ENERGIZING HEALTH

COLLABORATION SERIES

Health Collaboration Playbook

March 2014

CHECKLIST FOR GETTING A COLLABORATION STARTED

1. Choosing How to Organize (p. 5–11)
   - ☐ What is the goal of the organization? (p. 5–6)
     - ☐ Complete a research project or major milestone? (p. 6)
     - ☐ Create an enterprise to support a customer (p. 6)
   - ☐ How can we structure the organization? (p. 7–8)
     - ☐ For-profit (p. 8–11)
       - ☐ Are you making a product? (p. 8–9)
       - ☐ Are you providing a service? (p. 8–9)
     - ☐ Not-for-profit (p. 9–11)
       - ☐ Are you mission oriented (501(c)(3) (p. 9–11)
       - ☐ Are you member oriented (501(c)(6) (p. 9–11)
     - ☐ Collaborative contractual (p. 11)
       - ☐ Who is drafting the agreements? (p. 11)
2. Stance on Intellectual Property (p. 12–19)
   - ☐ Conduct an intellectual property inventory (p. 12–14)
     - ☐ Do participants agree to (p. 13–14)
       - ☐ Share prior intellectual property in order to develop new knowledge? (p. 13–14)
       - ☐ Share information generated in the collaboration in order to develop new products, services, or research? (p. 14)
   - ☐ Nature of Product (p. 14–15)
     - ☐ Is the product digital (software, creative work) or nondigital (device, pharmaceutical, practice change)? (p. 14–15)
     - ☐ Is there a fixed cost per transaction (skill or labor intensive) or a marginal cost per manufactured unit? (p. 15)
   - ☐ Industry Context (p. 15–17)
     - ☐ Are you creating something novel or improving an existing product or practice? (p. 15–16)
     - ☐ What is the regulatory burden? (p. 15–16)
     - ☐ Who is the customer? (p. 16)
   - ☐ Innovation Type (p. 17–18)
     - ☐ User-Driven Innovation: Consumers of the product are the empowered innovators (e.g., the Internet) (p. 17)
     - ☐ Distributed Innovation: Participants in an innovation process are the empowered innovators (e.g., Wikipedia) (p. 18)
     - ☐ Open Innovation: Members of participant organizations are coequal innovators (e.g., pre-competitive spaces) (p. 18)
   - ☐ Impact of Asserting (p. 18–19)
     - ☐ Do you have clearly defined patentable objects? (p. 18)
     - ☐ Do you have a mix of intellectual property and "knowledge" about how to use the property that isn't amenable to transfer? (p. 19)
3. Role of Technology, Innovation, and Data Practices (p. 20–35)
   - ☐ Using data to build a learning health care system (p. 20–22)
     - ☐ Generating a hypothesis (p. 21)

INTRODUCTION

Health care is slowly being drawn into a data-driven world. Calls for "learning health care" and "accountable care" are grounded in the idea that enriched data streams about individuals will lead to better, and potentially more value-driven, health care. A robust pipeline of new devices is leading to new streams of data about fitness and eating habits. Currently, smartphones and older devices developed for nonhealth uses are being repurposed to generate health data, even to the point of tracking typing errors over time from keyboards to infer the onset of dementia. Eventually, even the routine clinical encounter will be enhanced by medical management driven by a stream of applications and sensors that feed information to medical teams before consumers show up for an appointment.

Increasingly, our health status can be inferred from these kinds of data. And when we connect these data to more "traditional" health data, such as our medical records or genomes, we can create a composite of an individual's health as a state of being, not simply a set of episodes marked by visits to the clinic. Diseases like cancer, diabetes, obesity, arthritis, and autism are unlikely to be driven by single gene variations or environmental causes alone, but more likely arise due to environmental influences on genetically susceptible or predisposed people. Data about the individuals who suffer those diseases hold enormous promise to advance our understanding and ability to intervene effectively with treatments or prevention.

Typically, individuals have a weak relationship to the data that they generate. But our health is more personal, and therefore we need a new model for health data, one that comprehends the fundamental character of inquiry and research that we wish to accelerate. In the absence of a strong social contract regarding our health data, institutions have been highly conservative and have shut out entrepreneurs, research investigators, and in some cases, the patients who "generated" the data.

However, these institutions can change, through either internal leadership or external pressure. Patients increasingly have not just the ability, but the governmentally assured right, to their own medical records and data. Additionally, entrepreneurs and researchers can work directly with patients and participants. Patients also have the capacity to generate data on their own, via third-party genome sequencing services and clinical laboratories, and to make those data available to researchers directly via the Web.

This playbook serves as an outline for how entrepreneurs, researchers, patients, and institutions can collaborate and use data to drive research, create knowledge, and inform clinical care. Specifically, we seek to outline:

1. What are the organizational structures that facilitate collaboration and stakeholder goals?
2. What considerations and assertions should be made about intellectual property?
3. How can technology and standards benefit or hinder an organization's goals?

4. What is the pathway to sustainability?

We approach this playbook from the perspective of practitioners who have worked with industry, academia, not-for-profits, patient advocates, and government funding agencies, and seek to assist stakeholders across the space by identifying roadblocks, hurdles, misperceptions, and catalysts in order to better share knowledge. Ideally this playbook will be a learning document, and as it facilitates new cooperatives and collaborations we hope to use the experience of consumers to inform later iterations.

THE CHECKLIST

Through the course of assembling this playbook, conducting literature reviews, discussing past experiences with entrepreneurs, academics, and other stakeholders in the field, and through feedback it became clear that an organizing series of first principles—a road map—was needed. Many new and interesting ideas have failed to launch, thrive, or continue to operate because the array of groups had competing interests even within the context of a shared goal. We have provided this checklist, found at the beginning of the playbook, as an organizing framework for any organization. While not a strict table of contents, the checklist should be used to serve as a springboard for discussion and negotiation, with the playbook serving as a common reference guide, useful for establishing a shared language and joint understanding. We hope it serves the broader community well as we undertake this complex task.

CHOOSING HOW TO ORGANIZE

☐ What is the goal of the organization? (p. 5–6)
    ☐ Complete a research project or major milestone? (p. 6)
    ☐ Create an enterprise to support a customer (p. 6)
☐ How can we structure the organization? (p. 7–11)
    ☐ For-profit (p. 8–9)
        ☐ Are you making a product? (p. 8–9)
        ☐ Are you providing a service? (p. 8–9)
    ☐ Not-for-profit (p. 9–11)
        ☐ Are you mission oriented (501(c)(3) (p. 9–11)
        ☐ Are you member oriented (501(c)(6) (p. 9–11)
    ☐ Collaborative contractual (p. 11)
        ☐ Who is drafting the agreements? (p. 11)

Organizational Goal

The most important decision that the collaborators must make at the beginning of the process is to clearly define the desired outcome and ensure that all stakeholders and participants agree on that outcome. There are major differences between a foundation that seeks to provide a permanent source of funding for scientific research and patient care and a corporation that looks to develop a new biomarker, device, software program, or therapeutic. Common interest in a disease, a desire to reform an organization, or a will to improve health in a local community might be initial rallying cries among a group of collaborators with incredibly diverse goals. Over the life of a project, lack of concordance will manifest again and again, often coming to the fore through disagreements about intellectual property, technology, or whether an organization is listening to and respecting patients and research participants.

Lifespan and scope

Limited to specific aims
• Lasts until major milestones are accomplished

Is the project an enterprise?
• Created to support "customers" whether academic, consumer, professional, or others
• Serves until it is no longer needed in the market

Goals and Objectives

Generate new knowledge

Improve quality of life for people with a specific condition

Provide social and structural support to a community

Organizational Choice

Achieve early consensus

Make an informed, legal choice
• For-Profit ("C"/"B" corporation or LLC)
• Non-Profit (C3 or C6)
• Contractual consortium (Based in MOUs or CRADAs)

Figure 1. Guide to thoughtfulness about organizational structure

The first question to address is whether an organization or collaboration is looking to be limited to specific aims, such as researching a specific, tractable question or improving outcomes at a local institution, or whether the organization or collaboration is looking to create a sustainable enterprise. In the former case, the goals of the organization are short term and bounded. Projects defined by specific aims can operate through project plans, milestones, Gantt charts, and the like. In the latter case, the organization must think about its path to sustainability and go through the exercise of defining customers, whether those customers are academic, consumers, professionals, or others. Agreement on these issues allows groups to decide on organizational structures that most appropriately suit a project.

