hosted by academic or government entities. In addition, the lack of fees distinguishes Vivli from other data-sharing platforms, such as the Centers for Medicare & Medicaid's Virtual Research Data Center, and datasets, such as the Framingham Heart Study, which is supported by the National Heart, Lung, and Blood Institute (NHLBI). Because the need for datasets and other services varies by user, Vivli advances research by removing access barriers.

Occasionally, potential members from all sectors have expressed reluctance to sign Vivli's DUA, which is standardized across institutions and mandatory for enrollment. They are also sometimes reluctant to use Vivli's harmonized request form to obtain data from other entities. Neither the DUA nor request forms can be customized for each member unless Vivli's steering committee recognizes a need for revisions to the form.

Gaining the trust and confidence of academics continues to be a major aspiration. Out of the data from more than 5,900 clinical trials on Vivli's platform, the number of trials contributed by single investigators from academic institutions is small and continues to decrease. Academic institutions, usually founded to achieve freedom of speech and thought, are often governed by shared governance policies between university administrators and academics. As a result, academic institutions struggle to sign the institution-wide DUA without broad consent from faculty-led governance entities.

According to Dr. Li, academics express concerns about adequately anonymizing data and the financial costs of sharing their data with membership dues. Because of these barriers, individual academics and researchers from these institutions must either volunteer to join Vivli's platform as individual academics or consent to their institution signing an institution-wide DUA with Vivli.

Despite these barriers, there are notable exceptions, with universities such as Harvard, Johns Hopkins, and Duke University School of Medicine, joining the platform. In line with other Vivli members, each university has its policy on data they are willing to share with other entities. In Duke University's case, it is important to note that, within universities, individual schools, such as the School of Medicine, may sign limited data-sharing agreements with Vivli that do not apply to the entire institution and its other schools, departments, or colleges.

## FUTURE DIRECTIONS

Dr. Li, who added that Vivli prefers the Clinical Data Interchange Standards Consortium's Study Data Tabulation Model format that is used for data transmission to the National Cancer Institute,

hopes that researchers can and will standardize data across institutions and research entities. Clinical trial data with similar or identical formats would facilitate data sharing without the need to verify and standardize clinical trial data before they can be used on Vivli's platform. Many valuable datasets were collected and created before standardized datasets were widely adopted.

Additionally, Dr. Li added that funders of research and development should incentivize or require clinical trial investigators to share their data. For example, the Wellcome Trust mandates making data, code, or materials underpinning published research findings accessible to other researchers at the time of publication or as soon as possible during public health emergencies. They also strongly encourage sharing data in community health repositories for broader dissemination of articles and their data. The Bill & Melinda Gates Foundation also commits to sharing data through open datasets, curated resource collections, online data analysis and visualization tools, survey participation, and external reporting to trusted third parties. Finally, the NHLBI also requires research funding applications requiring \$500,000 or more in direct costs to submit a plan to share their final research data, with additional policies for studies conducting genomic research, regardless of costs, per the National Institute of Health's Genomic Data-Sharing Policy (NHLBI, 2014).

Rapid data sharing, Dr. Li added, is both an imperative and reality in the age of the COVID-19 pandemic. Before the pandemic, Vivli's steering committee members were aiming for releasing data 12-18 months after concluding a clinical trial—actions to combat COVID-19 exemplify the future possibilities in sharing data more expediently.

## 9

# CASE STUDY: THE LOUISIANA PUBLIC HEALTH INSTITUTE (LPHI) AND THE GREATER NEW ORLEANS HEALTH INFORMATION EXCHANGE (GNOHIE)

*Interviewees: Tom Carton, PhD, MS, Executive Director; Salvatore J. Peraino II, MS, Director of Business and Controller; and Kyla Mor, MSPH, Director of Product Development*

### ABSTRACT

The Louisiana Public Health Institute (LPHI) serves a unique function for the State of Louisiana, partly due to Hurricane Katrina's devastating effects. The 2005 Hurricane had singularly catastrophic effects on the state's health care infrastructure and human health. It led to the founding of the Greater New Orleans Health Information Exchange (GNOHIE) in 2010 as a data infrastructure apparatus using data and technology to facilitate innovative health data sharing between private hospitals and community health clinics. These endeavors inform the GNOHIE's objectives of improving population health, reducing health costs, and enhancing the patient experience. Today, with several full-time employees and financial support from membership dues and state and federal funding, the GNOHIE provides critical data services and produces insights from patient-level data and health systems interactions to inform population health efforts and improve integrated care quality in Louisiana. It has addressed technical and operational barriers that can be helpful to other nascent and mature health information exchanges (HIEs).

*Case Study at-a-Glance: Louisiana Public Health Institute (LPHI) and Greater New Orleans Health Information Exchange (GNOHIE)*

*Health systems, state agencies, and community-based organizations organize and exchange data for more efficient, coordinated care*

**Key Barriers Addressed**

- Data availability and interoperability across health systems and providers
- Variable interpretation of regulatory policies that govern different types of data and spheres in which data originates

**Specific Solutions for Data Sharing between Patients and Clinicians**

- Create a technical infrastructure for clinical data exchange to support care coordination
- Develop dependable utilities that meet specific needs of local health systems

**Insights for the Field**

- Plan for sustainability at the outset, especially for initiatives that begin via grant funding, as data sharing entails appropriate, on-going resources
- Integrate data sharing requirements into policies for health care payment models such as value-based care models

## BACKGROUND

In 2010, the LPHI, a 501(c)(3) nonprofit organization, was awarded a \$13.5 million Beacon Site grant by the Office of the National Coordinator for Health Information Technology to establish the Crescent City Beacon Community (CCBC) to improve population health in Jefferson and Orleans parishes in Louisiana. From 2010–2013, the CCBC worked with more than 150,000 patients and 150 providers to enhance primary care coordination through data sharing between private and community providers and reduce chronic disease burden (Khurshid and Brown, 2014). Among CCBC’s accomplishments were:

- Creating a real-time automatic encounter notification system to facilitate care transitions from emergency care to primary care clinics;
- Implementing clinical transformation programs at 16 clinics with significant improvement in several outcome measures, including coordination of care and completion of follow-up visits;
- Establishing the GNOHIE, its clinical data exchange platform; and
- Storing more than 900,000 patient records from hospitals and outpatient clinics (LPHI, 2016).

In 2013, after completing the initiative, the CCBC's administrative committee voted to form a separate nonprofit organization, the Partnership for Achieving Total Health (PATH) program, to manage the GNOHIE's operations and strategy. PATH functions as the organizational home for the GNOHIE. Since the closeout of the CCBC grant program, the GNOHIE has been transitioning to a stable financing model via a combination of membership dues and state and federal grants.

The GNOHIE's mission is to advance the Triple Aim of improving population health, reducing health costs, and enhancing patient experience. The GNOHIE advances its mission by connecting providers, facilitating secure exchange of patient data, and delivering actionable insights from the information that is exchanged. This aim is underpinned by a focus on health equity, especially after the 2005 hurricane illuminated stark disparities in receipt of timely, coordinated primary care.

Participating organizations sign agreements that cover the contractual and data-sharing parameters of membership. Once an agreement is signed, the GNOHIE technical team works with partners to establish a data connection, which are implemented in varying ways depending on data assets and needs of participating organizations. Participants send data to GNOHIE via the Health-Level 7 (HL7) interface that works with EHR software to support data exchange, such as clinical data extracts supplied by specialized population health management software, or a patient demographic file. GNOHIE sends notifications back to participants through EHR

and software interfaces, and through secure direct mail accounts. Analytics and reports are sent to participants through encrypted email or made available through user-specific access to a web-based business intelligence application. Uses of these reports and notifications are detailed below.

## DESCRIPTION

The GNOHIE team includes a staff of 10, including positions in technical operations, product management, compliance, and business and finance. Providing oversight and strategic advice is a six-person board of directors, comprised of health care leaders in Louisiana largely focused on under-resourced patient populations. Collectively, board members provide perspectives on health promotion and prevention; primary care access; health equity and community health services; patient, family, and community advocacy; care management; research; and policy.

The GNOHIE receives Admit, Discharge, and Transfer (ADT) data and clinical data from participating organizations and shares it with other treating providers to support care coordination and clinical management across care settings. Structured data elements conform to existing health data interoperability standards, such as HL7. The GNOHIE, in collaboration with its partners, has developed an extensive set of policies and standard operating procedures to govern data sharing and access, which resulted in the GNOHIE's terms and conditions policy governing its data use and sharing (GNOHIE, 2018a). The GNOHIE employs an "opt out" patient consent policy, which permits data sharing for patients who receive care at participating organizations. Patient information is only shared with other treating providers and, if patients do not want their information shared, they can opt out at any time. Participating organizations must comply with the necessary laws and regulations, including the Health Insurance Portability and Accountability Act (HIPAA). On the other hand, data providers retain their right to own their data while granting PATH, as a custodian of the data, the ability to share it with the GNOHIE. Meanwhile, the GNOHIE's obligations include compliance with the terms and conditions and laws and regulations; maintenance of the GNOHIE and

its services; protection against malicious software, viruses, and other threats; support through the training of personnel, telephone and email support; and periodic audits, investigations, and reports. In short, the GNOHIE complies with or exceeds the standards listed under HIPAA through its data security and transparency policies.

Transparency is an essential facet of health information exchanges nationally and for the GNOHIE. The GNOHIE provides a public complaint form for individual patients, member organizations, or whistleblowers to communicate concerns regarding opting out of sharing data, data privacy, customer service, and other issues. The form allows for complaints to be made anonymously, thereby reducing the barriers to filing a complaint. Additionally, the GNOHIE provides a landing page with an embedded opt-out form, its phone number, and a link to frequently asked questions from patients and families (GNOHIE, 2018b).

The constellation of policies and processes described above illuminate the technical and sociological complexity of maintaining a health information exchange. These policies accurately reflect the interviewees' experiences in creating the technological, legal, and policy structure of the GNOHIE. Given that the policies illustrate the GNOHIE's thoughtful consideration of real and potential issues in sharing and exchanging data based on their direct experience, this can provide guidance for similar health information exchange efforts—both nascent and mature.

