

## Cancer: From a Kingdom to a Commons

*Michael Mattioli*

### INTRODUCTION

In the late 1970s, decades before leukemia ended her life, the American writer Susan Sontag wrote, “Everyone who is born holds dual citizenship, in the kingdom of the well and in the kingdom of the sick.”<sup>1</sup> We prefer to spend our days in the good country, but sooner or later, said Sontag, “we are all obliged to identify as citizens of that other place.” In his 2010 Pulitzer Prize–winning book on cancer, Dr. Sidhartha Mukherjee invoked a similar geographic metaphor when he referred to the disease as a vast “empire.”<sup>2</sup> In a letter to a friend, the novelist Thomas Wolfe once called his cancer “a strange country.” By describing illness as a passport, these commentators conveyed the sense of dislocation that a cancer diagnosis can bring. But what if our doctors – our guides to this “other place,” such as it is – could, by working together, redefine its national character? What if the sick were not subjects of a sovereign, but instead, members of a commons that heals itself? This possibility is at the heart of a new and hopeful movement in cancer research.

This chapter explores a privately governed collaborative composed of doctors and hospitals that seeks to aggregate, manage access to, and draw insights from oncology treatment data.<sup>3</sup> The thesis behind this collaboration is simple: if cancer treatment

Michael Mattioli is Associate Professor of Law at the Indiana University Bloomington Maurer School of Law, Bloomington, Indiana.

<sup>1</sup> Susan Sontag, *Illness as Metaphor* 3 (McGraw Hill 1977).

<sup>2</sup> Siddhartha Mukherjee, 9 *The Emperor of All Maladies: A Biography of Cancer* (Scribner 2010).

<sup>3</sup> In a companion law journal article, I explore the implications of health data pooling on innovation policy more generally. Michael Mattioli, The Data Pooling Problem, 32 *Berkeley Tech. L.J.* (2017). This study builds upon the earlier work of commentators from diverse fields of expertise, including law, medicine, and political science. See, e.g., Louise Trubek et al., Improving Cancer Outcomes Through Strong Networks and Regulatory Frameworks: Lessons From the United States and the European Union, 14 *Journal of Health Care Law & Policy* 119 (2011). Although this chapter focuses on private data-pooling, other commentators have recently written on public-private partnerships designed to facilitate cancer data-sharing. See Jacob S. Sherkow, Cancer’s IP, 96 *North Carolina L. Rev.* (2018).

data could be aggregated on a large scale, scientists believe they would be able to select more effective treatments for particular cancers, and even for particular patients. Over time, this cooperative process could theoretically spur a “virtuous cycle” of medical advancement: more effective treatments would encourage more doctors, hospitals, and care centers to contribute even more data, which, in turn, would improve the quality of care, and so on.

This data-intensive approach to cancer treatment draws heavily upon the surprising power of correlations. For centuries, researchers have used the scientific method to grasp the *how* and *why* behind poorly understood phenomena – the underlying genetic or biochemical causes of an illness, for instance. While this approach can lead to useful treatments, it is not a particularly efficient or effective way to treat a single sick individual. It is often enough, rather, to understand and act upon a set of factors that *correlate with* a disease.<sup>4</sup> As two commentators from the field of computer science recently explained, “If millions of electronic medical records reveal that cancer sufferers who take a certain combination of aspirin and orange juice see their disease go into remission, then the exact cause for the improvement in health may be less important than the fact that they lived.”<sup>5</sup> Cancer research and treatment is an endeavor in which, to quote Voltaire, “the perfect is the enemy of the good.”<sup>6</sup>

Physicians have long embraced the power of correlations. Sidney Farber, the Boston physician and “grandfather” of cancer research, once stated in a congressional hearing that, “The history of Medicine is replete with examples of cures obtained years, decades, and even centuries before the mechanism of action was understood for these cures.”<sup>7</sup> In 1854 London, the physician John Snow noticed a correlation between the home addresses of cholera victims and the location of a community water pump. Snow could not prove exactly *how* cholera spread through the water – Louis Pasteur would not develop “germ theory” until the following decade – but he urged officials to remove the handle from the pump and the spread of the disease halted soon after. In the 1950s, the English researchers Austin Bradford Hill and Richard Doll similarly drew upon correlations to conclude that cigarette smoking causes lung cancer. The two men compared a data set of English physicians who reported that they regularly smoked against a national registry that listed the causes of death of English physicians. This comparison revealed a strong correlation between smoking and lung cancer.

<sup>4</sup> Some commentators offer a different, more descriptive term for these methods: “data-intensive science.” Tony Hey et al. (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research 2009).

<sup>5</sup> Victor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt 2013).

<sup>6</sup> This aphorism is widely attributed to Voltaire. Susan Ratcliffe (ed.), *Concise Oxford Dictionary of Quotations* 389 (Oxford University Press 2011).

<sup>7</sup> Hearings before the Subcommittee on Public Health and Environment of the Committee on Interstate and Foreign Commerce, House of Representatives, Ninety-Second Congress, First Session, on the National Cancer Attack Act of 1971, H.R. 8843, H.R. 10681, S. 1828, Serial No. 92-40, at 271.

Correlations in health data allow doctors to effectively treat diseases they do not yet fully understand.

Today, health data is at once more detailed and more scattered than it was in Snow's time. This diffusion of information is particularly pronounced in the field of oncology. One commentator characterized the entire corpus of cancer data recorded in the United States as "utterly fragmented."<sup>8</sup> Health care providers – from small practices to large hospitals – are primary sources of patient treatment data. (As described in Section 7.4, such data may include machine readings, the results of physical exams and chemical tests, a doctor's subjective observations, etc.) Hospitals typically send this data to outside electronic health record (EHR) vendors that store and manage access to it. Shadows of the same information could be reflected in other places, though, such as in a hospital's billing department or in a medical insurance provider's database.<sup>9</sup> Some individual patients also retain their own private treatment records.<sup>10</sup> Pharmaceutical companies maintain troves of records related to clinical trials performed on human subjects; academic researchers store (and often jealously guard) valuable scientific data sets related to their work; online services store health data generated by fitness and health gadgets and smartphone apps. The list of oncology data stewards goes on – from drugstores to social networks.

This fragmented informational landscape presents a collective action problem that necessarily precedes the search for useful correlations: how to aggregate health data from the many institutions that hold it. Today, a researcher who wishes to search for correlations between, say, oncology data held by two hospitals in different states faces significant difficulties. The researcher would need to first learn which hospitals hold the data she wished to analyze – a potentially costly and time-consuming project in itself. She would then need to negotiate with each hospital to obtain access to the data. This, too, could impose upfront costs, including the time and money involved with negotiations, drafting contracts, and the high likelihood that either hospital (or both) simply will not agree to share their data. (As discussed in Sections 7.2.2 and 7.6.3, institutions that hold useful data often have strong practical disincentives to disclose it.)

A recent effort led by the government of the United Kingdom to aggregate the health data of its citizens brought these challenges vividly to light. Launched in 2013, the UK National Health Services' National Programme for IT (nicknamed "care.data") aimed to centralize patient records stored by general practitioners and hospitals across England. The project's stated goal was to "assess hospital safety,

<sup>8</sup> Clifton Leaf, Why We're Losing the War on Cancer, *Fortune* (Mar. 22, 2004).