Second, groups should agree on the outcome that participants will be promised. Very few health research projects are likely to recover initial costs in the short term, so building an organization starts with understanding this and agreeing on a long-term vision. For some, generation of new knowledge is enough, and science can be advanced even if all that is accomplished is removing a hypothesis from broader consideration. From the patient or research participant perspective, people may wish to engage in order to build a shared experience and improve the lives of their peers. Most organizations will need to demonstrate a tangible, measured improvement to quality of life for those with a chronic condition, whether to market a product, raise donations, or attract further rounds of grants.

Third, collaborators need to decide on the structure of their organization and align that structure with their project duration and agreed-upon outcomes. This choice of legal framework is critical to the character of an organization, how it is perceived by the external world, and the mechanisms it will use to achieve short-term and long-term goals. Additionally, the choice of organizations structure informs how and what an organization owns (i.e., how it can assert and transfer intellectual property rights.)

Choices of Structure

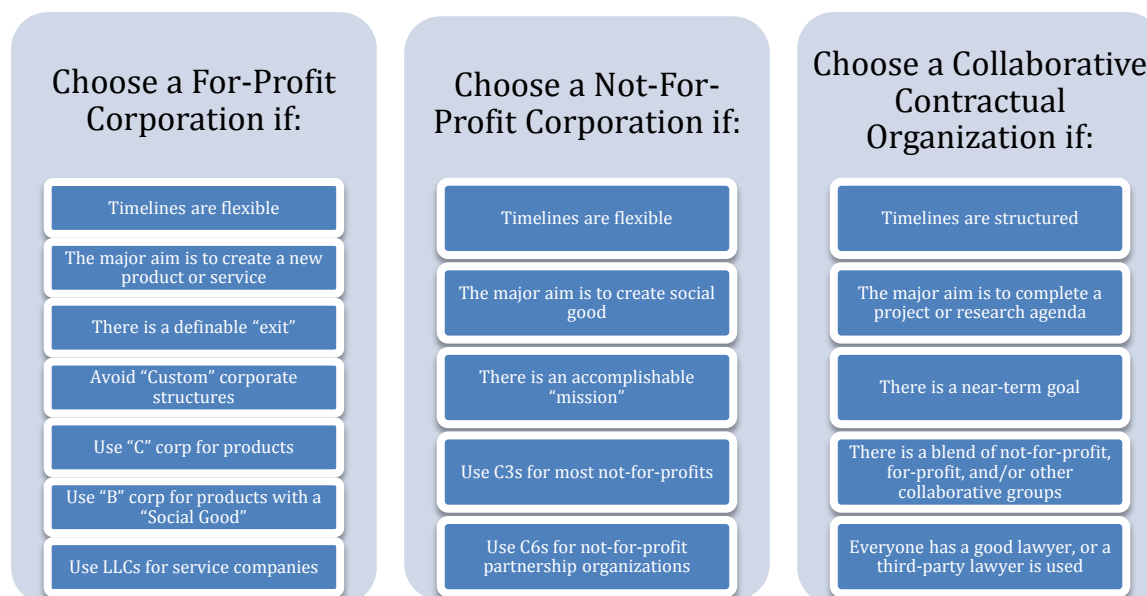| Choose a For-Profit Corporation if:	Choose a Not-For-Profit Corporation if:	Choose a Collaborative Contractual Organization if:
Timelines are flexible	Timelines are flexible	Timelines are structured
The major aim is to create a new product or service	The major aim is to create social good	The major aim is to complete a project or research agenda
There is a definable "exit"	There is an accomplishable "mission"	There is a near-term goal
Avoid "Custom" corporate structures	Use C3s for most not-for-profits	There is a blend of not-for-profit, for-profit, and/or other collaborative groups
Use "C" corp for products	Use C6s for not-for-profit partnership organizations	Everyone has a good lawyer, or a third-party lawyer is used
Use "B" corp for products with a "Social Good"		
Use LLCs for service companies		

Figure 2. Quick reference guide to organizational structures

There are three general classes of organizational structures from which to choose: a for-profit entity, a not-for-profit entity, and a collaborative contractual "virtual" entity. Each has its upsides and downsides, and should be considered against the two outcomes noted in the introduction for any given project: what kind of organization do you want to have (and for how long), and what kind of outcomes do you wish to return to participants?

A for-profit corporation is simple: it is intended to return a profit for its owners, who invest the funds that drive operations. These profits can take the form of dividends, stock sales, sale of the corporation, and more. It may or may not need a board of directors and often requires formal incorporation. There are many shapes and sizes of for-profit corporations, and it's not the goal of this playbook to delineate them in full. Instead, we will examine some of the most popular kinds of for-profit corporations (FPCs hereafter) and how some may or may not map to a data-driven health collaboration.

A not-for-profit organization (NPO hereafter) is an organization intended to achieve a social goal, not a financial one. The investors are not owners of the NPO, and should the NPO gain an excess of funds through grants, services, or other means, it must use those funds to further the social mission rather than return them as profit. In order to receive tax-exempt status, NPOs must have boards and formally incorporate.

A collaborative contractual organization can exist as a virtual project. These efforts may not have legal existence as a stand-alone entity, but instead are collaborations among existing parties formed by a set of interlocking contractual documents. The parties can

be FPCs, NPOs, or a mixture of the two, and usually are connected by memoranda of understanding or multilateral contracts in which the terms of the collaboration are laid out. Virtual projects often have stand-alone websites that create the look and feel of a distinct organization, but they do not have boards of directors or formally incorporate.

For-profit organizations: The creation of a for-profit entity should represent the clear consensus of the collaboration. If one party attempts to impose a for-profit context against the wishes of others, the odds of the collaboration succeeding drop, regardless of organizational structure. If the collaboration is intended to create a financial return on investment (not just become "sustainable" but make more funds than needed, and return those funds to owners), then an FPC is the only choice. However, settling on the appropriate form of FPC is a very important success factor.

Choosing the wrong sort of initial corporate structure can close off potential avenues to professional investment. While an early stage organization might include terms that seem attractive to initial "friends and family" investors, or structure their organization in a nontraditional manner that makes sense at the time, the creation of capital structures with unusual terms is a good way to prevent later funding.

Are you Creating a Product?

- Novel drug
- New device
- Software for clinical improvement or disease management
- Data repository for secondary use

Are you Providing a Service?