The GNOHIE acts as a conduit across the care continuum, providing and facilitating information exchange services for primary care and behavioral health, hospital and health systems, social services, health plans, accountable care organizations, and correctional health care providers. Importantly, this attenuates the need for each provider or system to set up its own distinct infrastructure. Many of the infrastructure features and functions are housed in the GNOHIE, as enumerated below. Key features that underpin the GNOHIE's activities and service offerings include the following:

- **Master Patient Index**, an interoperable platform that supports unique patient matching across records shared by participating organizations. This index enables the GNOHIE to send participating entities the most updated information on each patient when required and ensures health care data is ac-



curately attributed to the correct individual.

- **Data Integration** with various EHRs and population health management systems. These integrations enable bidirectional data exchange with participating organizations to supply patient health information to the GNOHIE and to deliver encounter notifications back to the provider. In turn, users can examine GNOHIE data within their existing systems and workflows or at the population health level, such as viewing rates of influenza-like illnesses.
- **Secure Direct Mail**, a secure, HIPAA-compliant web-based mail service that facilitates the exchange of protected health information and other data. This system can also include multiple direct mail accounts to ensure the flexible transmission of data. As part of the value proposition, this ability to send encrypted information enables an organization with less robust IT capabilities to log into the Exchange and initiate secure mail and secure data transfer.
- **Encounter Notifications**, powered by transmission of ADT data across care organizations and settings. Participating providers use the notifications to monitor hospital utilization; facilitate and track timely post-hospital follow-up appointments; and remain informed about care delivered at external facilities, including both hospital and ambulatory care settings.
- **Reporting and Analytics**, leveraging the same patient data delivered through encounter notifications and making it available to users in a longitudinal format for an entire population of patients. Reports include both aggregate and patient-level data to assist participating organizations to efficiently delegate tasks, such as scheduling of follow-up appointments, and to inform population health strategies and resource allocation.

Taken together, these utilities ensure accurate identification of patients as they receive care in various systems and sites, and support community providers as they monitor individuals' health needs. For example, if a patient is discharged from a hospital, a primary care provider utilizes GNOHIE-generated discharge notifica-

tions as a prompt to schedule follow-up appointments. The hospital is notified accordingly, via GNOHIE's ambulatory notifications, whether a follow-up visit occurred. Hence, care coordination information about the patient is transferred automatically without the need for duplicative data entry and validation processes. Consistent with the GNOHIE's stated mission, this supports both quality and safety, and can reduce administrative costs associated with provision of health care—an important aspect of the value they provide for their members and partners.

To support deployment and maintenance of the GNOHIE, the LPHI has benefitted from the cooperation between federal, state, and local governments. Their cooperative endeavor agreement was funded with a 90/10 matching fund from the state of Louisiana and the federal government, which emerged from the Health Information Technology for Economic and Clinical Health Act. Under this project, the Centers for Medicare & Medicaid Services paid 90% of funds related to the GNOHIE, while the state of Louisiana paid 10% of these funds. The GNOHIE also cooperates with the Louisiana state government on analyzing Medicaid data, which is shared for analysis then returned. These reports enable a statewide culture of learning and improvement, exemplified through insights from influenza surveillance analytics and an operational dashboard on Medicaid enrollees using HIE data, as two examples. Throughout the pandemic, rather than shifting health information exchange resources to COVID-specific needs, GNOHIE continued to focus on their core work to advance and expand care coordination capabilities, regarded by their stakeholders as a top priority.

Before the GNOHIE was founded, the team needed to create the requisite legal and regulatory scaffolding to ensure compliance in patient privacy data and sensitive data controls. Other efforts included creating rules governing information exchange, patient attribution, and navigating myriad technical and business implications of data sharing. These efforts required the GNOHIE team to navigate a complicated regulatory ecosystem, especially in the absence of accessible international data-sharing frameworks, such as HL7. Additionally, the incentive structure does not cover the real costs of data sharing, given the evolving and complex regulatory environment.

Even after the GNOHIE's founding, variable interpretations of federal laws and regulations have presented considerable barriers to data exchange activities. HIPAA has been a barrier for the GNOHIE due to its multiple grey areas, such as rules around sensitive data and various interpretations by entities' legal counsels. In aggregate, the law is often invoked as a reason not to share data, and there is no mechanism to provide authoritative clarification, unification, and alignment on these debates. Other primary legislative and regulatory texts, such as Title 42 of the Code of Federal Regulations Part 2, which regulates confidentiality of information related to substance use, continue to provide vague direction despite an update featured in the 2020 Coronavirus Aid, Relief, and Economic Security Act.

Government guidance, such as the Trusted Exchange Framework and Common Agreement (TEFCA), is, according to the GNOHIE team, considered an attempt to align health data sharing, but may lack the legal and regulatory power to mandate cohesive collaboration on data sharing (HealthIT.gov, 2021). Frameworks such as TEFCA need to be less broad and include strong regulatory mandates to ensure success. If not, the GNOHIE team observed, disparate interpretations of regulations and inconsistent efforts to expand the siloed health data-sharing ecosystem will persist.

## FUTURE DIRECTIONS

Despite HITECH's financial opportunities and its attempt to promote EHR integration, current financial and regulatory incentives do not enable and facilitate data sharing. According to the GNOHIE team, policies need to help align incentives to offset real and perceived costs of data sharing. The GNOHIE team observed that despite conversations at the federal level about prioritizing data sharing and interoperability, this priority was not attached to policies and actions to clarify and reform the regulatory environment. While they cited important benefits of HITECH and interoperability legislation, the accompanying incentive structure and operational aspects have received less attention. In particular, the GNOHIE team indicated that there are real costs associated with data sharing, and no penalties for not exchanging data, so these as-

pects create friction. Without this regulatory clarity, the GNOHIE team added, data-sharing vendors can still charge exorbitant fees to health systems to encode large amounts of data for exchange purposes. Because of these costs, changing vendors and incentivizing data-sharing efforts is tremendously difficult for exchanges like the GNOHIE. More recently, the requirements imposed by the information blocking rule may offer a stronger push to move the needle on interoperability (Federal Register, 2020).

Before embarking on data-sharing projects, interviewees stressed that organizations should understand that exchanging data is not solely a technology-based endeavor. Due to various societal, legal, and governance-related implications of data sharing, organizations should prepare resources and anticipate the consequences of participating in data sharing, and gain buy-in for the vision and value proposition for data sharing, even as the technical details are worked out. Furthermore, to realize its full potential, data sharing and exchange should not be relegated to a minor project or priority within a health system or organization. Integration with other major priorities such as payment for performance and value-based care initiatives could embed data sharing as part of an organization's priorities.

Development of an HIE—or any similar shared-data resource such as an All-Payer Claims Database or research data enclave—should offer something of real value to each data contributor. The upfront work of understanding and co-creating the benefits and anticipating the range of technical, procedural, financial, and sociocultural barriers has been demonstrated by the GNOHIE as a formula for success.

## 10

# CASE STUDY: THE SANFORD HEALTH SYSTEM AND SANFORD DATA COLLABORATIVE

*Interviewees: Benson Hsu, MD, MBA, VP, Enterprise Data and Analytics (former); Emily Griese, PhD, Vice President of Population Health and Clinical Operations; and Arielle Selya, PhD, Director of Data Exchange Core*

## ABSTRACT

The Sanford Data Collaborative (SDC) is an enterprise-wide initiative of Sanford Health System, created in 2015, and designed to unify the data sources and data analytics capabilities of the organization in support of improving population health outcomes. The SDC responded to unmet needs of the health system and now provides an important resource for the research community at the regional and national levels. SDC developers recognized that each data element in its electronic health record (EHR) was important, singularly and in combination, for helping Sanford Health System provide higher quality health care. Building this complex resource from the ground up entailed garnering collaboration at the executive level, engaging compliance and legal and personnel at the outset, and aligning it with health system priorities—including the movement to value-based care. The architects of the SDC recognized that providing access to both internal and external researchers would support their ability to maximize the utility of the data for the good of patients and would also meet demand in the research community for high-quality health data (Griese et al., 2017).

### *Case Study at-a-Glance: Sanford Data Collaborative*

*Health system fused its organization-wide data sources and data analytics capabilities to support collaboration with researchers and improve population health*

#### **Key Barriers Addressed**

- Wider use of health data by the system and collaborating researchers
- Broader access to data and insights to help unlock their fullest potential

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Help patients and systems recognize the value proposition in optimizing the use of health data
- Motivate system to improve data quality and support the health system's internal capability to feed insights from research back into care delivery

#### **Insights for the Field**

- “Data get better with use” axiom can guide efforts to unify disparate data sources to improve quality and outcomes
- Creating opportunities for researchers to use real-world data creates value for the health system

## BACKGROUND

The SDC is a dynamic data resource for researchers at the Sanford Health System as well as the academic research community outside of Sanford. The SDC integrates data from the EHR with other health system administrative data to create a secure enterprise-level data warehouse, consistent data dictionary, and data governance processes. It was established as Sanford Health consolidated its central functions after a series of mergers over multiple years, resulting in numerous data sources and silos. Creating a unified and coherent data infrastructure, including a consistent data dic-

tionary and data governance processes yielded a “single source of truth,” especially important for a health system that is geographically spread throughout a large rural region of the U.S. Consolidating the data also enabled the system to bolster and create coherence in its analytic capabilities, evolving in sophistication from descriptive analytics to predictive analytics (Hsu et al., 2017). Alongside the data resource itself, the team built the necessary processes and resources that would support its functions, including a privacy board, outreach function to promote use, and on-demand analytics tools (Griese et al., 2017).