<sup>9</sup> For more on the data held by insurance companies, and the potential this data holds, see Rebecca S. Eisenberg and W. Nicholson Price II, Promoting Healthcare Innovation on the Demand Side, 4 *Journal of Law and the BioSciences* 49 (2017).

<sup>10</sup> The *New York Times* recently reported on an MIT doctoral student and cancer survivor who, through great effort, collected and published a wealth of data generated on his treatment. Steve Lohr, The Healing Power of Your Own Medical Records, *New York Times* (Mar. 31, 2015).

monitor trends in various diseases and treatments, and plan new health services.”<sup>11</sup> The plan was beset with problems from the outset. Press reports and an academic study describe several causes of trouble, including gross mismanagement and conflicting interests: politicians and project managers pressed doctors across England for their patients’ data without giving them sufficient time – just two months – to provide their patients with an opportunity to opt out, and without giving enough thought to the legal implications for patient privacy.<sup>12</sup> Unsurprisingly, this led to a strong public backlash. Meanwhile, physicians widely opposed the plan for reasons that were more cultural. As one academic study explained, “General practitioners view the medical record as a reflection of their professional relationship with the patient and their status as protectors of the record. The record is sacrosanct.”<sup>13</sup> As a result of this widespread backlash, the project stalled and shows no signs of resuming operations.

During the same time period that the care data episode unfolded, several private initiatives formed in the United States with similar goals. Among these are CancerLinQ (the American Society of Clinical Oncology), Project Data Sphere (Celgene and Pfizer), Cancer Commons (a nonprofit group), and the Data Alliance Collaborative (Premier Healthcare). These collaboratives differ somewhat in their approaches, but they share seemingly elegant goals: to mediate the pooling of oncology data from various sources, to organize this information for research, and to provision access to it. In theory, institutions and individuals who contribute data to these groups could benefit by enjoying access to shared insights about how best to treat cancer. Meanwhile, outside researchers could benefit by having a “one-stop shop” for the data they need. It is a compelling institutional model reminiscent (if only superficially) of a patent pool or a copyright licensing collective.

These private data-gathering groups appear to be “commons” in the traditional sense that Nobelist Elinor Ostrom used that term. In her path-breaking book, *Governing the Commons*, Ostrom defined commons as institutions “governed by collective action whereby a group of participants can organize themselves voluntarily to retain the residuals of their efforts.”<sup>14</sup> Ostrom inspired a generation of scholars from many domains – this author included – to study how commons govern production and access to valuable intangible assets, such as patents and data. Ostrom and her intellectual followers have showcased the economic, political, and social advantages of these collaboratives, in particular, how they can dramatically reduce

<sup>11</sup> NHS England. Privacy Impact Assessment: Care.data (2014), [www.england.nhs.uk/wp-content/uploads/2014/04/cd-pia.pdf](http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-pia.pdf)

<sup>12</sup> Abandoned NHS IT system has cost 10bn so far, *The Guardian* (Sept. 13, 2013), [www.theguardian.com/society/2013/sep/18/nhs-records-system-10bn](http://www.theguardian.com/society/2013/sep/18/nhs-records-system-10bn)

<sup>13</sup> Presser L., Hruskova M., Rowbottom H. and Kancir J., Care.data and Access to UK Health Records: Patient Privacy and Public Trust, *Technology Science* (Aug. 11, 2015), <http://techscience.org/a/2015081103>.

<sup>14</sup> Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* 24 (Cambridge University Press 1990).

transaction costs and barriers to aggregation of complementary assets. As private cooperatives, commons also operate free from “the leviathan” of direct governmental control and the temptations for opportunistic behavior created in market environments. This body of scholarship suggests that oncology data commons may offer a practical solution to a pressing public health problem.

Drawing upon the knowledge commons analytical framework, this chapter presents an ethnographic study of CancerLinQ – an early-stage oncology data commons created by the American Society of Clinical Oncology (ASCO).<sup>15</sup> The purpose of this study is to explore, through a single deep case study, the extent to which privately governed medical commons may be able to solve the increasingly urgent problem of aggregating oncology data for large-scale analysis, and along the way learn more about the challenges they face, and how well they are poised to advance the state of cancer treatment.<sup>16</sup> These are urgent questions that merit deep study.<sup>17</sup>

Section 7.1 explains the methodology of this case study. Section 7.2 describes the background contexts that define the world of oncology data, including the nature of cancer itself, the institutions that generate oncology data, the recent ascendancy of data-intensive science (“Big Data”), and relevant laws and regulations. Section 7.3 describes the type of data aggregated by CancerLinQ, including how this data is generated, where it is typically stored, and what legal rules apply to its use. Section 7.4 explains the variety of institutions involved with CancerLinQ. Section 7.5 lays out the goals and objectives of the initiative, and some of the challenges that appear to lay ahead. A brief conclusion follows. These are still early days for efforts to aggregate privately held cancer treatment data; as such, this chapter’s primary goal is not to cast judgment on CancerLinQ but rather to identify the broad challenges and opportunities that this effort and others like it may face.

## 7.1 METHODOLOGY

This chapter’s approach follows Elinor Ostrom’s Institutional Analysis and Development (IAD) framework, as adapted by Strandburg, Frischmann, and Madison for the study of knowledge commons.<sup>18</sup> This involved the following:

- **A literature review.** To identify likely interview candidates and to gather general information about the state of data-intensive cancer research, I

<sup>15</sup> <http://cancerlinq.org>

<sup>16</sup> For a wider-angle view of data pooling as it relates to innovation policy generally, and for the results of my analysis of different health data commons, see Michael Mattioli, *The Data Pooling Problem*, 32 *Berkeley Tech. L.J.* (2017).

<sup>17</sup> Miguel Helft, *Can Big Data Cure Cancer?* *Fortune* (July 24, 2014), <http://fortune.com/2014/07/24/can-big-data-cure-cancer/>

<sup>18</sup> Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014).

surveyed recently published books, newspaper articles, and academic works related to this topic. This research also covered general interest publications on the so-called Big Data phenomenon and on cancer. From these sources, I identified private efforts designed to pool privately held oncology data, and the names of individuals involved with these efforts.

- **Semi-structured interviews.** I interviewed 10 professionals currently or formerly involved with ASCO's CancerLinQ project or with significant knowledge of it through related work in the field of oncology care. In keeping with the knowledge commons framework, these interviews were semi-structured and organized around the following topics: (1) the scientific, technological, and social contexts in which the project has taken form; (2) the various types of data and related informational assets the group seeks to aggregate and organize access to; (3) the "default" status of these assets; (4) the players involved, including corporations and health care institutions; (5) the project's goals; (6) rules and related internal governance mechanisms; (7) the technological infrastructure supporting the project.

All interviews were conducted by telephone, recorded with the permission of the interview subjects, and professionally transcribed. The average duration of the interviews was 45 minutes. Some interviews were supplemented with brief follow-up email exchanges. In keeping with Internal Review Board procedures, subjects were furnished with an information sheet describing the goals of this study. With the help of a research assistant, I reviewed and flagged portions of each transcript to identify common themes and topics.