- Establishing "best practices"
- Developing and implementing "guidelines"
- Managing organizational change or tech adoption
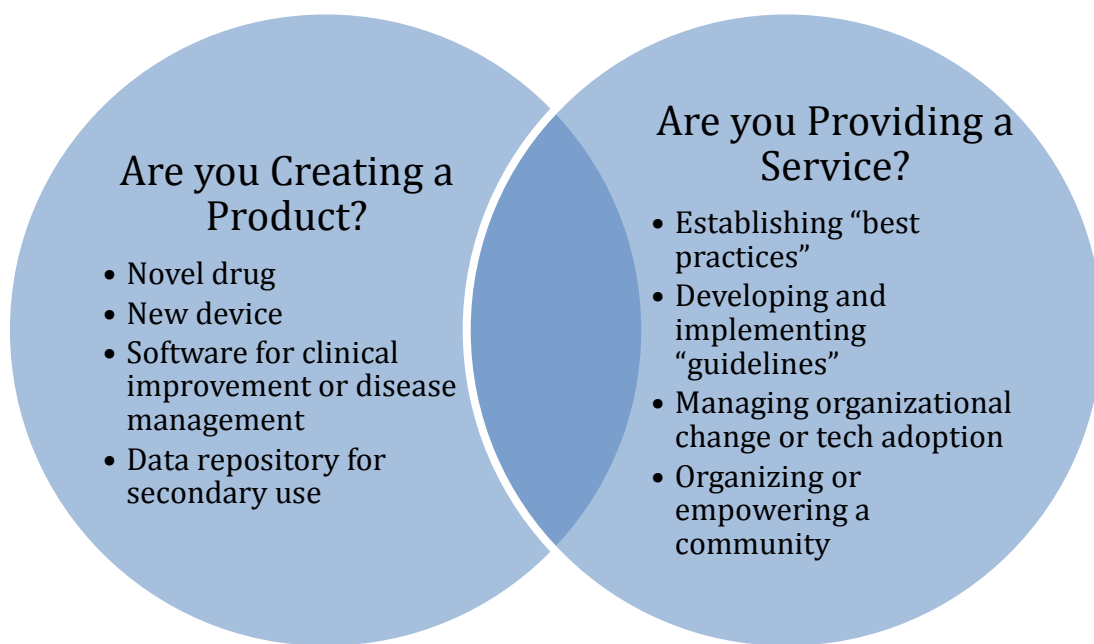- Organizing or empowering a community

Figure 3. Goals and aims of a for-profit organization

Most products leverage data in order to demonstrate value or attain regulatory approval. Devices and therapeutics always require data in order to reach the market, and increasingly, software vendors are pressured to demonstrate value to their potential client base. Services often require the creation of software and data, but don't treat them as the key value created—the "sale" is of the implementation, the training, and the

use of the software and data to get maximum value (either in decreased costs or increased effectiveness of care).

The product/service divide is an important one to get right at the beginning, as it will affect both the ideal organizational structure of the FPC and the universe of funding models available. Product companies are far more likely to attract venture investment than are service companies, for example. A general rule of thumb is that product companies are capable of high "multiples," with device developers, therapeutic investigators, and software pioneers making as much as ten times their revenues when sold. Service companies are generally felt to be capable of lower potential returns on early investments, but often are easier to "bootstrap" on small amounts of funding. Additionally, service companies can exist as sustainable entities without the pressure to either sell the FPC or make an initial public offering that comes along with venture investment.

Intellectual property choices also differ between product and service FPCs. For product companies, an aggressive IP strategy is almost a must. Medical devices and therapeutics typically involve a mixture of patent filings and data held as a trade secret as part of a strategy to survive the regulatory process. The case for aggressive IP strategies is less clear for software or data product companies due to emerging case law and trends toward open source and software as a service. However, investment and funding typically reward the pursuit of a broad patent portfolio, if only to be used defensively.

Product companies typically incorporate as classic corporations, with the Delaware "C" corporation structure an advisable choice. Investors are very comfortable with the Delaware C, as are corporate attorneys, and the transaction costs of setting up as a C corporation can be quite low. There also are variants of the for-profit product company that incorporate social benefit. Known variously as "B" corporations or "L3C" corporations, these are entities with both a profit and a social motive. Depending on the incorporation bylaws, the social mission may be as simple as "donate 1 percent of profits to charity" or far more ambitious, and—depending on the strength of its incorporation into the entity—far more enforceable. These kinds of entities are newer than the classic C corporation, and it's unknown what impact the choice of a social benefit corporation has on long-term prospects for financing and exit events or transfers of ownership.

Creating a services-based company is simpler. Service companies often begin as sole proprietorships—just the individual—though that may not work well in a collaborative context. The limited liability corporation, or LLC, is the classic model for a service company like a law firm, doctor's office, or similar professional model. Each state has its own requirements for LLC creation and most have forms online that facilitate incorporation.

Not-for-profit organizations: Organizing as a not-for-profit corporation (NPC) often is an attractive option. It removes the profit motive and frees the organization from income

taxes. However, it does not free the collaboration from the requirement to find funds and sustain itself over time, and imposes certain limitations on the sorts of activities an organization can engage in. Whether you're organized as a for-profit or a not-for-profit, you are still in the business of staying funded and paying the bills, or your project and its impact evaporate.

There are many versions of NPCs, but two stand out for IT-mediated health collaboration: the 501(c)(3) and the 501(c)(6) (hereafter C3 and C6, respectively).[1]

The C3 is the NPC that most people think of when they think "not-for-profit" and covers entities organized and operated exclusively for "religious, charitable, scientific, testing for public safety, literary, or educational purposes, to foster national or international amateur sports competition, to promote the arts, or for the prevention of cruelty to children or animals." The most likely justifications for C3 status in health collaborations would be either scientific or charitable, depending on the goals of the collaboration. A C3 will require a board of directors and incorporation in the state where the collaboration will be headquartered.

A key part of a successful C3 is a clear mission. It's easy to have gauzy social good in the mission statement, but a direct understanding of the goals of the organization is essential. It will make it easier to get the message out, and repeated. This is, unfortunately, also often ignored by not-for-profits. Many of the core entrepreneurial practices of for-profits like clear mission statements, attention to design and branding, and attention to market signals should be integrated in practice by C3s as well.

Funds for a C3 most often come from charitable gifts. These gifts can come directly from sources including individual donors, via fundraising campaigns, from charitable foundations or interested corporations, and grants from governments. C3s also can engage in contractually paid relationships with all of these entities. The most important distinction is that if a C3 brings in more funds than it spends, those funds must be recycled into the C3's mission rather than returned to investors.

Choosing a C3 is like choosing a Delaware C corporation: it's familiar, comfortable, and unlikely to cause a stir. But it's not the only option.

The C6 is a "membership" organization that allows other organizations or companies to join the cooperative and work for collective benefit. The C6 enjoys the same tax-free

---

[1] All not-for-profits will need a board of directors and financial management, including audits to ensure that the tax-free status is deserved. For all not-for-profits, the IRS maintains a public support test that must be passed each year by demonstrating that: 1) the organization receives a substantial part of its support in the form of contributions from publicly supported organizations, governmental units, and/or the general public; and 2) the organization receives no more than one-third of its support from gross investment income and more than one-third of its support from contributions, membership fees, and gross receipts from activities related to its exempt functions. A membership-fee organization, such as a parent-teacher organization, or an arts group with box office revenue are examples of publicly supported charities.

status as the C3 but can pursue different revenue models—most obviously, membership dues paid by the joining entities. This model has worked for collaborations related to Internet commerce (CommerceNet) and web standards (World Wide Web Foundation) and has real possibilities for data-driven health collaborations.

A core risk is that organizations will join when times are good and leave when times are bad. In a C6, it is very important to staff appropriately so as not to be exposed when the overall economy changes, or when the local market contracts. Another risk of C6 structures is that they often attract small vendors, or even large vendors, but not customers. In these cases, the membership is not composed of the complete ecosystem required to complete the social mission, just companies hoping to make more sales. C6s with this membership distribution face a tough route to success, and therefore ensuring that the paying and participating membership is diverse and representative is essential.

Whatever the choice, it's also good to contemplate an "exit" event for an NPC. In a for-profit, the exit is easy to define as either the sale of the company or a public offering. For an NPC, it might be the achievement of the social goal. Far too many NPCs either cannot define their goal well enough to know if it's been achieved, fail to achieve their goal, or in some cases achieve it but continue on. No NPC ever should become a permanent organization that exists for the sake of existing. If the goal is achieved, it should be recognized, celebrated, and disbanded.

Collaborative contractual organization: In many cases, a new organization does not need to be incorporated at all. If all the parties involved have their own funding and existing support structures, it often may be best to create a virtual project and move quickly to achieve the goals of the collaboration.