The on-demand tools were designed to support both internal and external research. The interviewees observed that, prior to the creation of the SDC, data requests and queries submitted by a research analytic team were often lower priority, with operational analytic requests taking precedence. These query requests also could be incorrectly translated by the business analyst running a research query, or incompletely formed based on the data available in the EHR. To this end, by leveraging and augmenting the built-in analytic capabilities of the EHR, anyone in Sanford Health can now run queries on the aggregated data. External researchers submit queries to a project manager. For approved requests, data are shared via a secure file transfer protocol. Such queries can help determine the feasibility of a potential research study and answer straightforward questions about such things as prescribing trends or disease incidence, for example.

The first generation of the SDC was developed for use with aggregated data. Using it for studies that need person-level data is comparatively more complex, and the SDC team has turned its attention to developing appropriate processes and audit trails, and working through attendant ethical, regulatory, and privacy-related aspects. To support this work and sustain leadership buy-in, the group seeks to address research questions that are also of interest to the health system. This alignment with the needs of the organization is integral.

## DESCRIPTION

Early work with the organization's legal department smoothed the way for the eventual success of the SDC. With the legal team, SDC leaders were able to work through challenging data use agreement (DUA) questions and develop policies about aggregation and de-identification. An insight gleaned from devising legal and governance processes that can be useful to others is the consideration of *who owns the knowledge* resulting from a given project or analysis? This is often framed through the lens of data ownership, but knowledge ownership is equally important to consider for any health system aiming to create a similar data utility. The SDC director reviews data requests for alignment and appropriateness now that an established DUA template is in place.

A critical early challenge was to garner support at the highest level of health system leadership. This was achieved by helping leaders see the possibilities in sharing the data and urging that it be leveraged to its fullest potential. The axiom in informatics that “data gets better with use” was on the minds of the creators (who were research and data analytics leaders in the organization). One of the interviewees for this case study observed, “every single touchpoint provides information about our patients and populations. If we just let it sit there, we are not fulfilling part of our responsibility as a health system.” The team actively cultivated engagement from across the enterprise, ensuring that the SDC could be analyzed by business personnel, as well as research and operations teams. The SDC has been used by engineers, public health researchers, and pharmacy personnel, each of which brings a different lens and questions of interest.

An important lever used to gain organizational buy-in was related to priming the organization for future health care transformation efforts, particularly value-based care. Preparing the data infrastructure to support system-wide priorities and bolstering research capabilities were two of many benefits, and deeper understanding of the data signaled to their community that they are careful stewards of important health information. A corollary benefit has manifested from this concerted work, in that researchers



regard this resource as a signal of Sanford Health's commitment to the community.

Significant attention was devoted to engaging and educating researchers and leaders of academic institutions in the region. Helping researchers understand the complexity of a seemingly straightforward query has helped garner bidirectional trust and has helped the SDC leaders forge meaningful connections between area researchers and service line leaders in the health system. This cycle helps improve the quality and relevance of the research and amplify the visibility of the SDC as a data resource. An important facilitator, then, is the role of a "boundary spanner," that is, someone who understands both research and care delivery, and can serve as a translator as well as an ambassador in the organization and externally.

As described above, the SDC offers unique value to Sanford Health System, its internal researchers, the broader research community, and the patients receiving care there. Each query offers the implicit opportunity to improve on the data resource and assure that processes are meeting the needs of all stakeholders—from the researchers to the leaders to the institutional review board and compliance personnel. A consolidated and curated data warehouse can accelerate research along the continuum from idea to intervention to eventual implementation of insights.

## FUTURE DIRECTIONS

The SDC team advised that others initiating a comprehensive data resource include their legal, compliance, and quality improvement leadership at the outset, along with research and clinical leaders. Despite early skepticism, the team urged a "just do it" mindset, addressing barriers as they arose, as the end result is worth the effort. One case study interviewee rebutted the notion that sharing or releasing data could diminish a system's competitive advantage and averred that it actually enhances it.

A second dimension of the advice for others pertained to technical facets of building the data collaborative. Interviewees observed that the inconsistency in health care data and analogous inconsistency in outcome measures creates complexity and frustration.

A unifying definition across payers and providers would facilitate measurement, whereas the current idiosyncrasies from one health system to the other creates friction in measurement, data sharing, and population-level care improvement. The advice is to go into this endeavor clear-eyed about these inconsistencies and from a data governance standpoint, aiming for steps that enable data integration, whether a common data model, use of standards, or a data governance apparatus to reconcile differences across data sources.

In short, the experience of the Sanford Health System and SDC illuminate a path so that other health systems undertaking similar endeavors do not have to “go it alone.”

## 11

# CASE STUDY: BLUE CROSS BLUE SHIELD OF NORTH CAROLINA

*Interviewees: Robert Emerson, PhD, VP and Head of Strategic Data Management; and Bradley Donovan, Director of Divisional Strategy*

## ABSTRACT

Blue Cross Blue Shield of North Carolina (Blue Cross NC) is the largest health insurer in the state of North Carolina, with more than four million covered lives as of 2020. With sizable market share, Blue Cross NC has led a statewide shift toward value-based care (VBC) in an innovative partnership with five health systems. This initiative, called Blue Premier, aims to align incentives for improving quality of care, and create a data infrastructure for standardized reporting of data to support clinical improvements and manage total cost of care. Initiated in 2019, the program has grown from five initial health system partners to eight, and has reported both quality improvements and cost savings. In developing Blue Premier, program leaders began with a single-site pilot to navigate the operational, legal, and technical barriers of data sharing, and used insights from this pilot to create a broader playbook to support data exchange and anticipate challenges as new systems were brought into the program.

### *Case Study at-a-Glance: Blue Cross Blue Shield of North Carolina*

*Shared clinical data between payer and eight health systems in the state to facilitate a large-scale shift toward value-based care*

#### **Key Barriers Addressed**

- Financial concerns, i.e., loss of competitive advantage if data are shared within the same market
- Concerns about misuse of data

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Ensure a compelling value proposition, such that goals for sharing data align with the strategic business goals of the organizations contributing data
- Start with data that all contributors have in common
- Engage the highest levels of leadership (CMO/CMIO) throughout the process

#### **Insights for the Field**

- Pilot the technical and operational aspects of data sharing with a contributing partner that is skilled, motivated, and adequately resourced to problem-solve when challenges inevitably arise

## BACKGROUND

Value-based care has been a cornerstone of the shift in health care payment reform over the last decade. Under this arrangement a physician, clinic, or health system contracts with an insurer (payer) to be accountable for a defined population of patients. Pre-defined benchmarks for cost and quality are often established as part of the contract, and those who deliver health care are thus incentivized to provide the highest quality care at a lower cost. Many VBC arrangements are “shared risk” in that doctors and hospitals share in the cost savings—receiving performance payments if targets are met, and share the losses if targets are missed.

In North Carolina, a collaboration called Blue Premier was initiated in 2019 between five health systems and a single payer, Blue Cross Blue Shield of North Carolina (Blue Cross NC). Blue Premier is a significant effort to apply VBC on a very large scale. For it to succeed, the health systems also needed to agree to share health data to support population health management. For independent physicians, this data-sharing and analytics capability is managed by a related partnership with Aledade, a private company that supports the creation and maintenance of accountable care organizations (McClellan et al., 2019; Reese, 2020).

Data shared between systems in the Blue Premier program and Blue Cross NC (as the payer) include encounter-level structured and unstructured clinical notes, and the structured data are typically formatted to Health Level 7 (HL7) standards, which is an international set of standards that guides the transfer of clinical and administrative data between software applications. Early efforts to pilot test the health data sharing focused on the use case of Admit, Discharge, and Transfer (ADT) data, as this is a core piece of administration and patient management, and the Continuity of Care Document, which helps systems exchange essential patient information (e.g., demographics, allergies, medications, lab results) in a seamless and standardized way. In the pilot, Blue Premier leaders identified important considerations when moving data from clinicians to payers, including the extent to which certain data may be filtered out before it is shared with the payer, such as any private-pay arrangements, or certain sensitive clinical information. A key lesson from the pilot was to identify a use case that had value for both the system and Blue Cross NC. The workflow for exchanging ADT data was comparatively straightforward, and the value proposition of reducing redundancies and administrative burden was compelling as a pilot.

## DESCRIPTION

With respect to governance and leadership, again, the VBC arrangement drove the parameters for sharing data. Negotiations and agreements were developed at the level of the chief medical

officers and other C-suite leaders, such as chief data officers, chief data analytics officers, and chief financial officers. The leadership of the participating organizations understood that there was a positive return on investment netted by creating a more efficient infrastructure for sharing data, particularly as federal regulations from the Centers for Medicare & Medicaid Services and the Office of the National Coordinator for Health Information Technology move in the direction of mandatory health data sharing.

As Blue Premier moved from a pilot to a larger initiative, new and different challenges were encountered. Addressing concerns from legal and privacy departments was cited as an early learning, in that the legal contracts needed to include mutually agreeable language related to data use and business associate agreements, as well as *a priori* agreements about how to handle out of network claims data, which requires upfront patient consent in order to be shared. An additional challenge was that the philosophies and processes at the health system level were highly variable. Some systems were more comfortable than others with the data release procedures, including the cadence of data release, and types of data to be shared. It was also the case that data were usually, but not always, harmonized to the established HL7 standards, which necessitated additional data extraction and standardization.

The level of comfort with data sharing also varied by medical subspecialty within a system, an important insight for other health care delivery organizations, since improving quality is a linchpin of this initiative as it is for other accountable care organizations structured around VBC agreements. Many medical specialties have specific performance measures designed to measure quality of care. When system partners come together to create accountable care organizations and craft attendant data-sharing agreements, the Blue Cross NC interviewee encouraged that the partners take these specialty-specific performance measures into account rather than developing new or separate quality measures. This will facilitate data exchange and reporting and will also yield administrative efficiencies and accelerate support from frontline clinicians. Moreover, many specialties, such as oncology, have many interdependencies with other clinical disciplines (e.g., radiology, surgery, primary care), and measuring quality must take these relationships into account.