## 7.2 BACKGROUND ENVIRONMENT: CONTEXTS

It is helpful to begin the study of any knowledge commons by surveying the environment in which it has formed. This section focuses on the most prominent landmarks in the oncology data landscape: cancer itself, the people and institutions that generate data in the pursuit of treating it, the recent ascendance of Big Data, the legal rules and regulations that apply to oncology data, and at the center of it all, patients.

### 7.2.1 *The Biological Context*

In his bestselling book, *The Emperor of All Maladies*, Dr. Siddhartha Mukherjee defines cancer with plainspoken eloquence: "Cancer is not one disease but many diseases," he writes, "We call them all 'cancer' because they share a fundamental feature: the abnormal growth of cells."<sup>19</sup> In practical terms, the *kind* of abnormal

<sup>19</sup> Mukherjee, *The Emperor of All Maladies* xiii (Author's note).

growth that we call “cancer” involves cells proliferating uncontrollably through parts of the body where they do not belong, and interfering with vital internal processes along the way. Cancerous cells may develop into tumors, and even more treacherously, they may spread widely throughout the body and mutate into different forms as they go. This process is called “metastasis.”

The problem’s cause is genetic. Francis Crick and James Watson’s discovery of DNA in the 1950s revealed that every cell of the human body contains a set of chemical blueprints for the entire organism. In simplified terms, every cell in the body divides according to what its local, internal copy of the blueprint says. When a healthy cell within, say, a human lung divides, it creates a duplicate of itself. Under normal circumstances, the original cell and its offspring both contain “correct” instructions and, as such, stop replicating when new lung cells are not needed. Cancerous cells, by contrast, contain mutated DNA that gives incorrect instructions on when cells should stop dividing. The result is runaway cell growth. The bulk of cancer research over the past 30 years has focused on understanding what causes DNA to mutate in this dangerous way, and what drugs and therapies can halt the process.

Cancer presents a puzzle of uncertainties to doctors who endeavor to treat it. As cancerous cells spread through the human body, they often become increasingly heterogeneous – that is, one mutation develops into *many* mutations, affecting different cells. Because different mutations may respond well to different drugs, cancer patients are often prescribed cocktails of therapies. For researchers who seek to understand which drugs work best and why, these combinational treatments present, in a sense, equations with too many unknown variables. If a patient responds well to a particular cocktail of therapies, this may be because one drug in the mix worked well, because several drugs worked well independently, or because there was a beneficial interaction between two or more of the drugs. The unknowns multiply when one considers that each patient’s unique physiology can also play a role, and that some cancers can evolve to resist certain therapies. As a subject interviewed for this chapter put it, “you are dealing with thousands of cancers and hundreds of therapies, and often times you have to use multiple therapies in combinations or else the cancer will just evolve around the therapy ... you just can’t test all of the variations and combinations in large trials.”<sup>20</sup>

As explained in greater depth in Section 7.2.3, a vast collection of treatment data could make this puzzle solvable. Although it may, for now, still be impossible to understand *why* a particular drug or a combination of drugs works well for a single patient, it may be possible to learn which kinds of patients (i.e., what shared characteristics) respond well to certain treatments through statistical analysis of large sets of treatment data. To make that possible, it is necessary to first pool treatment data from many patients.

<sup>20</sup> Marty Tenenbaum, CEO, Cancer Commons.

### 7.2.2 The Oncology Data Context

Health providers, academic research institutions, and pharmaceutical companies: if cancer is a kingdom, these institutions are its counties, towns, and precincts – the points of contact between “the emperor” and its subjects. For CancerLinQ and similar projects seeking to pool treatment data, these institutions are where the data is generated.

Oncologists routinely collect and generate a wealth of useful treatment data. Approximately 13,000 board-certified clinical oncologists practice in the United States today.<sup>21</sup> They treat patients in group practices, medical schools, hospitals, specialty cancer treatment centers, and other clinical settings. Most oncologists are members of the American Society of Clinical Oncology (ASCO) – a professional organization founded in the 1960s, which holds conferences and workshops and publishes journals and books related to oncology practice. Oncologists interviewed for this chapter explained that cancer care is a relatively small and close-knit community, and most doctors interact regularly through professional meetings, conferences, and symposia. Of course, these statements reflect the subjective perceptions of only the relatively small set of individuals interviewed for this study. As an empirical matter, it is difficult to gauge just how close-knit the oncology profession as a whole truly is.

Academic researchers also generate and maintain access to sets of oncology data.<sup>22</sup> For these experts, scholarly publications represent a pathway to future grants, wider recognition, and tenure. As Jorge Contreras has astutely observed, however, “journal publication has proven to be inadequate for the dissemination of genomic data.”<sup>23</sup> In part, this has led federal grant-awarding agencies such as the National Cancer Institute, which is part of the National Institutes of Health (NIH) to require the scholars they fund to release their data publicly. While such data research requirements may have sped the dissemination of data in some fields of medical science, experts interviewed for this study reported that this is not the case in the field of oncology. One prominent professor of medicine commented that in practice, academic researchers obligated to disclose their data often obfuscate useful data in a deluge of ancillary information. “When the NIH promotes sharing,” the expert explained, “people will share but then they make it really hard for anybody else to

<sup>21</sup> This chapter was written in 2015 and 2016. Key Trends in Tracking Supply of and Demand for Oncologists, ASCO (Jan. 2013), [www.asco.org/practice-research/facts-and-figures-diversity-oncology](http://www.asco.org/practice-research/facts-and-figures-diversity-oncology) (reporting the precise figure as 13,084 oncologists). See also *The State of Cancer Care in America: 2015*, [www.asco.org/sites/www.asco.org/files/2015ascostateofcancercare.pdf](http://www.asco.org/sites/www.asco.org/files/2015ascostateofcancercare.pdf) (reporting additional relevant statistics and a census of the oncology workforce).

<sup>22</sup> It is helpful to appreciate that there is a great deal of overlap between the foregoing domains: many practicing oncologists hold professorships at major universities and frequently publish in academic journals, for instance.

<sup>23</sup> Jorge L. Contreras, Constructing the Genome Commons, in *Governing Knowledge Commons* (Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg eds., Oxford University Press 2014), at 103.

figure out what the hell to do with it.”<sup>24</sup> The expert went on to broadly characterize academic researchers as reluctant to share data. “You’d think that the places that have the least interest in sharing are the ones that are for profit or commercial, but actually, it’s just the opposite,” the subject said, adding “It’s the nonprofits – especially academic institutions – where the only incentive is nonfinancial and completely competitive.” As discussed in Section 7.6.2, such disincentives to share data appear to pose a challenge for nascent data-pooling efforts.