This kind of collaboration is simply a set of agreements among parties. Demonstration projects often take these forms. But it is just as essential for an agreement-based organization to have consensus on goals and aims as it is for a formally incorporated organization. Contracts between the parties should express these goals and aims, so that all parties are protected from either unintended "gifts" of products created collaboratively or unintended "enclosure" of those products. These contracts can be bilateral (between each party and each other party) or multilateral (between all parties via a single, shared contract). If you create a project like this, make sure to have an independent lawyer look over the contracts on your behalf, especially if you are less skilled in drafting such documents than other parties in the collaboration.

- ☐ Conduct an intellectual property inventory (p. 12–14)
    - ☐ Do participants agree to (p. 13–14)
        - ☐ Share prior intellectual property in order to develop new knowledge? (p. 13–14)
        - ☐ Share information generated in the collaboration in order to develop new products, services, or research? (p. 14)
- ☐ Nature of Product (p. 14–15)
    - ☐ Is the product digital (software, creative work) or nondigital (device, pharmaceutical, practice change)? (p. 14–15)
    - ☐ Is there a fixed cost per transaction (skill or labor intensive) or a marginal cost per manufactured unit? (p. 15)
- ☐ Industry Context (p. 15–17)
    - ☐ Are you creating something novel or improving an existing product or practice? (p. 15–16)
    - ☐ What is the regulatory burden? (p. 15–16)
    - ☐ Who is the customer? (p. 16)
- ☐ Innovation Type (p. 17–18)
    - ☐ User-Driven Innovation: Consumers of the product are the empowered innovators (e.g., the Internet) (p. 17)
    - ☐ Distributed Innovation: Participants in an innovation process are the empowered innovators (e.g., Wikipedia) (p. 18)
    - ☐ Open Innovation: Members of participant organizations are coequal innovators (e.g., pre-competitive spaces) (p. 18)
- ☐ Impact of Asserting (p. 18–19)
    - ☐ Do you have clearly defined patentable objects? (p. 18)
    - ☐ Do you have a mix of intellectual property and "knowledge" about how to use the property that isn't amenable to transfer? (p. 19)

Conduct an Intellectual Property Inventory

Technology increasingly allows for firms to collaborate with those who used to be their customers, and sharing models make it possible to both increase global inclusion and bring dividends back to the firm, staff, shareholders—and patient participants. In particular, we examine methods for that sharing technology via IP among large institutions, for-profit companies, not-for-profit entities, and groups of passionate patients, which will lead to faster access to good clinical practices, faster routes to drug discovery, and better quality and length of life. Opportunities abound to leverage IP and knowledge in novel ways toward solving the problems of attacking a rare disease, or a chronic disease of the global poor. Sharing-oriented models often involve public- and private-sector partners, and the coordination challenges, cultural differences, and misalignment of incentives would benefit from the discipline and analysis of a clearly articulated sharing strategy.

The potential for sharing of intellectual property rights (IPR) acquired by companies, universities, and individuals in life sciences and health research to drive down costs in the health care system is a deeply attractive idea, and often an implicit assumption inside collaborations. The "sharing" of creative works, data, and inventions via the IPR system should be a means of increasing capacity on the ground in patient-powered research networks in ways that "traditional" systems might not create. Collaborative systems can result in shared power, and a shared responsibility to leverage knowledge in local hands.

When collaborators join to form a new organization they should be clear in defining what intellectual property (IP) they believe they currently own, how they wish to be compensated for that IP, and what they perceive the benefits are to the new organization of the IP they are bringing to the table. Ideally, stakeholders each should conduct formal inventories and identify what they view as ideal terms of exchange. For example, a patient might offer their clinical records or patient-reported data with an agreement that they are going to be offered a higher standard of care. A researcher might offer their ability to generate publications and novel findings generated from prior published and unpublished research. An existing company may bring a portfolio of patents that they believe should be licensed by the new organization or can be exchanged for equity or offered as a donation. Failure to define and formalize prior art in the outset of collaboration can severely impede an organization's ability to be sustainable in the short and long term.

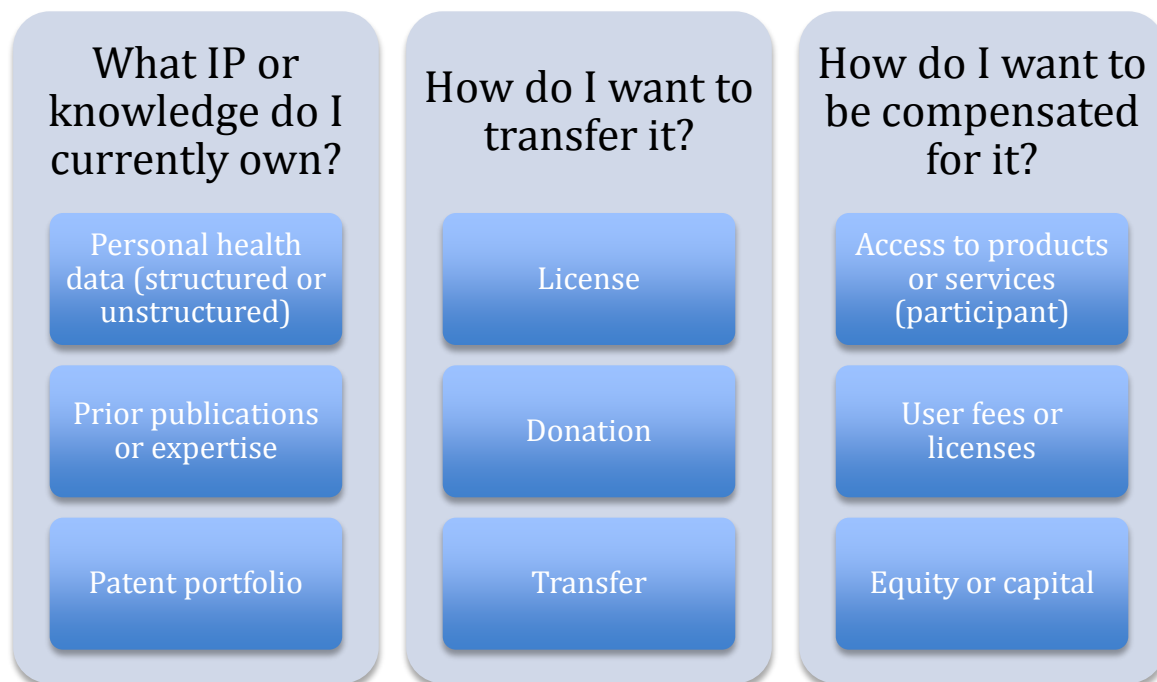| What IP or knowledge do I currently own?	How do I want to transfer it?	How do I want to be compensated for it?
Personal health data (structured or unstructured)	License	Access to products or services (participant)
Prior publications or expertise	Donation	User fees or licenses
Patent portfolio	Transfer	Equity or capital

Figure 4. Intellectual property inventory framework

It can be difficult to separate intellectual property from the context in which it was originally developed. In addition to their societal benefit, the patents developed by massive clinical research facilities are intended to benefit that facility through

commercial use; articles in a peer-reviewed journal are written to advance the recognition and publication record of the author; and software at venture-backed startups is coded in order to create return on investments. The secondary benefits derived from IPRs are an essential factor to contemplate as we examine potential models for sharing them, and significantly affect the ability of sharing-oriented models (SOMs) to act as engines of innovation in health.