Overall, Blue Premier's ability to overcome these challenges was facilitated by the VBC arrangement itself, and the fact that VBC conferred benefits to participating systems. In short, meaningfully improving quality and reducing costs was predicated on data sharing. An unanticipated but beneficial opportunity to leverage their early work on standardizing data for sharing came about when the COVID-19 pandemic hit, as Blue Premier collaborators were able to use their standardized lab data to guide plans for virtual care and telehealth, including finding capacity in their systems for different patient care modalities.

## FUTURE DIRECTIONS

Reflecting on the early success of Blue Premier, the case study interviewees recognized two important contributing factors: Blue Cross NC's unique position in the market and the fact that it insures half the population of North Carolina. The state also has a significant proportion of health systems that are academic/research institutions, creating an environment that is open to innovation and offers technological acumen. Some of these factors are configurable in other states or regions, even as other factors are unique. The interviewees further emphasized the fundamental importance of understanding the incentives driving each partner and then aligning the incentives across all collaborators. In their case, the desire for more fluid data that could be applied for multiple synergistic needs, coupled with the shared imperative to take on VBC contracts helped to drive collaboration and a desire to build "the infrastructure of the future."

Notwithstanding the success of Blue Premier, the case study interviewees believed that, in spite of the important work the partners did to standardize data across their systems, the better way forward would be for all health data to be in a pre-standardized to a pre-specified format, to avert the need for back-end harmonization by each individual health system or data compiler (such as a health information exchange). A second recommendation they encouraged was the development of a national governance structure to oversee the (re)use of data by third parties that are not covered by the Health Insurance Portability and Accountability Act. Ulti-

mately, the initiation of a comprehensive shift to VBC offered an opportunity to improve the care while improving the data infrastructure.



## 12

# CASE STUDY: MAYO-GOOGLE PARTNERSHIP

*Interviewees: John Halamka, MD, MS, President, Mayo Clinic Platform; and Jeff Anderson, PhD, MBA, Director of Accounts, Clinical Data Analytics Platform*

### ABSTRACT

Security and privacy are foundational to digital health care innovation. In September 2019, Mayo Clinic entered a 10-year partnership with Google to design a framework for the ethical secondary use of clinical data. This new infrastructure has two components: 1) the Mayo Clinic Cloud, which houses patient records, and 2) the Mayo Clinic Platform, a controlled enclave in which Mayo can share a subset of its clinical data, de-identified, and allow collaborators to link to it or supplement it with their own data for advanced analytics, including the development and training of artificial intelligence (AI) systems. This, unique approach called "data under glass" exemplifies a federated learning model. With algorithms permitted into the enclave and data never leaving the home institution, the Mayo-Google partnership illustrates an approach to how health systems and technology companies can partner to facilitate knowledge generation while addressing privacy and cybersecurity concerns. It also promotes data collaboration and knowledge generation by off-setting the costs of procuring, managing, and storing large amounts of data needed for algorithmic development.

### *Case Study at-a-Glance: Mayo–Google Partnership*

*Partnership between health system and industry that blends a federated learning approach with a data enclave that ensures privacy and security*

#### **Key Barriers Addressed**

- Insufficient use of available data at an enterprise level to support care improvement and generate insights
- Operational challenges of procuring and maintaining sufficient data necessary to develop and refine complex algorithms
- Lack of trust, particularly in large technology companies

#### **Specific Solutions for Data Sharing between Patients and Clinicians**

- Develop policy and technical controls to support de-identification of structured and unstructured data
- Use a data enclave, bringing users to data rather than releasing data to users, to enable assiduous data stewardship

#### **Insights for the Field**

- Unlock information in unstructured health data by leveraging the technical/computing capabilities of well-resourced technology firms
- Ensure transparency around health system–technology company partnerships as an essential component of building trust

## BACKGROUND

Mayo Clinic is an academic medical center and integrated health system comprised of three "shields"—patient care, education, and research. In September 2019, Mayo announced a 10-year partnership with Google. The partnership establishes a cloud-computing infrastructure, called Mayo Clinic Cloud, which is aimed at benefiting all three shields. The Cloud offers Mayo the ability to centrally store 1.2 million patient records (Furst, 2021). Through this partnership, Mayo also gains access to Google's AI toolsets, engineering talent, and security experience. Although Mayo considered all

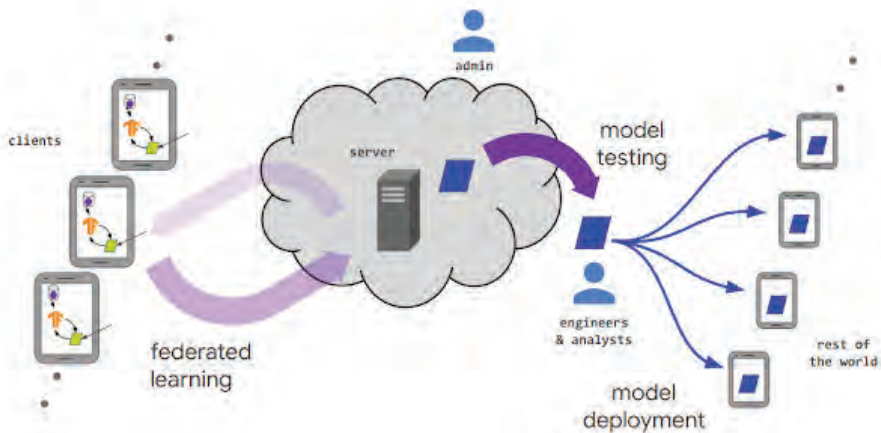
cloud providers, the health system chose Google because of the technology company's combinational strengths in these areas. These factors were important given Mayo's reliance on secondary use of data to support translational, clinical, and epidemiological research studies.

Overall, Mayo's recognition of the increasing digitalization of health care prompted the health system to undertake this transformation. The Mayo Clinic Platform is Mayo's strategic initiative to improve health care through insights and knowledge derived from data and partnerships. In 2020, it pursued three major initiatives:

- Clinical Data Analytics Platform – the process by which internal and external collaborators access the de-identified data in Mayo Clinic Cloud to discover new cures and treatments
- Home Hospital Platform – cloud-hosted components that enable serious and complex care at a distance
- Remote Diagnostic and Management Platform – ingestion of novel data from wearables and home-based devices that is combined with AI algorithms to deliver care recommendations to providers and patients.

Prior to engaging with Google, Mayo Clinic unified data from 70 diverse care sites into a longitudinal patient data store called the Universal Data Platform. This critical step made it easier for Mayo to migrate its structured and unstructured data to a private cloud container within the Google Cloud. The arrangement is analogous to renting a storage unit within a warehouse in which one puts their belongings and secures it with a lock. The warehouse owner cannot open the storage unit, and ownership of the belongings remains with the owner of the storage unit. Although Mayo Clinic built Mayo Clinic Cloud within the Google Cloud, Google is not able to access Mayo data independently since Mayo holds the key. Thus, Google is unable to combine Mayo patient data with data sourced from Google applications such as Search, Gmail, Google Maps, and YouTube.

For Mayo's internal operations, a patient-identified copy of Mayo's data is stored in the private cloud container under Mayo's control and not accessed by third parties. Another copy of the data, which is de-identified, is stored in a private cloud container that



**FIGURE 6** Diagram of a Federated Learning Model

SOURCE: Google AI Blog. 2021. *Google Research: Looking Back at 2020, and Forward to 2021*. Available at: <https://ai.googleblog.com/2021/01/google-research-looking-back-at-2020.html> (accessed December 6, 2021).

can be accessed by authorized third parties with Mayo's control for analytics. Third parties can develop novel algorithms, validate existing algorithms, and perform data analyses, but the data never leaves the Mayo container. Third parties can only take wisdom with them in the form of the finished algorithm or completed analysis. This is called "data behind glass."

The mechanism in its entirety represents a federated learning model (see *Figure 6*). Unlike a centralized data-sharing model in which a singular database hosts all of a person's accumulated health information and is the locus of aggregation and computing, the federated learning model allows Mayo to maintain physical and logical control of its data while selectively inviting investigators in and having the ability to audit what they do. The resulting learnings are distributed across private, academic, and federal research entities and are exchanged accordingly for further development (McMahan and Ramage, 2017).

## DESCRIPTION

Underpinning this partnership is a well-defined governance structure that consists of several layers of oversight for Mayo Clinic Cloud and Mayo Clinic Platform. A multi-stakeholder task force called "One Table" reviews data access requests and reports to Mayo's board of governors, who also weigh in on decision making. Nonetheless, executive leadership notes that the tradeoff with establishing a multi-level governance structure is speed and efficiency. However, investing in clear-cut decision-making processes is time well spent. In addition, the Health Data and Technology Advisory (DaTA) Board was created in 2021 and now has 11 members. Members are a diverse group of Mayo patients who live in the Rochester area, charged with providing perspectives and opinions on how potential AI and health technology applications, including the Google partnership and data sharing, will impact individual patients and the community as a whole.

Management of the partnership with Google is governed by a joint steering committee. Technical controls are complemented by policy controls. Due to privacy and ethical concerns that stem from third-party involvement, the steering committee is responsible for establishing a combination of policy and technical controls for regimenting data access and auditing. The technical controls block a third-party user's access if they connect to the cloud in an unsanctioned way. Policy controls prohibit partners from combining Mayo patient data with other data that could increase the risk of re-identification.

De-identification comes with challenges. From January to April of 2020, Mayo de-identified its structured data, including problem lists, medications, allergies, laboratories, and demographics. While the Health Insurance Portability and Accountability Act (HIPAA) mandates the removal of 18 types of direct and indirect identifiers, such as the patient's name, phone number, and, in some cases, ZIP code to render the data sufficiently de-identified, Mayo navigated instances where the identity of a patient could be deduced based on the combination of data available (HHS, 2015). As a result, Mayo employed both computer- and human-mediated mechanisms to redact datasets and the concept of "bin size" to assess if the data

is sufficiently de-identified. Studying domestic and international privacy law, Mayo came to the conclusion to use a bin size of 10, which means that a dataset was considered sufficiently de-identified if the data could be thought to be any one of 10 individuals in the database. Once Mayo performed de-identification, the health system sought certification from a third-party expert to verify that the data had a low likelihood of re-identification.