Pharmaceutical companies also generate and manage a wealth of oncology data. They do so most often in the course of developing and testing new commercial drugs, diagnostic tools, and related methods. Because this information often has commercial value, pharmaceutical companies typically exert their dominion over it as much as possible by, for instance, restricting access through physical and digital barriers; nondisclosure agreements; asserting trade secret protection; and when possible and advantageous, patenting new methods and molecules.<sup>25</sup>

### 7.2.3 *The Big Data Context*

The recent ascendance of Big Data has introduced new types of workers and institutions to oncology research. The most publicly visible among these are technologists and entrepreneurs, who today are applying a Silicon Valley-style ethos to cancer research. In 2014, for instance, Netflix’s chief product officer presented a public lecture exploring how the same algorithms that recommend films to consumers could one day suggest useful treatments to cancer patients.<sup>26</sup> Marty Tenenbaum, an engineer and entrepreneur who helped bring the Netscape web browser into existence in the 1990s, recently founded “Cancer Commons” – a data-pooling effort with goals similar to those of ASCO’s CancerLinQ. In 2013, Bill Gates invested heavily in a Cambridge-based company that uses DNA sequencing to recommend more effective cancer treatments.<sup>27</sup> Verily, a subsidiary of Alphabet Inc. founded in 2015, is attempting to compose a detailed data-based portrait of the traits of a healthy human.<sup>28</sup>

Less visible but equally significant are many new technology companies helping to generate and mediate access to health data. Today, a variety of smartphone apps help cancer patients chart their treatments and vital statistics, for instance. The San Francisco-based technology company 23andMe is one of several companies to offer genetic screening for certain genetic health risks, and has furnished consumers with

<sup>24</sup> Interview Subject 42.

<sup>25</sup> The FDA publishes a helpful summary of its drug review policies on its website. See *The FDA’s Drug Review Process: Ensuring Drugs Are Safe and Effective*, [www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm](http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm)

<sup>26</sup> Neil Hunt, Big Data Meets Cancer (video), *TedXBeaconStreet*, <https://youtu.be/IWu6XWzbPBs>

<sup>27</sup> Matthew Herper, Bill Gates Invests in Cancer DNA Sequencing Firm, *Forbes* (Jan. 8, 2013), [www.forbes.com/sites/matthewherper/2013/01/08/bill-gates-invests-in-cancer-dna-sequencing-firm/](http://www.forbes.com/sites/matthewherper/2013/01/08/bill-gates-invests-in-cancer-dna-sequencing-firm/)

<sup>28</sup> Verily Inc., [www.verily.com](http://www.verily.com) (discussing “The Baseline Study”).

portions of their “raw” DNA code, which may reveal still undiscovered health information in the future. Individual companies typically govern this consumer-oriented data through the use of form contracts (“end-user-license agreements”) that govern how consumer data may be used and, in some scenarios, shared.<sup>29</sup>

These new services have introduced a new class of professionals to the world of oncology data: data scientists. While this job title is not well defined, data scientists typically hold degrees in computer science, engineering, and mathematics, and they are often hired by companies to help organize, manage, and analyze data. A data scientist hired by ASCO to assist with CancerLinQ explained in an interview for this chapter that his expertise involves removing any personally identifying information from patient treatment records (“anonymizing” or “de-identifying”), cleaning minor errors from data that was recorded incorrectly or incompletely, transcribing data into common formats to aid analysis, grouping data into useful clusters based on its characteristics, and finally, divining useful patterns from statistical noise.

#### 7.2.4 *The Legal and Regulatory Environment*

The sources and stewards of oncology data described in the foregoing paragraphs – health care providers, academic research institutions, pharmaceutical companies, and technology companies – operate in a complex legal and regulatory environment that influences how oncology data can be collectively governed.

At the outset, it is helpful to appreciate that oncology data is often stored in computer databases protected by technological barriers that are difficult to circumvent, such as passwords and encryption mechanisms. A patchwork of federal and state laws prohibits gaining unauthorized access to such databases. The institutions that generate and store oncology data also typically contractually forbid their employees, partners, and customers from disseminating it. Because patent and copyright protection for data is generally quite thin, those two bodies of law do not represent a significant deterrent to its dissemination.<sup>30</sup> Trade secret law may, however, offer some recourse for data-holding firms. Unlike copyright and patent law, which apply to carefully defined categories of subject matter, trade secret protection can apply to many types of information. In simplified terms, the requirements for protection are typically that the information is valuable because it has been the subject of reasonable efforts to keep it secret. In theory, this could include many types of oncology treatment data and related information. (Leading legal

<sup>29</sup> See US experience with doctors and patients sharing clinical notes, *BMJ* 2015; 350:g7785, [www.bmj.com/content/350/bmj.g7785](http://www.bmj.com/content/350/bmj.g7785). Another new project gaining traction in the health care community called OpenNotes provides new ways for doctors to share their notes and related data directly with patients through the Internet.

<sup>30</sup> Copyright protection may sometimes apply to compilations of data or data classifications, but such protection does not represent a significant barrier to unauthorized copying in this context.

commentators have noted, however, the challenges of maintaining secrecy over such information as well as the policy challenges presented by trade secrecy as applied to clinical information.)<sup>31</sup> In addition to longstanding state laws designed to protect trade secrets, the recent passage of the Defend Trade Secrets Act of 2016 provides a federal cause of action for misappropriation of trade secrets.<sup>32</sup> Ultimately though, perhaps the laws most important to oncology data are not those that forbid outsiders from taking it, but rather, those that govern its voluntary disclosure.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) prohibits hospitals from disclosing patient names, zip codes, treatment dates, and other pieces of potentially identifying information.<sup>33</sup> Some state and federal privacy laws provide similar protections to patients by imposing civil liability on those who disclose patient information without permission.<sup>34</sup> As a result, before a health care provider can contribute its data to a commons such as CancerLinQ, it will typically need to remove identifying information prior to disclosure.

In theory, antitrust law could also potentially discourage the formation and operations of oncology data pools. Antitrust authorities have long recognized that certain types of information-sharing arrangements between commercial firms can have anticompetitive effects that violate the Sherman and Clayton Acts.<sup>35</sup> Historically, such exchanges involved patents and other forms of organizational knowledge, but a data-sharing commons focused on commercially valuable health information could raise similar concerns, depending on how it was structured – if it conditioned access to one set of data on the licensing of substitutive data, for instance. While this possibility is interesting to consider, it seems remote in the case of efforts like CancerLinQ. If potential members and partners of a cancer data-pooling group *perceived* antitrust liability as a risk, however, that could present a challenge for group organizers to overcome.

Policymakers have recently taken some important steps to encourage controlled and limited disclosures of privately held health data. The 2010 Patient Protection and Affordable Care Act has helped to encourage the development of “accountable care organizations” (ACOs) across the country.<sup>36</sup> These hospitals and care centers

<sup>31</sup> Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 *Yale L.J.* 177, 194 (1987); J. H. Reichman and Pamela Samuelson, *Intellectual Property Rights in Data?*, 50 *Vand. L. Rev.* 51, 60 (1997).

<sup>32</sup> 18 U.S.C. §1836 et seq.      <sup>33</sup> 45 C.F.R. § 164.514(e) (2013).

<sup>34</sup> An interesting reflection of the stigma that cancer once carried is the 1966 Freedom of Information Act’s specific exemption for “treatment for cancer,” describing disclosure of the fact of treatment as “an unwarranted invasion of personal privacy.” 1966 Freedom of Information Act. 5 U.S.C. § 552 et seq.

<sup>35</sup> As David J. Teece has written, “meetings and exchanges of technical information . . . can cause antitrust suspicion.” Teece, *Information Sharing, Innovation, and Antitrust*, 62 *Antitrust L.J.* 465, 474 (1994). The central concern is that some forms of information sharing may dampen competition for products and services.