Successful deployment of these models in health will require careful attention to the desired outcomes of a deployment, the types of intellectual property deployed, and the implementation of the model in a locally relevant context. This playbook outlines key areas for practitioners to evaluate in choosing an SOM for deployment, and a simple process that can increase the odds of shared IP driving measurable progress toward collaborative goals.
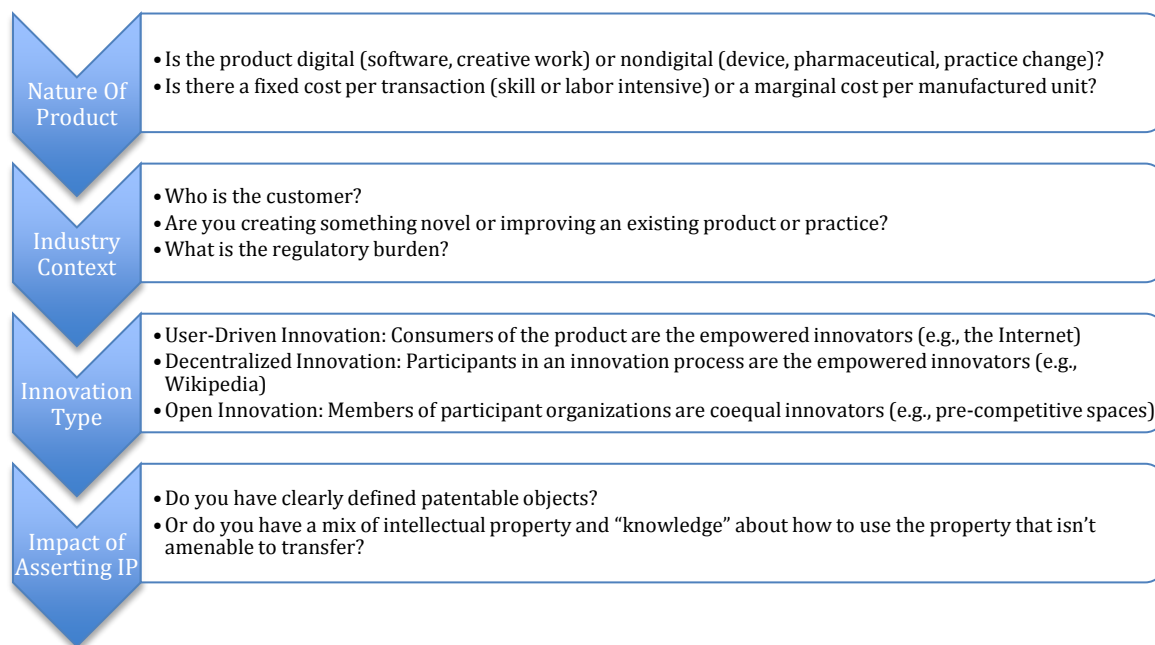
Nature Of Product	• Is the product digital (software, creative work) or nondigital (device, pharmaceutical, practice change)? • Is there a fixed cost per transaction (skill or labor intensive) or a marginal cost per manufactured unit?
Industry Context	• Who is the customer? • Are you creating something novel or improving an existing product or practice? • What is the regulatory burden?
Innovation Type	• User-Driven Innovation: Consumers of the product are the empowered innovators (e.g., the Internet) • Decentralized Innovation: Participants in an innovation process are the empowered innovators (e.g., Wikipedia) • Open Innovation: Members of participant organizations are coequal innovators (e.g., pre-competitive spaces)
Impact of Asserting IP	• Do you have clearly defined patentable objects? • Or do you have a mix of intellectual property and "knowledge" about how to use the property that isn't amenable to transfer?

Figure 5. Four-factor model

Nature of Product

First, consider the underlying character of the technology—is it digital or nondigital, and what is its marginal cost of production and distribution? These questions can help to untangle ideas about software from ideas about drugs and diagnostics; creative works from complex chemical processes; and sharing models that implicitly assume Internet distribution from those that require supply chain and physical distribution. Second, what is the actual intellectual property's context in a specific industry—is it oriented toward consumers or intermediate payers? Is it highly regulated, either to use the IPR at stake, or to get products to market? Is it a discovery field, like pharmaceuticals, or an engineering field, like software for sharing patient data? Third, what kind of innovation is the invention trying to create—is it user-led, as we see in everything from skateboarding to agriculture? Decentralized, as we see in Wikipedia and GNU/Linux? Or open, as we

see in the informal knowledge flows between firms? Fourth and finally, consider the impact of the class of IPR leveraged—are we talking about copyrightable works, patented inventions, trade secrets, or data and databases? Or, as is more common, is there a mix of assets that need to be assembled to achieve the desired goals, and how best might we deal with each class of assets?

Impact of underlying character of technology: Software is natively digital. This means some very simple economies apply to its reproduction, distribution, and use in a networked world: essentially, there is no cost involved. And although creating software is of course not free , the rapid innovation cycles have led to better and better tools for software creation and co-development, better languages for novice programmers, and a general explosion in the amount of software available at no, or low, cost. Thus, the marginal costs of production have fallen over the same time period that the marginal costs of distribution and reproduction have disappeared.

This clearly is not the case in all industries, especially those involving complex manufacturing systems and value chains. A drug is not a digital item. It is built of physical goods, in a highly customized value chain dependent on a network of suppliers and shippers, in a globalized manufacturing and supply system. Its marginal costs of creation, distribution, and use are very high compared to software, and even generic reproductions of a drug don't come for free. Thus, drug companies pursue IP in the form of patents on designs, composition of matter, forms of compounds, trade secrets in manufacturing processes, and more. Sharing a manufactured item, then, is more complex than sharing a piece of software—and "sharing" itself means something different.

Industry Context

Many collaborations in health care that deal with manufactured goods center around the launch of a new therapeutic or the discovery of a new biomarker. These problems are exceedingly intensive to solve, and the loss of control of intellectual property has enormous repercussions. The cost of discovering a drug is itself a fiercely fought topic, but estimates range from a few hundred million dollars to more than a few billion dollars spent for each drug that makes it to market. We impose rigorous controls on the safety and efficacy of new medicines, and this control imposes costs via regulatory burden. Thus, the marginal cost of creation is enormous. Atop that is the cost of production: the penalty for making a drug poorly is much higher than that of poorly making a shoe, so that "good manufacturing processes" are imposed on makers as they produce the drug in quantity. The manufacturing burden is even more expensive when the medicine is a "biologic" medicine rather than a traditional small-molecule chemical. Additionally, some development costs are imposed by biological complexity—we simply are not very knowledgeable about the human body, and thus most attempts at drug interventions in humans fail. Taken together, these costs place an enormous burden on the protection of new medicines and diagnostics. These protections nearly always take the form of patents that are used to recover the costs of development.

On the other hand, developing guidance for clinicians has less harsh costs to "go to market" and a much lower regulatory burden. Organizations seeking to improve clinical practice can conduct and disseminate studies and work with clinical societies to determine methods to optimize care. Typically in this process, peer review in a high-impact journal is an essential component of attaining customer acceptance, and in order for a guideline to be fully implemented the authors of a paper must rigorously describe the underlying data set. Some room for IP assertions remains in using a computer or software program to execute a model, or an organization's ability to refine and improve upon a published model and package the refinements or the implementation of guidelines as a service.

Recently, a handful of organizations have emerged that sell services around patient engagement, based on a number of studies published in the comparative effectiveness literature showing that engaged patients (as measured by a self-report or questionnaire) attain improved outcomes, are more adherent to medication, and have lower risk of rehospitalization. Organizations seeking to develop IP for patient engagement, clinical improvement, or guidelines implementation should understand that their customers typically are purchasing the ability to change behavior or organizations, not necessarily the intellectual property being leveraged for that change. Ownership or a closed model around the underlying IP is therefore less essential.