From April to August of 2020, Mayo de-identified the unstructured data, including clinical notes and reports, which program leaders admitted was a much more difficult task. The process required the removal of text that could enable easy re-identification. For example, if a note included the term "this senator" or specified "a star quarterback" from a named sports team, the note would be considered not sufficiently de-identified. Similarly, text may be typed into notes in a way that compromises privacy, such as the presence of phone numbers typed in a non-standard form (i.e., 5674328999). All such issues had to be addressed before the data could be certified as de-identified.

Operating amid growing skepticism and scrutiny of third-party collaborations, Mayo leadership has had to be conscientious about questions of feasibility and fostering stakeholder buy-in. In the interest of building trust and transparency, Mayo shares details of partnerships as soon as agreements are finalized with a number of leading publications and holds seminars with the broader health care community to gather feedback on its policies and procedures. In partnership with the Healthcare Information and Management Systems Society, it conducted a nationwide survey gauging consumers' attitudes vis-à-vis data sharing. Mayo Clinic Platform also works with community advisory boards, comprised of both patients and non-patients, to advise on topics such as Mayo's genomics data sharing policy. Equally important was ensuring the comfort of Mayo's own research community. This required educating the internal research community about the research benefit of these tools and addressing their concerns. Mayo also espouses the guiding principle of partnering only with external groups whose values align with its internal research practices.

## FUTURE DIRECTIONS

The interviewees for this case study cite the use of a federated learning model as the key factor to affording Mayo Clinic the agility and functionality to meet not only its data analytics goals, but also its stewardship responsibilities to patients. For health systems looking to emulate the Mayo-Google model, the interviewees advise starting small, thinking big, and moving fast. All of Mayo's data projects start as limited pilots, which are only expanded after lessons learned are thoroughly reviewed and risks mitigated. External partnerships and coalitions bring diverse experiences to innovation projects, so it is important for health care institutions to seek alliances with others.

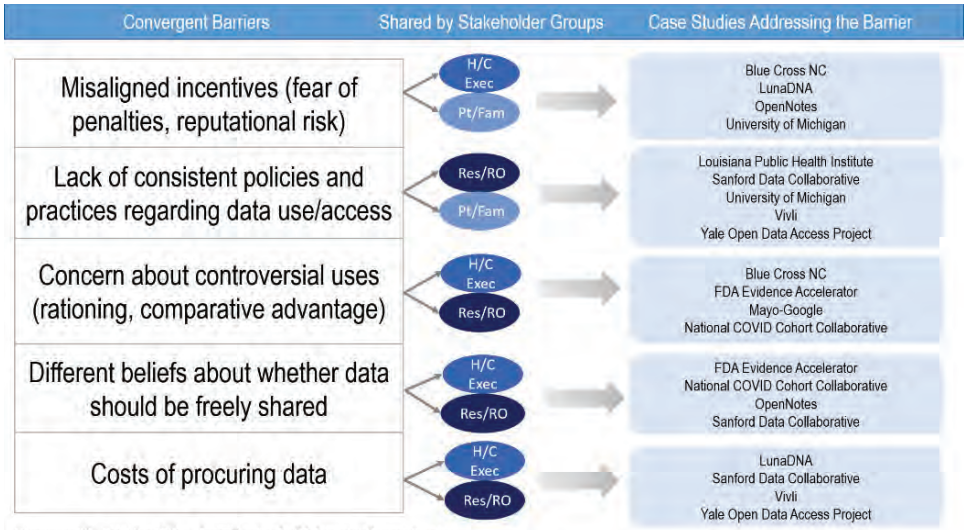
# 13

## CONCLUSION

The 11 exemplars described in this publication illuminate the possibilities for improving care, patient experience, and research when sharing health data among different stakeholders. Data exchange, when thoughtfully executed, can also support efforts to improve patient safety and even control health care costs through more efficient and effective use of information. Nine of the 11 cases highlighted in this publication were active and operational prior to the COVID-19 outbreak, and the ensuing pandemic only served to underscore the vital importance of maximizing all of the sources of health data at our collective disposal in support of better outcomes and continuous learning. The progenitor Special Publication underpinning this case study project, *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust*, elucidated many different cultural, financial, regulatory, and operational barriers to data sharing, but these examples offer many ways to attenuate the barriers and do so in a sustainable manner.

As shown in *Figure 7*, the case studies address multiple convergent barriers identified in *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust*. Blue Cross Blue Shield of North Carolina (Blue Cross NC) and Luna respond to financial barriers often expressed by health system executives—namely, the absence of a compelling business model for sharing data, as well as concerns about reputational risk and a related concern that retaining singular control of data is a way of maintaining competitive advantage in one’s market. Yet, these two cases, along with the Sanford Data Collaborative (SDC) and Mayo-Google partnership, indicate that a visible approach to data sharing can serve to bolster reputation and differentiate systems. However, health care markets vary widely at the local level, and additional insights





**FIGURE 7** Mapping of Case Studies to Priority Barriers

are needed from other health systems and entities where the market competition is more diffuse. Each case also offers viable and transferable lessons about maintaining stringent data privacy and security.

Concerns expressed by researchers and research oversight leaders in *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust* spanned different categories (e.g., regulatory, operational, cultural), but all barriers centered on a fundamental issue of heterogeneity. Different beliefs, different approaches, different processes, different costs, and different regulations (or their interpretation) impede researchers from maximizing opportunities to use data as a means of improving the evidence base and conducting impactful research. Nearly all of the cases profiled in this Special Publication take on at least one of these elements of “difference” and provide blueprints for more consistent, standardized approaches that are well-vetted by contributors. The University of Michigan (U-M) case shows the added value of standardized approaches to accelerate research, even in a single institution. Other case studies with roots in academia, Vivli and the Yale University Open Data Access (YODA) Project, crafted consistent approaches to data sharing to ease operational and fi-

nancial barriers. Each has a transparent review and request process, and in Vivli's case there is no upfront cost for the researcher to request and use data. In contrast, Luna employs a subscription model, charging institutions and granting user licenses to those who seek to analyze the de-identified, aggregated data.

A chief barrier expressed by patients and families was recognition—both of the value of data that they provide, and recognition that sharing clinical data with patients would accrue widespread bidirectional benefits. These aspects are at the core of participatory medicine, shared decision making, and patient empowerment. For patients who have tried to access their own health data, the Open-Notes case study provides a welcome solution—one that will only become more ubiquitous with the 2016 passage and 2021 implementation of new legislation prohibiting information blocking (Federal Register, 2020). With respect to financial disincentives, another barrier raised by the patients and families group, Luna is one of multiple such entities that seeks to apply a new business model of compensating patients for their data.

In consideration of this barrier raised by the patient stakeholders, all interviewees were asked whether and how patients were engaged in the development or implementation of the case activities; rarely did the respondents describe active and sustained patient involvement. The U-M team sought patient input, both via research studies and participation in community engagement “studios,” and the YODA project team engages patients as advisory board members. The SDC seeks to draw connections between research and health care outcomes as an element of its value proposition to patients receiving care at Sanford Health System. Each of these is a beneficial tether between patients and the use of their data, however, more concentrated and sustained engagement of diverse patients, families, and communities as data-sharing efforts mature is an important area needing attention.

The two use cases that emerged as a consequence of COVID-19, the National COVID Cohort Collaborative (N3C) and the COVID-19 Evidence Accelerator (EA), show how epochal events can spur rapid cultural, operational, and philosophical change. The common cause of unlocking the biology of the novel coronavirus and treatment of COVID-19 hastened collaboration and willingness to

share data and insights on a scale not previously seen. Similarly, Louisiana Public Health Initiative (LPHI) was spurred to rethink data availability, access, and sharing as a consequence of Hurricane Katrina. Those who might ordinarily “compete” on data or guard it for individual or a single organization’s needs embraced the value of sharing it for good. Taken together, these case studies illustrate the silver lining of opportunities born from crises. One key opportunity is to leverage messaging about the positive impact of data sharing and increase public awareness of its value. A second opportunity is to retain the efficiencies that manifested from the urgency of data exchange, including streamlined approaches to research review, protocol development, and contracting and other administrative elements.

## THEMES ACROSS CASE STUDIES

Conversations with case study representatives yielded several themes. Some of these were intangible and philosophical in nature, and others were more tangible and operational. One common thread among the cases was the sense of a *strong moral imperative*—a deeply felt sensibility that democratizing and sharing health data required their action. Leaders of these entities took it upon themselves to develop new and different approaches based on an intrinsic belief that democratizing data is an essential element of improving health and health care.

A second pervasive theme, which was more concrete in nature, was the central importance of investing time at the outset to address legal, regulatory, and technical barriers. Multiple interviewees held that early conversations with their organization’s legal teams were a critical facilitator of their success. Particularly for those venturing into new terrain, the opportunity to openly discuss issues with legal and regulatory officials helped identify and anticipate issues and engage in shared problem solving. This built trust for the data-sharing work as a whole and helped solidify enterprise-wide engagement.

Similarly, bringing together different personnel with health information technology or health information management responsibilities in the organization helped address questions related to

workflow, access controls, and related data security issues. These foundational conversations were common in all of the permutations explored in this compendium—that is, whether it was sharing data across different health insurers, preparing for use of OpenNotes, developing a health information exchange, or developing a repository between academia and industry—the importance of garnering upfront cooperation and input cannot be overstated.

An additional commonality in many cases was gaining organizational buy-in from influencers. Having an influential champion or sponsor for ideas that come from within an organization is a well-established tenet in business and organizational management literature. While they may not be at the highest level of the organization, these are individuals who can help overcome resistance to new ideas and changes. The experiences of OpenNotes, University of Michigan, and SDC, as well as that of Blue Cross NC and the N3C, all illustrate the importance of a champion, whether it is a specialty care leader, a chief medical informatics officer, or even a funder.