<sup>36</sup> A helpful summary and map displaying these organizations is available at [www.oliverwyman.com/content/dam/oliver-wyman/global/en/files/archive/2013/OW\\_HLS\\_ACO\\_maps.pdf](http://www.oliverwyman.com/content/dam/oliver-wyman/global/en/files/archive/2013/OW_HLS_ACO_maps.pdf)

beneficially flip the traditional economics of patient care: private insurance companies or Medicare make higher payments to ACO doctors with strong track records for high-quality patient treatment. The quality of treatment may be measured by, for instance, how often a doctor's patient visits the hospital or an emergency room following treatment.<sup>37</sup> This payment model necessarily requires hospitals to report on the rates of success they have with treating patients. While the accountable care system involves such data being shared between the care provider and an insurance company, several subjects interviewed for this chapter expressed the hope that this shift could make hospitals more willing to share patient treatment data more generally – including with promising projects such as CancerLinQ.

The Department of Health and Human Services (HHS) has also recently advanced a number of projects to facilitate widespread health data sharing. One project, called the Blue Button Initiative, is designed to help patients gain easier access to their personal health records.<sup>38</sup> Beginning in 2014, the Health IT Policy Committee (a public advisory body on health-related information technology organized by the HHS) has regularly held workshops on health data sharing. Experts from government, private industry, and academia have participated in these workshops. The topics they have examined include health data information exchanges, personal privacy, and national security. Oncologists interviewed for this chapter expressed optimism that these recent steps might help get more clinical data available for oncology pooling efforts such as CancerLinQ.

### 7.2.5 *The Patient Context*

In 2014, there were approximately 14.7 million cancer patients in the United States and 1.6 million new patients per year.<sup>39</sup> These people are at the very core of oncology research, and yet they remain largely disconnected from the processes and systems that gather and use their data. As a cancer patient interviewed in the *New York Times* recently commented, “The person with the least access to the data in the system is the patient.”<sup>40</sup> The reasons are partly cultural and partly financial. Traditionally, doctors simply did not regard cancer patients as people who needed to understand their situation in great detail. Not long ago, doctors typically reported bad news not to patients but, instead, to their families. A sick person's role was more or less passive. Only recently has a cultural shift occurred in which patients expect to be active participants in their care. Although that change has occurred, an important economic reality remains: because doctors profit from treating patients, sharing data

<sup>37</sup> Patient Protection and Affordable Care Act § 2706 (March 23, 2010).

<sup>38</sup> The Blue Button Initiative at [www.healthit.gov/patients-families/your-health-data](http://www.healthit.gov/patients-families/your-health-data); see also [www.hhs.gov/news/press/2014pres/02/20140203a.html](http://www.hhs.gov/news/press/2014pres/02/20140203a.html)

<sup>39</sup> National Cancer Institution webpage, at [seer.cancer.gov/statfacts/html/all.html](http://seer.cancer.gov/statfacts/html/all.html) (last accessed June 6, 2017).

<sup>40</sup> Steve Lohr, The Healing Power of Your Own Medical Records, *New York Times* (Mar. 31, 2015).

with a patient makes it easier for that patient to go elsewhere. Subjects interviewed for this chapter explained that, as a result, medical culture is highly resistant to information sharing and the burden to obtain information (data or otherwise) remains squarely on patients. This may change, however, under plans such as the Blue Button initiative, and the accountable care reimbursement payment system, under which some hospitals and treatment centers will be unable to function without sharing treatment data more broadly.

### 7.3 GOALS, OBJECTIVES, AND HISTORY

Subjects interviewed for this chapter traced the genesis of CancerLinQ to a 2010 report published by Institute of Medicine (IOM) that explores the idea of a “Rapid Learning System for Cancer Care.”<sup>41</sup> Drawing heavily upon the published work of leading oncologists, the report defines such a system as follows:<sup>42</sup>

A rapid learning healthcare system (RLHS) is one that uses advances in information technology to continually and automatically collect and compile from clinical practice, disease registries, clinical trials, and other sources of information, the evidence needed to deliver the best, most up-to-date care that is personalized for each patient. The evidence is made available as rapidly as possible to the users of a RLHS, which include patients, physicians, academic institutions, hospitals, insurers, and public health agencies. A RLHS ensures that this data-rich system learns routinely and iteratively by analyzing captured data, generating evidence, and implementing new insights into subsequent care.<sup>43</sup>

The report discusses how an RLHS might work in practice, including how data could be collected from various sources, such as hospitals, and meaningfully analyzed.<sup>44</sup> A former president of ASCO explained that the IOM report was a watershed moment for CancerLinQ:

The IOM report was a clarion call that [ASCO] really should be looking at this because it was a definable public good to bring together data from a broad array of sources and use them to both help instruct the care of the individual patients and to use that data collectively as a way to improve the overall health care of the nation. So it was clearly partially the fact that technically it was no longer possible to do, but also just because we thought it was the right thing to do for patients – individually and collectively.<sup>45</sup>

<sup>41</sup> IOM (Institute of Medicine), *A Foundation for Evidence-Driven Practice: A Rapid Learning System for Cancer Care: Workshop Summary: Workshop Summary* (National Academies Press 2010).

<sup>42</sup> See Amy P. Abernethy et al., Rapid-Learning System for Cancer Care, *J. Clinical Oncology* 28(27): 4268 (Sept. 2010).

<sup>43</sup> Abernethy et al., Rapid-Learning System for Cancer Care.

<sup>44</sup> Paul Howard, Can We Build a Kickstarter for Cancer? *Forbes* (Aug. 1, 2013), [www.forbes.com/sites/theapothecary/2013/08/01/can-we-build-a-kickstarter-for-cancer/](http://www.forbes.com/sites/theapothecary/2013/08/01/can-we-build-a-kickstarter-for-cancer/)

<sup>45</sup> George Sledge, Stanford.

The interview subject added that around the same time the report was published, experts at ASCO were growing aware that such a system was feasible. “We looked at [the report] and realized,” he recalled, “that . . . there may well be capacity in the system to link together a lot of these databases from a technical standpoint in a way that we would not have been able to do previously.”<sup>46</sup>

CancerLinQ has two goals: first, to mediate among sources of oncology treatment data – primarily hospitals and cancer treatment centers – and second, to act as a central location for the storage, analysis, and distribution of shared informational resources – that is, treatment guidance and quality metrics. As the chief medical officer at ASCO explained, “The overarching goal for CancerLinQ is to allow doctors to learn from every clinical counsel with every cancer patient, so that they can be better informed for the management of all cancer patients.”<sup>47</sup> When asked what ASCO would need to do to reach this goal, the subject focused more on technological tasks than institutional challenges:

Setting up the health IT platform that enables us to actually capture information from the clinical care of every cancer patient, and then using that much larger experiential database to try to develop insights, assess trends, make inferences, hypotheses, that can then be pursued or could even result in immediate changes in clinical care, depending upon how robust the observation is.<sup>48</sup>

One subject said he believed that ASCO is well suited to coordinate the activities of an oncology data pool because it is less influenced by factors that could lead corporate institutions to behave opportunistically. “Pharmaceutical companies are very interested in having information about how people are using their drugs,” he stated, adding that,

insurers of course are very interested in having their hands on all of the information because it allows them to know rates to charge . . . There is nothing wrong with those things, but they are not things that are necessarily going to improve patient care, so from a public good standpoint we viewed this as something important to have an honest broker, and we viewed ASCO as an honest broker.<sup>49</sup>

CancerLinQ has developed its technological infrastructure swiftly. In 2012, ASCO began work on a prototype designed to demonstrate the technological feasibility of aggregating treatment data from multiple care centers.<sup>50</sup> The prototype was completed soon after, and ASCO demonstrated the system publicly on March 27, 2013, at the National Press Club.<sup>51</sup> The demonstration was a proof-of-concept in which an

<sup>46</sup> George Sledge, Stanford.    <sup>47</sup> Richard Shilsky, ASCO.    <sup>48</sup> Richard Shilsky, ASCO.