Additionally, some organizations may want to develop new software platforms or enabling technologies for researchers, patients, or clinicians. While there may be some aspects of these systems for which organizations should assert IPR, many components of these systems can be shared freely between collaborators, and eventually within the broader customer base. A number of companies have made a decision to release products as open source or the equivalent, and allow a user community to modify and improve their developments. These companies are self-sustaining because of their success in deploying and supporting the solutions that their user community continually improves. In the health care sector, electronic health records (EHR) organizations, clinical decision support systems, and data commons for research recently have started to trend toward fully open-source products supported through software-as-a-service business models.
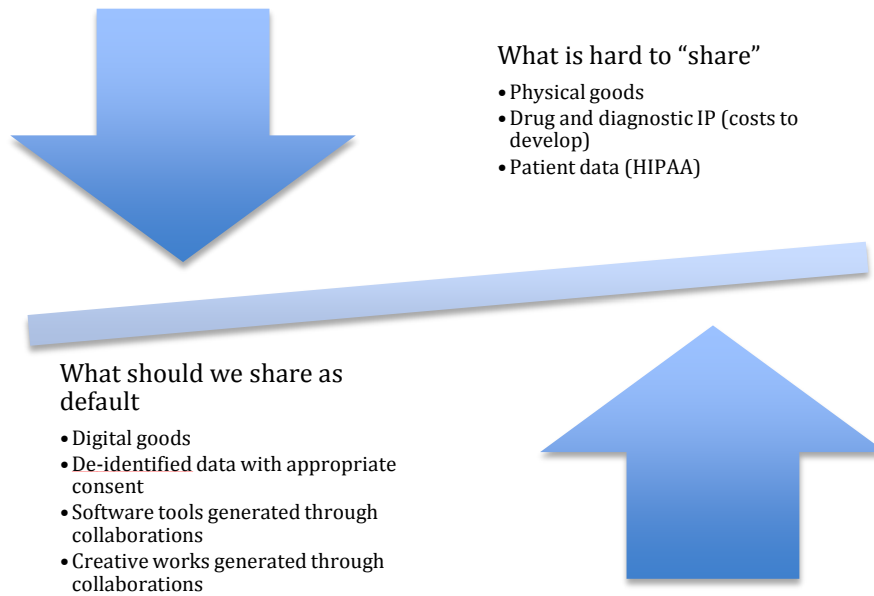
What is hard to "share"
- Physical goods
- Drug and diagnostic IP (costs to develop)
- Patient data (HIPAA)

What should we share as default
- Digital goods
- De-identified data with appropriate consent
- Software tools generated through collaborations
- Creative works generated through collaborations

Figure 6. Changing the Default State to "Share"

Type of Innovation

In much the same way that we must examine the underlying character of the technology itself, we also must examine the character of the innovation that a given sharing-oriented model wishes to inspire. The theories of innovation that dominated policy in the late twentieth century typically held that successful returns on investment in research and development sprung from investment in basic research by universities and companies that was then translated into products and protected by IP.

But the advent of a global citizenry empowered by technology and telecommunications has led to the emergence of forms of innovation that would have been deemed impossible by the older economic theories, the digital commons of software and culture being very good examples. These new forms have distinct contours, outcomes, benefits, and limitations, and various sharing-oriented models often are more suited to one form than another. It is important to map the triad of the technology desired to be shared—and the innovation desired to emerge—to the model to be used, if the goal of actually increasing innovation in health is to be achieved.

While there are multiple versions of networked innovation theory, three strands dominate: user-driven innovation (UDI), distributed innovation (DI), and open innovation (OI).

User-driven innovation (UDI) is built on the idea that end users develop a significant number of technical innovations, and that corporations often are simply a vector to bring those innovations to market in a smart package. UDI is observed across technologies with vastly different marginal costs of production and distribution, from software to

skateboards to farming irrigation systems to complex laboratory equipment. The core insight is that if the end user of a product is not happy with a product, she is likely to "hack it" and make it suit her needs, or invent an entirely new product, often without regard to intellectual property rights.

Distributed innovation (DI) references observations from free software, Wikipedia, and similar systems. DI theory explains a disruptive aspect of networks, which is the innovation power of a group of otherwise unconnected users whose individual actions "snap together" into a coherent knowledge product through a combination of norms, lightweight technology, technical standards, and (more often than not) commons-based approaches to intellectual property.

DI is notable for the absence of a central authority assigning tasks and embrace of commons approaches to intellectual property, as well as a very strong role for norms and standard technical platforms. Incentives to participate vary widely among DI communities, making it difficult to create or design DI systems *de novo*. Rather, the communities are formed by many different individuals, participating for very different sets of reasons and incentives, who self-organize around challenges and tasks.

Open innovation (OI) observes the way that increased knowledge flow leads to increased innovation, but encourages a move from informal knowledge "leakage" from inside a firm to a purposeful, intentional flow of knowledge both in and out of the firm. Knowledge flow becomes something that companies desire in an OI context rather than something to prevent or protect, letting the outside world generate knowledge that is internally useful. OI and sharing-oriented models are a tantalizing combination, as the knowledge that is flowing often is easier to share than the products emerging as a result of OI practices, but often is in conflict with existing business structures. Trade secrecy as a form of IP, for example, does not play well with intentional disclosure. And patents held by third parties not participating in sharing-oriented models can make OI troublesome to implement, even when many parties wish to collaborate.

Impacts of Asserting IP and IP Instruments

For nondigital technologies with health implications, several kinds of instruments have been proposed, created, and implemented with varying degrees of success. The most popular form of IP in this space is the patent. In a non–sharing-oriented model, the patent is used to exclude others from practicing, making, importing, and otherwise using an invention. But since patents often are incremental to one another and interdependent, there also is widespread practice of cross-licensing and strategic non-assertion in a nonsharing context.

Sharing-oriented models for patents can include direct investment in patenting technologies with the intent of sharing, buyback of patents in order to share the technology, and "pull mechanisms" such as advance purchasing committees and prizes. These models tend to assume a central authority paying for the patent, but perhaps a liberal (though "nonstandard" compared to FLOSS or CC) license to practice. Models

that assume a more significant institutional presence include patent clearinghouses, pools, and exchanges. There also are standard public patent licenses ranging from simple "one to many" offers to licenses specifically targeted at "nonmarket" or development use, and strategic nonassertion in order to enable health priorities.

Other approaches include liability rules (which outline pre-patent divisions of revenues among collaborators through contract mechanisms), mapping and landscaping to help users find clear routes around patent thickets, public-private partnerships to share for-profit and not-for-profit uses of joint patents, and regulatory processes such as compulsory licensing or accelerated patent review that affect development. Nondigital technologies also often are subject to trade-secret regimes or contractually based sharing regimes, of which Materials Transfer Agreements in chemistry and biology are perhaps best known and quantified.

The explosion of shared and shareable "Big Data" has significant potential to affect all forms of innovation. Cheap, embeddable, and mobile sensors make significant elements of lifestyle and environment quantifiable and open up an entirely new avenue for generating data on health. To date, technologies and practices regarding "Big Data" are more focused on reusing data "exhaust" than on sharing-oriented models. A thorough discussion of data sharing should cover the investments in infrastructure and education required to allow patients and individuals to take full advantage of the opportunities created by a data-driven society.

ROLE OF TECHNOLOGY, INNOVATION, AND DATA PRACTICES

- ☐ Using data to build a learning health care system (p. 20–22)
    - ☐ Generating a hypothesis (p. 21)
    - ☐ Validating an approach (p. 21)
    - ☐ Measuring an outcome (p. 22)
- ☐ What sort of data do you need? (p. 22–23)
    - ☐ Clinician generated (p. 22)
    - ☐ Patient generated (p. 23)
    - ☐ Researcher data (p. 23)
- ☐ How can the data be collected? (p. 23–24)
    - ☐ Physical sensors (p. 23–24)
    - ☐ Electronic data capture (p. 23–24)
- ☐ What regulations are the data subject to? (p. 24–25)
    - ☐ HIPAA (p. 24–25)
    - ☐ Privacy and consumer data protections (p. 24–25)
    - ☐ Human subjects research protections (p. 24–25)
- ☐ What methods will you use to: (p. 26–30)
    - ☐ Store data? (p. 26)
    - ☐ Clean data? (p. 27–28)
    - ☐ Harmonize data? (p. 27–30)
- ☐ Whose standards will you use and support? (p. 30–32)
    - ☐ HL7 standards (p. 30)
    - ☐ Meaningful Use standards (p. 30–31)
    - ☐ Research standards (p. 31–32)
- ☐ Identifying and controlling for bias (p. 33–34)
    - ☐ Understanding perceptions around data bias (p. 33–34)
    - ☐ Developing practices to control for data bias (p. 33–34)
- ☐ How will your organization change clinical practice? (p. 34–35)
    - ☐ Informing guidelines and care (p. 34–35)
    - ☐ Publishing research (p. 34–35)
    - ☐ Creating and assessing technology (p. 35)

Using Data to Build a Learning Health Care System

Although we have seen the emergence of novel sensors, an explosion in our capacity to generate, store, and process data, and the ability to create technologies to improve health and care, a great deal of innovation still is necessary to create collaborations that eventually will build and operate a learning health care system.[2] Information on individual experiences with illness and wellness often are disparately scattered among various actors and frequently not collected at all. Collaboration can bring stakeholders together to work toward a common goal, improving outcomes and care, but it is useful to define some of the technologies that either are necessary components or fortunate byproducts along the path.