## PERSISTING CHALLENGES, "MAGIC WAND" INSIGHTS, AND ADVICE FOR OTHERS

Interviewees were asked about barriers they overcame as well as those that endure, framed as “if you had a magic wand and could change one aspect that would facilitate data sharing, what would you do?” This yielded a range of beneficial insights regarding the work ahead. The entreaty to *revisit regulations that undergird health data* was raised multiple times, as was the observation that *the business model and incentives for sharing data need reexamination*. Both of these notions are consistent with insights outlined in *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust* and provide additional substantiation for thoughtful reexamination and modernization of the entire policy landscape for health data—including policies that impact research, clinical care, payment models, interoperability, and information exchange. For example, the Health Insurance Portability and Accountability Act (HIPAA) was passed in 1996, long before health data were digitized, and before contemporary data sharing became a routine practice and important need for research and care. The

more recent passage of the European Union’s General Data Protection Regulations (GDPR) and California Consumer Protection Act (CCPA) empower individuals to control how their health data will be used, albeit these regulations have introduced new complexities, particularly with regard to use and sharing of data between nations. With respect to the business model and incentives for health data sharing, many of these efforts rely on at least some external grant funding (including OpenNotes, YODA, U-M, LPHI, and N3C), which can often be sustained up to a point, but is less optimal than stable, predictable funding from an organization’s core budget.

While advice to others did not necessarily yield a generalizable concrete blueprint, many interviewees *urged perseverance coupled with a “just do it” mindset*, averring that the work is eminently worth pursuing, despite the multifaceted challenges. Others emphasized the importance of *aligning the work of data sharing with the goals of all stakeholders*—whether in a single organization, or across different organizations. For example, the Mayo–Google partnership is congruent with Mayo Clinic’s strategic initiative to improve health care through data-derived knowledge. Similarly, in engaging regional health system partners, Blue Cross NC sought to identify specific issues of importance to each partner to augment the value proposition for collaboration.

## AREAS FOR FUTURE EXPLORATION

As noted above, a barrier expressed by researchers and research oversight leaders is the lack of shared principles about data ownership and sharing, and heterogeneity of beliefs about whether data should be shared. This perspective permeates how researchers interact with one another, but also influences attitudes about sharing data and knowledge with patients who volunteer to give their data and their time for research. Return of study results to participants is an important but inconsistent aspect of the research enterprise, and personal relevance of research is often a motivation for patients to volunteer for studies in the first place. As such, an objective of this work was to identify an organization that shares results with study volunteers as often as possible, and at an individualized versus summarized level. However, this proved elusive. Some or-

ganizations share study findings on an *ad hoc* basis, but this area warrants additional scrutiny since it can bolster patients' sense of trust in research and demonstrates respect for the contributions of research volunteers. Dedicated funding for this step in the research process, and an expectation by major funders that researchers will share data with participants can remediate this gap. It is encouraging that the National Institutes of Health's All of Us Research Program has specifically adopted the policy and practice of sharing data with all of their study participants, and the Program's experience can illuminate a pathway for others (NIH, 2021).

The case studies in this publication are U.S. based, but the global implications of data sharing and exemplars from other countries represent key next steps in this work. Different attitudes toward the public utility of data and national data repositories used by other countries indicate that future work could compare and contrast examples of building trust and addressing barriers with a global lens. A recent report authored by several European academies of science identified another data-sharing challenge that was exacerbated during the COVID-19 pandemic—the lack of unified regulations in the European Union and other countries (ALLEA, 2021). The report's authors observe that GDPR has impeded data sharing and international collaboration with countries that have less stringent privacy provisions than those in the GDPR. The report recommends recommitting to broader international discussion and coordination about regulations that can be better aligned and facilitate reciprocity in data sharing.

Finally, the experiences of Vivli, YODA, N3C, and the Evidence Accelerator, as well as other entities not profiled in this report (e.g., Project Data Sphere, Sentinel Initiative, TriNetX) demonstrate that there are multiple different workable models and approaches to sharing data, based on the nature of the collaboration and relationships between stakeholders. The opportunity to undertake a more robust comparison of the comparative advantages and disadvantages of federated, centralized, and intermediated approaches is another potential next step.

## CONNECTIONS BETWEEN CASE STUDIES AND THE PROGENITOR SPECIAL PUBLICATION

This Special Publication is relatively circumscribed in its breadth and scope, in that the case studies were identified through discussions among workgroup stakeholders who collaborated to produce the previous publication, *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust*, and the project was only intended to yield 10–12 cases. There are many other entities engaged in innovative approaches to health data sharing, including MIMIC (a freely accessible critical care database), the Light Collective (a patient advocacy and education group created to support responsible use of health data), newer ventures such as Truveta (a company formed in 2020 as a collaboration among health systems who agree to share their data to improve patient care), and initiatives specific to a certain type of data, like the Medical Imaging Data Resource Center (Light Collective, 2021; MIMIC, 2021; Truveta, 2021; MIDRC, 2021)<sup>1</sup>. As such, the selected case studies are meant to show what is possible and offer a springboard for others interested in advancing both the philosophical importance of health data sharing, and pragmatic approaches to doing so.

*Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust* enumerated several priority actions that could be achieved in one-to-three years, including this case study compilation. The compilation itself can provide insights for another of the action steps, the creation of a *public education campaign or similar national conversation* expressing the value and benefits of health data sharing, that helps individuals make better informed decisions about how they share health information, and helps organizations identify and uphold the highest ethical and technical standards for data use. The progenitor publication also included action steps related to *payment and financing*—both a reconsideration of how current payment models could thoughtfully integrate data sharing as a lever for supporting better care at

---

<sup>1</sup> The Light Collective. 2021. *Home*. Available at: <https://lightcollective.org/> (accessed May 9, 2021). Medical Imaging Data Resource Center (MIDRC). 2021. *Home*. Available at: <https://www.midrc.org/> (accessed October 3, 2021). Medical Information Mart for Intensive Care (MIMIC). 2021. *Home*. Available at: <https://mimic.mit.edu/iv/> (accessed April 29, 2021). Truveta. 2021. *Home*. Available at: <https://truveta.com> (accessed May 9, 2021).

lower cost and a deeper examination of viable business models that account for complexity of health data, estimations of its value, and the advantages of moving toward greater data sharing. In particular, the business model must account for the costs of *not sharing*, and the potential *loss* of competitive advantage if data are shared. Cases profiled in this publication provide a useful starting point for addressing many of these key aspects of the business case. Finally, the need for *supportive government policies* was called out in the previous publication and is echoed in many of the case studies. Newer regulations related to the 21st Century Cures Act and information blocking legislation are helpful, yet given the proliferation of health data sources, perhaps the most important new regulations will be modernization of the HIPAA Privacy Rule and expansion of the Genetic Information Nondiscrimination Act to include all medical information (Federal Register, 2020). The process of updating HIPAA Privacy Rule is underway; however, the timetable for the regulatory revisions is unclear, as is the resulting translation of any new regulations into on-the-ground process change. HIPAA became part of both the lexicon and culture of privacy in health care and modernizing it in support of a broader embrace of data sharing will take time and effort on the part of all stakeholders, including government, health system leaders, compliance personnel, clinicians, and of course, patients and families.

## THE PATH FORWARD

Given significant public and private investments in health research, health care solutions, and data aggregation and management utilities across the health ecosystem, and the unparalleled computational techniques that enable rapid analysis of exabytes of data, a trust fabric that supports the use and sharing of data is imperative. The concept of who owns health data remains contentious, especially given the liquidity and shifting value of data points when they are used alone or in combination with multiple sources such as financial or geolocation data. The traditional notion of what constitutes health data is particularly germane in consideration of health equity, as numerous social factors also help characterize the experience of health. Thus, ownership is an elu-



sive concept, but many stakeholders have a vested interest in the responsible stewardship and reinforcement of trust through their words and actions. Trade associations exist to support health information management. Many individual advocates and patient advocacy organizations offer positions on health data sharing, and some health and health care associations have issued statements about data use and data privacy. However, there is not a singular, visible advocacy organization that has health data sharing as its sole purpose—which may impede progress on regulatory and educational fronts. *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust* suggested that a consortium of organizations could advance dialogue and catalyze action. Though the case study interviewees were not asked about this explicitly, the collaborative tenor and enthusiasm of the participating organizations suggests that a unified consortium could accelerate progress and leverage the lessons of the exemplars featured here.

During the development of the previous publication, workgroup members were asked to consider not only the benefits of data sharing, but the consequences of *not sharing data*. The lost opportunities to maximize what can be learned at the individual and population level should serve to motivate action. Moreover, more robust health data sharing could help redress fundamental issues of data equity and algorithmic fairness by ensuring that larger and more diverse populations are represented in large data repositories and ensure that solutions harvested from health data truly confer societal benefit. As data become more proliferative and diverse, it is worth recalling that “Big Data” has often been characterized by a set of “V’s” with volume, variety, and velocity being commonly mentioned, along with veracity and value. Perhaps a more beneficial concept of Big Data with regard to sharing would include four A’s: health data that are *accessible, affordable, analyzable, and actionable* by all stakeholders—such a framework could galvanize public trust, forge better outcomes, and yield meaningful progress toward an equitable learning health system.