<sup>49</sup> George Sledge, Stanford.

<sup>50</sup> <http://cancerlinq.org/about> (“2012: Prototype development begins /Aims to demonstrate the feasibility of CancerLinQ and obtain insights for the full system”).

<sup>51</sup> Press Release, ASCO, ASCO Completes Prototype for CancerLinQ, Marking First Demonstration of a “Learning Health System” to Transform Cancer Care, [www.asco.org/about-asco/press-center/news-releases/asco-completes-prototype-cancerlinq%E2%84%A2-marking-first](http://www.asco.org/about-asco/press-center/news-releases/asco-completes-prototype-cancerlinq%E2%84%A2-marking-first) (Mar. 31, 2013).

ASCO director searched and analyzed a set of records of more than 100,000 breast cancer patients originating from four cancer centers – Maine Center for Cancer Medicine, Marin Specialty Care and Marin General Hospital (California), Space Coast Cancer Center (Florida), and Tennessee Oncology – before an audience.<sup>52</sup> Following this demonstration, the project garnered significant attention in the national press, including reports in the *Wall Street Journal* and the *LA Times*, and accolades in a White House press release on the important role that Big Data can play in solving national problems.<sup>53</sup>

As of this writing, CancerLinQ is swiftly progressing. In early 2015, ASCO's president, Peter P. Yu, announced that CancerLinQ would go into operation later in the year with the support of eight community oncology practices and, possibly, seven large cancer centers. Yu stated that, thanks to these commitments, CancerLinQ will house 500,000 patient records at launch.<sup>54</sup> In a May 2015 press release, ASCO announced that the number of committed member institutions had expanded to fifteen.<sup>55</sup> According to another 2015 press report, ASCO is investing heavily in the project's future, having allocated funds in its budget in the "eight-figure" range over the next five years.<sup>56</sup> As of this writing, the CancerLinQ website indicates that the project is operational and overseen by CancerLinQ LLC, a subsidiary of ASCO.<sup>57</sup>

In addition to reporting on how CancerLinQ has developed and what its goals are, this brief history revealed the central members of the CancerLinQ community – a key element of the knowledge commons analytical framework. The key actors are ASCO personnel, oncologists at hospitals and universities who sit on the project's

<sup>52</sup> The video can be viewed at [www.asco.org/quality-guidelines/transforming-cancer-care-through-big-data-asco-unveils-cancerlinq-prototype](http://www.asco.org/quality-guidelines/transforming-cancer-care-through-big-data-asco-unveils-cancerlinq-prototype); Jo Cavallo, Building CancerLinQ: The Road to Faster, More Efficient Treatment Delivery, *The ASCP Post* (Mar. 15, 2013), [www.ascopost.com/issues/march-15-2013/building-cancerlinq-the-road-to-faster-more-efficient-treatment-delivery.aspx](http://www.ascopost.com/issues/march-15-2013/building-cancerlinq-the-road-to-faster-more-efficient-treatment-delivery.aspx)

<sup>53</sup> Ron Winslow, Big Data for Cancer Care – Vast Storehouse of Patient Records Will Let Doctors Search for Effective Treatment, *Wall St. J.* (Mar. 27, 2013); Melissa Healy, Cancer trials aim to shore up 'precision medicine's' base of evidence, *L.A. Times* (June 1, 2015), [www.latimes.com/science/science-now/la-sci-sn-trials-precision-medicine-20150601-story.html](http://www.latimes.com/science/science-now/la-sci-sn-trials-precision-medicine-20150601-story.html); Press Release, "Data to Knowledge to Action" Event Highlights Innovative Collaborations to Benefit Americans, *White House Office of Science and Technology Policy* (Nov. 12, 2013), [www.whitehouse.gov/sites/default/files/microsites/ostp/Data2Action%20Press%20Release.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/Data2Action%20Press%20Release.pdf)

<sup>54</sup> Tony Berberabe, CancerLinQ to be Rolled Out to 15 Community Practices, Large Cancer Centers, *Onclive* (Jan. 21, 2015), [www.onclive.com/web-exclusives/CancerLinQ-To-Be-Rolled-Out-to-15-Community-Practices-Large-Cancer-Centers](http://www.onclive.com/web-exclusives/CancerLinQ-To-Be-Rolled-Out-to-15-Community-Practices-Large-Cancer-Centers). Centers include Inova Comprehensive Cancer & Research Institute (Virginia), South Coast Centers for Cancer Care (Massachusetts), New England Cancer Specialists (Maine), Medical Oncology Hematology Consultants (Delaware), Cancer Treatment Centers of America (Pennsylvania), Marin Cancer Care (California), Space Coast Cancer Center (Florida), and Michiana Hematology-Oncology (Indiana).

<sup>55</sup> CancerLinQ™ Takes a Big Leap Forward (May 29, 2015), <https://am.asco.org/cancerlinq-takes-big-leap-forward>

<sup>56</sup> Nick Mulcahy, ASCO: "High-Risk" Project "Must Be Done": Tech Platform Is Best Known in Finance, *Medscape* (Jan. 22, 2015), [www.medscape.com/viewarticle/838543](http://www.medscape.com/viewarticle/838543)

<sup>57</sup> <http://cancerlinq.org> (last visited June 20, 2017).

various advisory boards, and the practices that have committed to donate their data to the project. The group's decision not to solicit data contributions from other potential sources mentioned in the original IOM report, such as pharmaceutical companies, academic researchers, or individual patients, may reflect how relatively closed off those domains are to health data transactions as compared to the clinical environment. This is perhaps the broadest way that the environment in which the project formed appears to have shaped its goals.<sup>58</sup>

#### 7.4 ATTRIBUTES: THE CHARACTERISTICS OF ONCOLOGY DATA

The primary resource relevant to CancerLinQ is oncology treatment data from doctors, hospitals, and other care providers. An interview subject involved with the project helpfully divided this resource into two types: structured data and unstructured data. Structured patient treatment data includes objective, machine-recorded information such as “laboratory test results or the dosages of medicines prescribed, or patient vital signs,” he explained, as well as medical history data, laboratory data, and imaging data from radiological studies, whereas unstructured data is generated and recorded more casually and based upon more subjective observations – a clinical physician's handwritten notes, for instance.<sup>59</sup>

Structured data is typically incorporated into a patient's electronic health record, which, at most hospitals, is digitally stored and managed by an outside vendor.<sup>60</sup> Unstructured data, meanwhile, is sometimes stored locally within a hospital or may be appended to a patient's EHR as a digital image. This data is often replicated in other places as well, such as a hospital's billing department or an insurance provider's customer database. “If the doctor bills for their services, all that sort of information gets converted, if you will, into a lot of different kinds of codes that are used to submit the claim to insurance,” one subject explained. Another subject explained that one of the most important types of data in this area, patient mortality information, can be found in the social security death index, as well as newspaper obituaries, and sometimes even copies of handwritten condolence notes stored in patient files. In summary, oncology patient treatment data is scattered, but most of it resides in a patient's EHR maintained by a company on contract with the patient's care provider.