---

[2] The simplest definition of a learning health care system is one in which all actors have access to the appropriate data to make the appropriate decisions at all times.

Fundamentally, a learning health care system requires a framework for evaluating the positive or negative consequences of interventions among populations. There are multiple components to this framework, including the ability to: collect data from populations in clinical settings as well as free-living populations; aggregate that data in a secure repository; contextualize that data within existing knowledge; and conduct analysis on that data. Multiple technologies have been created across the range of the analytic process, but these systems rarely are integrated, outside of some of the most extreme inpatient settings such as critical care, or in demonstration programs in telemedicine, such as the Whole System Demonstrator project or the Delivering Assistive Living Lifestyle at Scale (DALLAS) program in the United Kingdom.
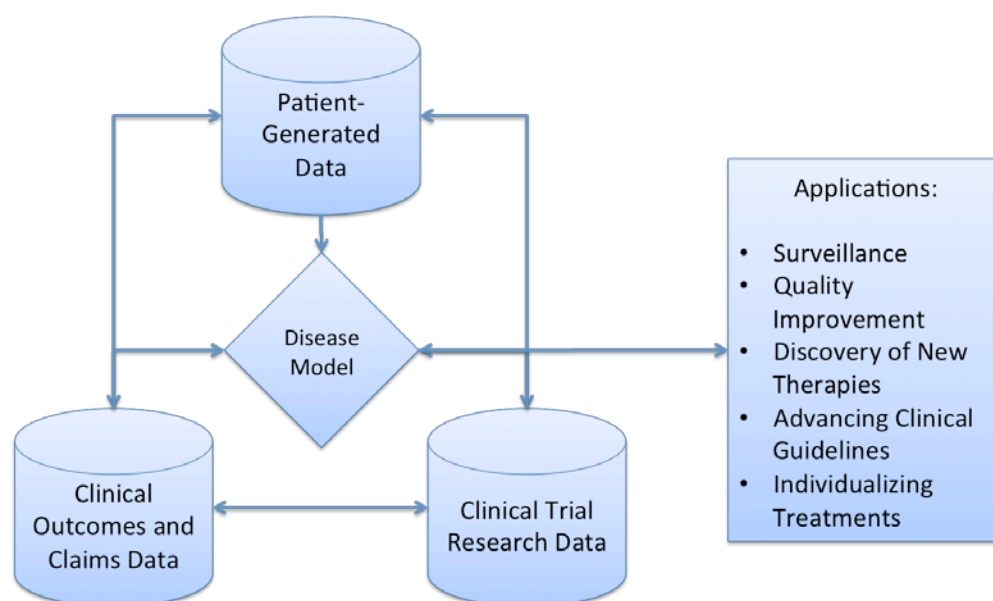
Figure 7. Data-driven learning health care system

The role of data in such a system is to accomplish three primary tasks:

1. *Generating hypotheses*: Data, both qualitative and quantitative, can be used to down-select from a universe of possible research questions to a set of questions, or hypotheses, that we are able to test and explore in the short term. In the absence of data, many stakeholders in health care currently use heuristics, mental models, or educated guesses in order to prioritize new research. While simple observations grounded in data aren't a substitute for the formal products of the scientific process, data still can be a useful guide in early stages of any collaboration.
2. *Validating an approach*: Data can be used during the process of implementation to determine whether collaboration is headed in the right direction or off on the wrong track. The ability to frequently assess collaboration from a data-driven perspective allows investigators to collect early signals of whether or not an intervention is meaningful in a population; model how a disease might affect a community or an individual; and

determine whether or not a user community might exist for a product or solution through examining things like usage statistics or uptake.

3. *Measuring an outcome*: The most critical utility for data is the ability to concretely measure outcomes of interventions. The standard model for identifying a meaningful outcome has been the randomized control trial, but the ability to collect massive amounts of data from continuous data streams may enable new organizations by providing tools for robust assessment that do not rely on strict inclusion or exclusion criteria. Rather, outcomes may be able to be determined from observing free-living populations.

Data Sources

A new collaboration should assess what sorts of data needs it has at the beginning of the effort, and what data needs might arise during the course of the effort. While five or ten years ago most efforts relied primarily on clinically generated data or data acquired through research, the emergence of participatory research paradigms and components of the Affordable Care Act that place data control in the hands of patients have simplified patients' ability to be part of formal research processes. When initiating a new effort, it is useful to consider the strengths of each data type: clinician generated, patient generated, or researcher generated.
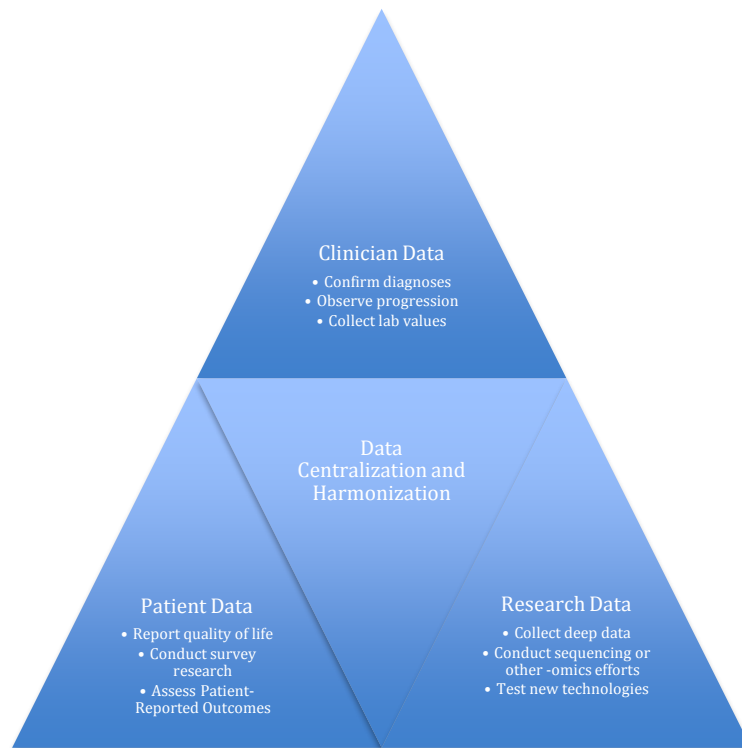

Figure 8. Data type centralization and harmonization

Clinician generated data is highly useful for getting data certified by a medical professional. Data generated in the clinical environment contains information on

medications prescribed to patients, dates of diagnosis, and formal diagnostic records. Additionally, clinical data contains laboratory diagnostic values, which can be difficult for a patient to easily remember or may be presented to a patient not as a discrete value but as an indicator of health status. Difficulties in accessing clinically generated data typically have to do with data labels, data regulations, and the willingness of clinicians to export data to third parties.

Patient data can be useful for annotating and amending clinical data on such topics as medication adherence or resolution of diagnoses between multiple providers. In addition, patients are uniquely suited for reporting on quality of life and other Patient-Reported Outcomes (PROs) and may be able to report these outcomes with higher frequency than clinical data. Last, patients can take part in survey research or other qualitative research, like focus groups.

Research data is best suited for deep investigation into patient characteristics. For example, outside of the research environment it can be difficult and costly to collect proteomic or genomic data. Additionally, the research setting has high utility in the deployment and testing of new technologies, as rigorous research practices more adequately control for unforeseen events.