## REFERENCES

- All European Academies (ALLEA). 2021. *International Sharing of Personal Health Data for Research*. <http://www.doi.org/10.26356/IHDT>.
- Bell, S. K., T. Delbanco, J. G. Elmore, P. S. Fitzgerald, A. Fossa, K. Harcourt, S. G. Leveille, T. H. Payne, R. A. Stametz, J. Walker, and C. M. DesRoches. 2020. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Network* 3(6):e205867. <https://doi.org/10.1001/jamanetworkopen.2020.5867>.
- Bennett, T. D., R. A. Moffitt, J. G. Hajagos, B. Amor, A. Anand, M. M. Bissell, K. R. Bradwell, C. Bremer, J. B. Byrd, A. Denham, P. E. DeWitt, D. Gabriel, B. T. Garibaldi, A. T. Girvin, J. Guinney, E. L. Hill, S. S. Hong, H. Jiminez, R. Kavuluru, K. Kostka, H. P. Lehmann, E. Levitt, S. K. Mallipattu, A. Manna, J. A. McMurry, M. Morris, J. Muschelle, A. J. Neumann, M. B. Palchuk, E. R. Pfaff, Z. Qian, N. Qureshi, S. Russell, H. Spratt, A. Walden, A. E. Williams, J. T. Wooldridge, Y. J. Yoo, X. T. Zhang, R. L. Zhu, C. P. Austin, J. H. Saltz, K. R. Gersing, M. A. Haendel, C. G. Chute, and N3C Consortium. 2021. The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction. *medRxiv [Preprint]*. <https://doi.org/10.1101/2021.01.12.21249511>.
- Centers for Medicare & Medicaid Services (CMS). 2000. *Medicare and Medicaid Programs; Patient Protection and Affordable Care Act; Interoperability and Patient Access for Medicare Advantage Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of Qualified Health Plans on the Federally-Facilitated Exchanges, and Health Care Providers*. Available at: <https://www.federalregister.gov/documents/2020/05/01/2020-05050/medicare-and-medicaid-programs-patient-protection-and-affordable-care-act-in->

- teroperability—and (accessed November 26, 2021).
- DesRoches, C. M., S. K. Bell, Z. Dong, J. Elmore, L. Fernandez, P. Fitzgerald, J. M. Liao, T. H. Payne, T. Delbanco, and J. Walker. 2020. Patients managing medications and reading their visit notes: A Survey of OpenNotes participants. *Annals of Internal Medicine* 17(1):69–71. <https://doi.org/10.7326/M18-3197>.
- Dey, P., J. S. Ross, J. D. Ritchie, N. R. Desai, S. P. Bhavnani, and H. M. Krumholz. 2017. Data sharing and cardiology: Platforms and possibilities. *Journal of the American College of Cardiology* 70(24):3018–3025. <https://doi.org/10.1016/j.jacc.2017.10.037>.
- Federal Register. 2020. *21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program*. Available at: <https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification> (accessed December 27, 2021).
- U. S. Food & Drug Administration (FDA). 2020. *Real-World Evidence*. Available at: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (accessed May 28, 2021).
- Furst, J. 2021. *Meeting the Challenges of a Historic Year: 2020 Mayo Clinic Performance Highlights*. Available at: <https://newsnetwork.mayoclinic.org/discussion/meeting-the-challenges-of-a-historic-year-2020-mayo-clinic-performance-highlights/> (accessed May 28, 2021).
- Golbus, J. R., W. N. Price, and B. K. Nallamotheu. 2020. Privacy gaps for digital cardiology data: Big Problems With big data. *Circulation* 141(8):613–615. <https://doi.org/10.1161/CIRCULATIONAHA.119.044966>.
- Greater New Orleans Health Information Exchange (GNOHIE). 2018a. *Agreement Regarding Terms and Conditions for Health Information Exchange Participation*. Available at: [https://gnohie.org/wp-content/uploads/2018/06/GNOHIE\\_TC\\_4.30.18.pdf](https://gnohie.org/wp-content/uploads/2018/06/GNOHIE_TC_4.30.18.pdf) (accessed November 26, 2021).
- GNOHIE. 2018b. *Patient Opt-Out Request Form*. Available at: <https://gnohie.org/contact-main/patient-opt-out-contact-form/> (accessed November 26, 2021).

- Griese, E. R., B. S. Hsu, and D. A. Pearce. 2017. Extenuating circumstances: Data sharing: Leading data-driven population health. *South Dakota Medicine* 70(8):374-375. Available at: <https://research.sanfordhealth.org/-/media/research/files/services/sanford-data-collaborative/data-sharing/leading-data-driven-population-health.pdf?la=en&hash=B330128AAA3A6C794642B11FD1303F634BADF65E> (accessed November 29, 2021).
- Google AI Blog. 2021. *Google Research: Looking Back at 2020, and Forward to 2021*. Available at: <https://ai.googleblog.com/2021/01/google-research-looking-back-at-2020.html> (accessed December 6, 2021).
- Haendel, M. A., C. G. Chute, T. D. Bennett, D. A. Eichmann, J. Guinney, W. A. Kibbe, P. R. O. Payne, E. R. Pfaff, P. N. Robinson, J. H. Saltz, H. Spratt, C. Suver, J. Wilbanks, A. B. Wilcox, A. R. Williams, C. Wu, C. Blacketer, R. L. Bradford, J. J. Cimino, M. Clark, E. W. Colmenares, P. A. Francis, D. Gabriel, A. Graves, R. Hemadri, S. S. Hong, G. Hripscak, D. Jiao, J. G. Klann, K. Kostka, A. M. Lee, H. P. Lehmann, L. Lingrey, R. T. Miller, M. Morris, S. N. Murphy, K. Natarajan, M. B. Palchuk, U. Sheik, H. Solbrig, S. Visweswaran, A. Walden, K. M. Walters, G. M. Weber, X. Tanner Zhang, R. L. Zhu, B. Amor, S. G. Michael, L. M. Portilla, N. Qureshi, C. P. Austin, K. R. Gersing, and N3C Consortium. 2021. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *Journal of the American Medical Informatics Association* 28(3):427-443. <https://doi.org/10.1093/jamia/ocaa196>.
- HealthIT.gov. 2021. *Trusted Exchange Framework and Common Agreement*. Available at: <https://www.healthit.gov/topic/interoperability/trusted-exchange-framework-and-common-agreement> (accessed December 6, 2021).
- U.S. Department of Health & Human Services (HHS). 2000. *Standards for Privacy of Individually Identifiable Health Information*. Available at: <https://www.federalregister.gov/documents/2000/12/28/00-32678/standards-for-privacy-of-individually-identifiable-health-information> (accessed November 26, 2021).
- HHS. 2015. *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. Available at:

- <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html> (accessed May 28, 2021).
- Hsu, B., D. Nowak, and J. Smith. 2017. Reframing analytics: Transforming insights into action. *NEJM Catalyst*. <https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0467>.
- Kain, R., S. Kahn, D. Thompson, D. Lewis, D. Barker, C. Bustamante, C. Cabou, A. Casdin, F. Garcia, J. Paragas, A. Patrinos, A. Rajagopal, S. F. Terry, A. V. Zeeland, E. Yu, Y. Elrich, and D. Barry. 2019. Database shares that transform research subjects into partners. *Nature Biotechnology* 37: 1112–1115. <https://doi.org/10.1038/s41587-019-0278-9>.
- Khurshid, A., and L. Brown. 2014. How a Beacon Community Program in New Orleans helped create a better health care system by building relationships before technology. *eGEMs: The Journal for Electronic Health Data and Methods* 2(3):3. <http://dx.doi.org/10.13063/2327-9214.1073>.
- Louisiana Public Health Institute (LPHI). 2016. *Crescent City Beacon Community*. Available at: <https://lphi.org/work/clinical-transformation/crescent-city-beacon-community-ccbc/> (accessed May 28, 2021).
- Li, R., J. Wood, A. Baskaran, S. Neumann, E. Graham, M. Levenstein, and I. Sim. 2020. Timely access to trial data in the context of a pandemic: The time is now. *BMJ Open* 10:e039326. <https://doi.org/10.1136/bmjopen-2020-039326>.
- LunaDNA. 2020a. *Luna Terms of Use*. Accessed at: <https://lunadna.freshdesk.com/support/solutions/articles/43000454664-lunadna-terms-of-use> (accessed on November 26, 2021).
- LunaDNA. 2020b. *LunaDNA SEC Offering Circular*. Accessed at: <https://support.lunadna.com/support/solutions/articles/43000486909-lunadna-sec-offering-circular> (accessed on November 26, 2021).
- McClellan, M., M. Alexander, M. Japingam, and R. S. Saunders. 2019. North Carolina: The new frontier for health care transformation. *Health Affairs Blog*. Available at: <https://www.healthaffairs.org/doi/10.1377/hblog20190206.576299/full/> (accessed May 28, 2021).
- McMahan, B. and D. Ramage. 2017. Federated learning: Collaborative machine learning without centralized training data. *AI Google*

- Blog*. Available at: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html> (accessed May 28, 2021).
- Minakshi, R., R. De Vries, P. Nong, S. L. R. Kardia, and J. E. Platt. 2020. Do people have an ethical obligation to share their health information? Comparing narratives of altruism and health information sharing in a nationally representative sample. *PLoS One* 15(12):e0244767. <https://doi.org/10.1371/journal.pone.0244767>.
- National Academy of Medicine (NAM). 2020a. *NAM Leadership Consortium: Collaboration for a Value & Science-Driven Learning Health System: Evidence Mobilization Action Collaborative January 23, 2020 Meeting Highlights*. Available at: [https://nam.edu/wp-content/uploads/2020/01/1.23.20-Meeting-Highlights\\_Final.pdf](https://nam.edu/wp-content/uploads/2020/01/1.23.20-Meeting-Highlights_Final.pdf) (accessed December 6, 2021).
- NAM. 2020b. *NAM Leadership Consortium: Collaboration for a Value & Science-Driven Learning Health System: Digital Health Action Collaborative February 20, 2020 Meeting Highlights*. Available at: <https://nam.edu/wp-content/uploads/2020/01/2.20.20-Meeting-Highlights-FINAL.pdf> (accessed December 6, 2021).
- National Center for Advancing Translational Sciences (NCATS). 2021. *Clinical and Translational Science Awards (CTSA) Program*. Available at: <https://ncats.nih.gov/ctsa> (accessed November 26, 2021).
- National Center for Data to Health (NCDH). 2021. *CD2H: Harmonizing the Informatics Community*. Available at: <https://cd2h.org/> (accessed December 6, 2021).
- National COVID Cohort Collaborative (N3C). 2021a. *Home*. Available at: <https://covid.cd2h.org/n3c> (accessed December 6, 2021).
- N3C. 2021b. *Data Use Request (DUR) Process*. Available at: <https://covid.cd2h.org/dur> (accessed December 6, 2021).
- National Heart, Lung, and Blood Institute (NHLBI). 2014. *NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies*. Available at: <https://www.nhlbi.nih.gov/grants-and-training/policies-and-guidelines/nhlbi-policy-for-data-sharing-from-clinical-trials-and-epidemiological-studies> (accessed September 8, 2021).
- NHLBI. 2021. *Biologic Specimen and Data Repository Information Coordinating Center*. Available at: <https://biolincc.nhlbi.nih.gov/home/> (accessed September 8, 2021).