<sup>58</sup> There are other data-pooling efforts that seek contributions of data from individual patients, however. The largest of these is Cancer Commons.

<sup>59</sup> Dr. Richard Shilsky, ASCO.

<sup>60</sup> According to a recent federal study conducted by the Office of the National Coordinator for Health Information Technology, about three out of four non-federal acute care hospitals have a basic EHR system. Office of the National Coordinator for Health Information Technology, *Adoption of Electronic Health Record Systems among US NonFederal Acute Care Hospitals: 2008–2014, ONC Data Brief No. 23* (Apr. 2015), [www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf](http://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf)

CancerLinQ intends to use the treatment data it collects to generate a secondary asset: comparative assessments of the quality that member hospitals and doctors provide. As explained earlier (Section 7.2.4), such “quality metrics” can play an important role in how some hospitals and insurance companies reimburse doctors. As a result, hospitals may find CancerLinQ’s metrics valuable, both to track their quality internally and for reporting purposes. The next section contains more information about how the project intends to use these metrics as an incentive to potential data contributors.

## 7.5 GOVERNANCE

### 7.5.1 *Formal Governance Structure*

CancerLinQ is governed by ASCO directors and by a collection of advisory boards populated by oncologists; academic researchers; and, perhaps surprisingly, employees of several large pharmaceutical companies.<sup>61</sup> A cursory look at these boards and their membership reveals the project’s far-reaching constituency. The Board of Governors (the core leadership of CancerLinQ) includes ASCO’s president and the president of a major cancer research center; a Business Strategy Committee includes employees of GlaxoSmithKline and Novartis; a committee designed to offer advice on interactions with physicians includes physicians who work at several cancer care centers around the country; another committee geared toward technology includes a professor from Harvard Medical School and a doctor employed at Merck Research Laboratories. More committees advising on regulatory compliance, patient outcomes, and more subjects include professors from the University of Pennsylvania, the University of Michigan, Duke, and leading cancer centers. ASCO’s chief medical officers briefly listed several key challenges these boards expect to grapple with in the months leading up to CancerLinQ’s official launch: “data quality, privacy, and security.”<sup>62</sup>

### 7.5.2 *Incentives*

A central part of ASCO’s vision for CancerLinQ is delivering useful information back to the members who contribute data. As explained throughout this chapter, this will include information that can be used to make treatment recommendations. A second enticement to join will be reports or “metrics” describing the quality of a contributor’s medical services. Especially in light of the growth of the accountable

<sup>61</sup> Office of the National Coordinator for Health Information Technology, *Adoption of Electronic Health Record Systems* (“Leading thinkers from an array of relevant fields engaged to advise on design and implementation. Advisors include leading oncologists, patient advocates, privacy advocates, health outcomes researchers, ethicists, health IT experts, and many others.”)

<sup>62</sup> Richard Shilsky, ASCO.

care business model, such data may hold great value. As ASCO's chief medical officer explained, "We think that one of the big incentives is going to be that . . . in order to optimize their reimbursement [doctors] are going to have to be able to demonstrate that they provide quality care and that they continuously improve their quality. And so one of the major focuses of CancerLinQ will be to develop and provide useful metrics on quality of care to clinicians and to federal government and other agencies."<sup>63</sup> In this sense, the subject explained, CancerLinQ acts as an information exchange between its members: "We will be able to return to the physician on a regular basis a dashboard report that shows what is the quality of their performance against . . . standard measures and [how] they can use that information to report to . . . their private insurers what their quality is, how it compares to other physicians."<sup>64</sup> Another subject explained that such data "could be used by the insurers or the government as a way of judging quality as we migrate . . . to payments based upon outcomes and quality. You know, having these sort of process measures in place will be very important."<sup>65</sup> The president of a member cancer care center said he believed this could be a powerful incentive to join the effort.<sup>66</sup>

### 7.5.3 Openness

Interview subjects reported that CancerLinQ has thoughtfully addressed difficult questions of openness and access from its earliest days. A former leader at ASCO who co-chaired a committee on the project provided examples of such questions:

Governance is important at multiple levels. At one level, if you are getting data from an institution, or from an individual doctor, or from a group practice somewhere, who has access to that data? Is access given just in access to CancerLinQ staff who are aggregating and putting the data together? Do physicians have access to other physicians' data? How does ASCO give out access to health-services researchers who are interested in looking at important questions? At the end of the day, if ASCO has data that would be of interest to an insurer or a pharmaceutical company, what level of data do you allow access to? . . . So I actually view the governance issues as the most important issues surrounding CancerLinQ.<sup>67</sup>

Such concerns prompted the organizers of CancerLinQ to set up the Data Governance Committee very early on in the project's history. "It was one of the first things we did when we set up CancerLinQ," he stated. "And CancerLinQ actually has a full-time staff member who does nothing but Governance."<sup>68</sup> The committee is composed of oncologists at leading hospitals and universities around the country.<sup>69</sup>

At the time of this writing, ASCO has elected not to publicize its policies and rules governing who may join CancerLinQ as a contributor of data and who may join as a

<sup>63</sup> Richard Shilsky, ASCO. <sup>64</sup> Richard Shilsky, ASCO. <sup>65</sup> George Sledge, Stanford.

<sup>66</sup> Interview Subject 21 (commented on condition of anonymity). <sup>67</sup> George Sledge, Stanford.

<sup>68</sup> George Sledge, Stanford. <sup>69</sup> [www.cancerlinq.org](http://www.cancerlinq.org).

user of data. As a result, one can only infer these policies from ASCO's public statements and from the comments of others in the oncology community: CancerLinQ project leaders have consistently stated that their ultimate goal is to include the data from "all cancer patients in the United States" in its pool, for instance. These statements support an inference that the group is open to any and all cancer care providers who wish to donate useful data.

The more nebulous question at this time is who may join CancerLinQ purely as a *data user* – that is, a user who gains access to the anonymous oncology data contributed by member institutions. According to materials ASCO has published on the CancerLinQ website, the project's organizers anticipate providing access to outside researchers, but there are no specifics on the steps that such a researcher would need to take once CancerLinQ launches. A researcher (and CEO of a patient-centered oncology data pool) stated that he had requested access to the data that the project's organizers had gathered for its 2013 proof-of-concept to no avail. Commenting generally, he added, "[Many organizations] that you may think will share data are really silos."<sup>70</sup> Because CancerLinQ is still in early days, however, it seems appropriate to withhold any judgment with respect to this issue.