- National Institutes of Health (NIH). 2021. *Precision Medicine Initiative: privacy and trust principles*. All of Us Research Program. Available at: <https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-privacy-and-trust-principles> (accessed May 28, 2021).
- OpenNotes. 2021. *OpenNotes Compliance Date Extended to April 5, 2021*. Available at: <https://www.opennotes.org/news/open-notes-compliance-date-extended-to-april-5-2021/> (accessed on November 26, 2021).
- OpenNotes. 2020. *New Survey Data Reveals 54 Million People are Able to Access Clinicians' Visit Notes Online*. Available at: <https://www.opennotes.org/news/new-survey-data-reveals-54-million-people-able-to-access-clinicians-visit-notes-online/> (accessed May 27, 2021).
- Ouellette, P. 2013. A look at the New Orleans HIE security infrastructure. *Health IT Security*. Available at: <https://healthitsecurity.com/news/a-look-at-the-new-orleans-hie-security-infrastructure> (accessed May 28, 2021).
- Pew Charitable Trusts. 2021. *Most Americans Want to Share and Access More Digital Health Data*. July 27. Available at: <https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2021/07/most-americans-want-to-share-and-access-more-digital-health-data> (accessed July 29, 2021).
- Price, W. N., M. E. Kaminski, T. Minssen, and K. Spector-Bagdady. 2019. Shadow health records meet new data privacy laws. *Science* 363(6426):448-450. <https://doi.org/10.1126/science.aav5133>.
- Reese, E. C. 2020. *The healthcare value imperative: All eyes on North Carolina's move to value-based payment*. Health Care Financial Management Association. Available at: <https://www.hfma.org/topics/hfm/2020/february/healthcare-value-imperative-north-carolinas-move-value-based-payment.html> (accessed May 28, 2021).
- Reagan-Udall Foundation and Friends of Cancer Research. 2021. *COVID-19 Evidence Accelerator*. Available at: <https://evidenceaccelerator.org> (accessed November 26, 2021).
- Reagan-Udall Foundation and Friends of Cancer Research. 2020. *COVID-19 Evidence Accelerator Summary*. Available at: [https://evidenceaccelerator.org/sites/default/files/2020-05/COVID-19\\_](https://evidenceaccelerator.org/sites/default/files/2020-05/COVID-19_)

- Evidence\_Accelerator-SUMMARY\_4.10.20.pdf (accessed September 8, 2021).
- Roe, L., A. Abernethy, J. Franklin, E. Sigal, J. Allen, A. Bhat, C. Rodriguez-Watson, and S. Winckler. 2021. *Accelerating Evidence Generation by Convening Diverse Stakeholders Across the Real-World Data Ecosystem*. Available at: <https://evidenceaccelerator.org/sites/default/files/2021-08/Ev%20Acc%20Lab%20Focus%20Science%20Poster%202021%20Final%20with%20OCC%20review.pdf> (accessed December 6, 2021).
- Ross, J. S., J. Waldstreicher, S. Bamford, J. A. Berlin, K. Childers, N. R. Desai, G. Gamble, C. P. Gross, R. Kuntz, R. Lehman, P. Lins, S. A. Morris, J. D. Ritchie, and H. M. Krumholz. 2018. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Science Data* 5:180268. <https://doi.org/10.1038/sdata.2018.268>.
- Ross, T. R., D. Ng, J. S. Brown, R. Pardee, M. C. Hornbrook, G. Hart, and J. F. Steiner. 2014. The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. *eGEMS: The Journal for Electronic Health Data and Methods* 2(1):1049. <https://doi.org/10.13063/2327-9214.1049>.
- Spector-Bagdady, K., R. Hutchinson, E. O'Brien Kaleba, and S. Khetarpal. 2020. Sharing health data and biospecimens with industry — A principle-driven, practical approach. 2020. *New England Journal of Medicine* 382(22):2072-2075. <https://doi.org/10.1056/NEJMp1915298>.
- Spector-Bagdady, K., R. G. De Vries, M. G. Gornick, A. G. Shuman, S. Kardia, and J. Platt. 2018. Encouraging participation and transparency in biobank research. *Health Affairs* 37(8):1313-1320. <https://doi.org/10.1377/hlthaff.2018.0159>.
- Sun, L. H. 2014. Boston hospital pilot gives patients electronic access to their therapists' notes. *The Washington Post*, May 18. Available at: [https://www.washingtonpost.com/national/health-science/boston-hospital-pilot-gives-patients-electronic-access-to-their-therapistsnotes/2014/05/18/2d891bac-cfe5-11e3-a6b1-45c4dffb85a6\\_story.html](https://www.washingtonpost.com/national/health-science/boston-hospital-pilot-gives-patients-electronic-access-to-their-therapistsnotes/2014/05/18/2d891bac-cfe5-11e3-a6b1-45c4dffb85a6_story.html) (accessed December 6, 2021).
- University of Illinois Chicago (UIC). 2018. *Increasing Patient Access to Medical Records*. Available at: <https://healthinformatics.uic.edu/blog/increasing-patient-access-to-medical-records/> (ac-



- cessed May 27, 2021).
- University of Michigan Medical School (UMMS). 2018. *Data Policy Checklist*. Available at: <https://research.medicine.umich.edu/data-policy-checklist> (accessed November 26, 2021).
- UMMS. 2019. *Data & Biospecimen Sharing*. Available at: <https://research.medicine.umich.edu/ourunits/data-office-clinical-translational-research/data-access/data-biospecimen-sharing> (accessed November 26, 2021).
- U.S. Congress. 2016. *H.R.34 – 21st Century Cures Act*. Available at: <https://www.congress.gov/bill/114th-congress/house-bill/34> (accessed on November 26, 2021).
- U.S. Department of Veterans Affairs (VA). 2017. *What Happens When Patients Access Their Mental Health Providers' Notes?* Available at: <https://www.research.va.gov/currents/0217-8.cfm> (accessed November 26, 2021).
- Vivli. 2021. *Our Members*. Available at: <https://vivli.org/members/ourmembers/> (accessed May 28, 2021).
- Vivli. 2020. *What is Vivli?* Available at: <https://vivli.org/faq/what-is-vivli/> (accessed May 28, 2021).
- Weeks, J. and R. Pardee. 2019. Learning to share health care data: A brief timeline of influential common data models and distributed health data networks in U.S. health care research. *eGEMs: The Journal for Electronic Health Data and Methods* 7(1): 4. <https://doi.org/10.5334/egems.279>.
- Whicher, D., M. Ahmed, S. Siddiqi, I. Adams, M. Zirkle, C. Grossmann, and K. L. Carman, *Editors*. 2020. *Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust*. NAM Special Publication. Washington, DC: National Academy of Medicine. Available at: [https://nam.edu/wp-content/uploads/2020/11/Health-Data-Sharing-to-Support-Better-Outcomes\\_prepub-final.pdf](https://nam.edu/wp-content/uploads/2020/11/Health-Data-Sharing-to-Support-Better-Outcomes_prepub-final.pdf) (accessed May 27, 2021).
- Wiens, J., S. Suchi, M. Sendak, M. Ghassemi, V. X. Liu, F. Doshi-Velez, K. Jung, K. Heller, D. Kale, M. Saeed, P. N. Ossorio, S. Thadane-Israni, and A. Goldenberg. 2019. Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine* 25(9):1337-1340. <https://doi.org/10.1038/s41591-019-0548-6>.
- Yale University Open Data Access Project (YODA). 2021. *Welcome to the YODA Project*. Available at: <https://yoda.yale.edu/welcome->

yodaproject (accessed September 8, 2021).

YODA. 2019. *Procedures to Guide External Investigator Access to Clinical Trial Data*. Available at: <https://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data> (accessed May 28, 2021).

YODA. 2014. *Data Holders*. Available at: <https://yoda.yale.edu/data-holders> (accessed May 28, 2021).

# APPENDIX A

## KEY INFORMANT INTERVIEW QUESTIONS

### NAM Data Sharing Project – Key Informant Interview Questions (updated 8/31/2020)

#### 1) Overarching rationale/motivation:

1. What was the motivation for sharing data with this/these group(s)?
2. What barrier(s) were you specifically seeking to address/overcome?
3. What financial or other incentives spurred this arrangement between the partners?

#### 2) Nature of the Data Sharing

1. Describe the type(s) of data that are shared
2. What major facilitators led to the success of your data sharing work?
3. With whom were data shared, and in what form?

#### 3) Nature of the Partnership and Stakeholders

1. Describe the partnership, including the process for decision making, garnering cooperation, and gaining consensus.
2. Who drove the collaboration, and what level of leadership did you have at the table?
3. What public commitments were in place?
4. How did you address resistance?
5. Were patients at the table and, if so, what role did they play?

#### 4) Challenges during the Process

1. Describe the legal and consent arrangements, and how they may have evolved before the data exchange and once the data sharing had occurred.
2. If a party wanted to remove their data, or withdraw from the data-sharing partnership, what would that look like?
3. What obstacles did you encounter before/during/after data were exchanged? How were they addressed?
4. Did you encounter any new challenges once the data sharing had commenced, and if so, how were they resolved?
5. Was there any regulatory guidance that helped or hindered this work? If so, how?

#### 5) Looking toward the Future

1. If you could wave a magic wand, what kind of guidance or policy recommendations would be helpful to ameliorate the barriers to more robust and reasonable data sharing?
2. What advice do you have for others who want to undertake something similar—a key take home message for others?