Perhaps the foregoing discussion of openness is most useful as a provocation – an impetus to speculate about the position ASCO may soon be in as a data pool administrator. There are, it seems, prudent reasons for ASCO to be reluctant to share patient treatment data very broadly. For one thing, it is becoming increasingly possible to re-identify patient records that have been made anonymous.<sup>71</sup> It could be problematic legally and professionally if the identities of patients represented in the pool of data were ever uncovered. (HIPAA, discussed in Section 7.6, represents one such risk.) Then again, tools are available – both legal and technological – that would permit data pools such as CancerLinQ to provide access while guarding against such risks. End user license agreements could forbid researchers from attempting to de-identify health data, for instance. Likewise, the data pool could decide to share only highly generalized information (i.e., trends, analytics, and summaries rather than individual records) with certain researchers in positions of lower trust.

## 7.6 CHALLENGES

### 7.6.1 *Legal and Regulatory Challenges*

In light of CancerLinQ's early stage, this chapter's primary goal is not to assess or evaluate it, but rather, to identify potential challenges that this group and other

<sup>70</sup> Marty Tenenbaum, Cancer Commons.

<sup>71</sup> See, e.g., Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 *UCLA L. Rev.* 1701 (2010).

efforts like it may face. While several areas of law and regulation are relevant to CancerLinQ, the one that appears to present the greatest costs and risks is HIPAA. As explained earlier in this chapter, this law requires institutions to remove personally identifying information from health records before disclosing them. A patchwork of state and federal privacy laws impose similar requirements on health care institutions that seek to share data as well.

While it would be a simple matter to redact patient treatment records of names, dates of treatment, home addresses, and other identifiers, removing this information also removes a wealth of the useful underlying data – the period of time over which a patient was treated, or state of residence for instance. To remedy this problem at the proof-of-concept stage, ASCO employed a data scientist to creatively modify patient data without removing its underlying utility. Commenting for this chapter, the data scientist explained that these steps involved shifting treatment dates by a consistent offset (e.g., 27 days), replacing a zip code with the zip code of an adjacent community, or altering a patient’s birthdate by a year or two. Importantly, patient names are frequently replaced with numerical identifiers, permitting the same patient to be recognized within the system and examined over the entire treatment period. Manipulating data in this way is costly and difficult to automate because it often requires the skills and judgment of a human data expert. (Mattioli 2014). It remains unclear who will bear these costs when CancerLinQ is no longer in the prototype stage.

### 7.6.2 Cultural Challenges

Although doctors and hospitals may be more likely than pharmaceutical companies or academic researchers to part with useful data, the medical culture remains largely antithetical to data sharing. As an editorial in *Genome Magazine* recently explained, there are “longstanding institutional barriers to gathering patient data, ranging from patient privacy to competition for patients to the idea the data has proprietary value.”<sup>72</sup> Subjects interviewed for this chapter painted a consistent picture: “We have approached many, many medical institutions, large cancer centers, especially the big ones,” explained the CEO of one data-pooling project, “They are very, very protective of their data. Because they think they are big enough to be able to not need anyone else’s data, so they will not share their data. And they will argue strongly that competing strongly is the best way to move science forward. It is strongly in the culture.”<sup>73</sup> Another subject summarized the culture of medical data in one word: “dysfunctional.”

<sup>72</sup> Eric Celeste, Collective Action, *Genome Magazine* (Aug. 24, 2014) available at <http://genome.mag.com/collective-action/#.WTcIwxPytE4> (last visited June 6, 2017).

<sup>73</sup> Marty Tenenbaum, CEO Cancer Commons.

### 7.6.3 Commercial Challenges

Alongside the institutional reluctance woven in the culture of medical cancer care are challenges related to EHR vendors who act as stewards over most of the data in which CancerLinQ is interested.<sup>74</sup> “You need to have an EHR vendor or an institution that is willing to share the data,” stated an ASCO director, adding that they are “notoriously proprietary” and poor at sharing useful data even within their own organizations. According to interviewees, this practice is strategic: it benefits EHR vendors to keep data in specialized formats that do not integrate well with other systems because this tends to discourage hospitals from moving to competing vendors. This problem is aggravated, the subject explained, by the fact that hospitals have few choices when selecting EHR vendors to work with. “A fairly small number of corporations that are responsible for the electronic health records in the United States,” he said, adding, “none of them will speak to each other.” Other subjects made consistent comments about the difficulty of obtaining data from EHR vendors and some explained that even when data is obtainable, it may not be immediately usable because it is in a proprietary digital format.

The president of a cancer treatment institution that is a CancerLinQ member explained that the difficulty of working with large EHRs and hospitals is why CancerLinQ has pursued smaller cancer care centers: ASCO “looked at private practices to start this project rather than academic institutions [because] the data’s probably more easily extractable from the private practice EHRs and trying to get discrete information out of a big hospital system can be very tedious.”<sup>75</sup> Other subjects interviewed consistently reported that data held by smaller private practices is typically subject to fewer institutional barriers.

## CONCLUSION

The analysis of patient treatment data culled from hospitals, practices, and other private institutions across the country could profoundly improve how successfully cancer is treated. This possibility is, in part, a product of new developments in technology – the mass digitization of health information, advances in digital storage, and the new methods of organizing and analyzing information that go by the name Big Data. It is puzzling, however, that most accounts of this phenomenon focus on technology when the more significant story to be told is about institutions.

This chapter has focused on the early years of a prominent data pool that is still taking form. As such, it has centered on early-stage issues, such as gathering buy-in from potential contributors in a challenging regulatory and competitive environment. If one

<sup>74</sup> Some of the information reported in this discussion also appears in a law review article that serves, in part, as a counterpart to this chapter. There, the information is placed in a broader context with respect to innovation policy. Michael Mattioli, *The Data-Pooling Problem*, 32 *Berkeley Tech. L. J.* (2017).

<sup>75</sup> Interview Subject 21 (commented on condition of anonymity).

were to study CancerLinQ longitudinally, other sets of important institutional issues would likely arise. These could include resolving conflicts between members; establishing a framework for collective decision making; rules governing ownership of, and royalties generated by, intellectual property developed from the data pool; and so forth. Other issues and challenges still unknown rest on the horizon.

For now, it is helpful to consider today's challenges. This chapter has shown that medical data is not an abstract resource that can simply be gathered and mined; rather, it is the product of human effort and insight, scattered and stewarded by different organizations, each of which is motivated by different incentives, beholden to different interests, and averse to varying risks. Convincing these organizations to aggregate their data and to collectively govern access to it despite legal, competitive, and cultural pressures not to: this is the true challenge that CancerLinQ and projects like it face. There is a basis for optimism: if more care centers move to the accountable care payment model, the performance information that CancerLinQ will share with its members could offer a powerful incentive to contribute data. In addition, efforts such as the Blue Button Initiative could help create a cultural shift within the world of medicine, making care providers more comfortable with sharing properly de-identified patient data. Perhaps most importantly (although hardest to quantify) is the sense of forward inertia and excitement surrounding the idea of pooling medical data. There appears to be a widespread consensus among oncologists that access to a vast pool of patient treatment data would spur meaningful improvements in patient care. Perhaps this is a reflection of the fact that most doctors understand that medical advances have often emerged not from traditional scientific models of understanding, but from the statistical wisdom found in correlations. If enough care providers can be convinced to join together, the kingdom of cancer may one day become a true commons – a new country defined by data, governed collectively, and most importantly, a place from which more visitors return.