Data Collection

Fundamentally, any data stream is comprised of an underlying set of "sensors" that create a statistical space.[3] Sensors can take the form of hardware systems made up of accelerometers for measuring movement and activity, voltmeters for measuring skin impedance, typical medical sensors such as an electroencephalograph (EEG) or electrocardiogram (ECG) array, or next-generation sensors based on optical measurement of temperature, pulse oximetry, or other physiologic status measurements (e.g., blood pressure). Further, laboratory based assays and arrays measuring chemical composition of blood, absence or presence of proteins, RNA/DNA or other molecular diagnostics can contribute to an overall "sensor system."

---

[3] All sensor systems have the potential to introduce error into a data set, whether it is caused by electromagnetic interference, miscalibration of a probe, inefficient protein binding, or simple human error in entering in a diagnosis. Therefore, datasets are "statistical" spaces with accompanying measures of accuracy.
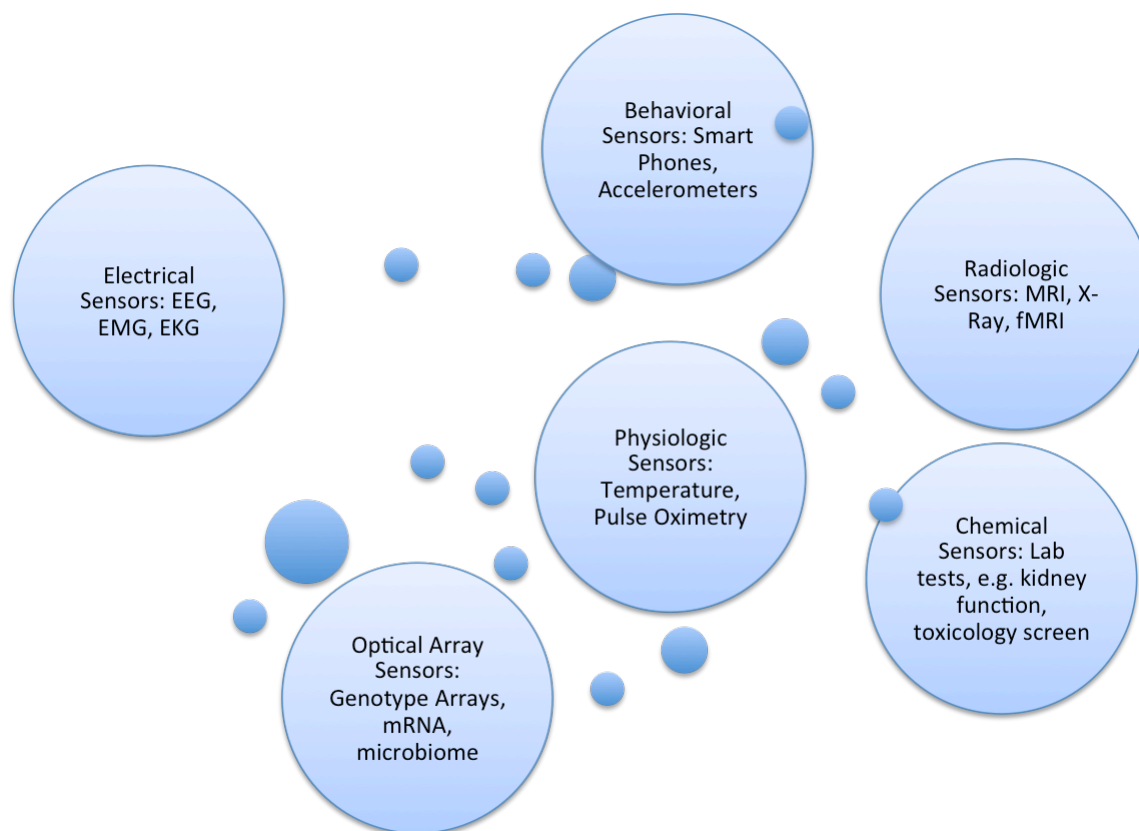
Figure 9. Emergence of a sensor ecosystem in clinical practice

In addition, software systems are contributing to continuous streams of sensor data. ICD-10 diagnostic codes can be readily complemented by data contributed from clinical providers. The advancement of natural language processing continues to add relevance to clinical notes, allowing the most traditional form of narrative data to be used to add structure to data captured through the EHR system. A number of stakeholders in clinical systems have made the case for robust capture of metadata through EHR systems to capture components of the practice of medicine in in-patient settings. The extraction of data from systems that identify who is providing care to what patients at any point in time could be used to inform how differences in staffing processes or frequency of contact between patients and clinical staff contribute to outcomes. However, standards have yet to be identified or emerge from major clinical systems in a consistent manner. Last, in both the outpatient and clinical settings, a number of methods are used to allow individuals to contribute patient-reported outcome data, both through structured questionnaires and validated instruments and through providing narrative context.

Regulation of Data

There is additional complexity in the creation of systems for storing, sharing, and manipulating biological and medical data. A patchwork of regulation controls the permissions by which data generated from clinical care can be accessed for research, including local determinations by ethics boards; application of the Common Rule and

interpretation of the Health Insurance Portability and Privacy Act (HIPAA) in the United States; the European Data Protection Directive; and the Helsinki Accords and Belmont Reports defining research ethics and medical care. New policy frameworks, including those put forth by the Personal Genome Project and Portable Legal Consent, expand the potential for an individual to contribute their data to a shared commons, lowering the legal barriers for reuse of data, but technical systems for data storage and reuse in many cases still are built under varied interpretations of current policy. Some academic centers have chosen to take a dual approach, blending technological implementations with policy safeguards. The University of Michigan's Honest Broker Office has been set up to facilitate researcher reuse of data and provide oversight of data systems. The Electronic Data Methods Forum, a project from the U.S. Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality, has created an issue brief on data governance of personal health information (PHI) for research that represents neutral interpretation of current regulation and thought leadership on this issue.

Table 1. Data regulation

Class	Name	Details
Framework/Policy	Belmont Report	Identifies basic protections and ethical guidelines for U.S. human subjects research (1979)
	Declaration of Helsinki	Establishes international principles for human subjects protection in medical care and research (1975)
	Federal Policy for the Protection of Human Subjects (Common Rule)	Institutionalizes guidelines for human subjects protection across all programs funded by fifteen federal agencies (1991)
	U.S.-EU Safe Harbor	Creates a self-certification regime for US organizations to comply with European Commission's Directive on Data Protection (1998)
Laws	Health Insurance Portability and Accountability Act (HIPAA)	Limits access to personal health information without prior authorization. Sets standards for de-

		identification of health data (1996)
	Genetic Information Non-discrimination Act	Prevents an organization from discriminating against an individual on the basis of biomolecular data (2008)
	Patient Protection and Affordable Care Act ("Obamacare")	Guarantees individuals the right to access a copy of their own health records (2010)
Institutional Safeguards	Institutional Review Board	Provides a mechanism for institutions to independently determine the ethics of an individual investigator's research proposal
	Honest Broker Office	Ensures institutional compliance with regulations regarding data protections by acting as a "clearinghouse" for requests for data and carrying out oversight

Storing, Cleaning, and Harmonizing Data

Storage of data is additionally complicated due to the size and sort of data generated by clinical systems. Genomic data files and raw imaging files in particular are immense outputs generated by high precision machines. While personal storage devices may allow for an individual to move around a copy of, for example, an image of a radiologic file, an actual data file generated by magnetic resonance imaging (MRI) contains information that is not visible to the human eye and ultimately not rendered in the PDF or film that a clinician may consult. Nevertheless, this information is of high utility to research systems. For example, the data generated by an MRI machine can be used to train algorithms that detect the onset of rheumatoid arthritis or other joint injury before the point that a defect is visible to a radiologist. By discarding information that may not seem immediately clinically relevant due to storage concerns, we frequently remove the ability of a learning health care system to evaluate pertinent data in retrospect.
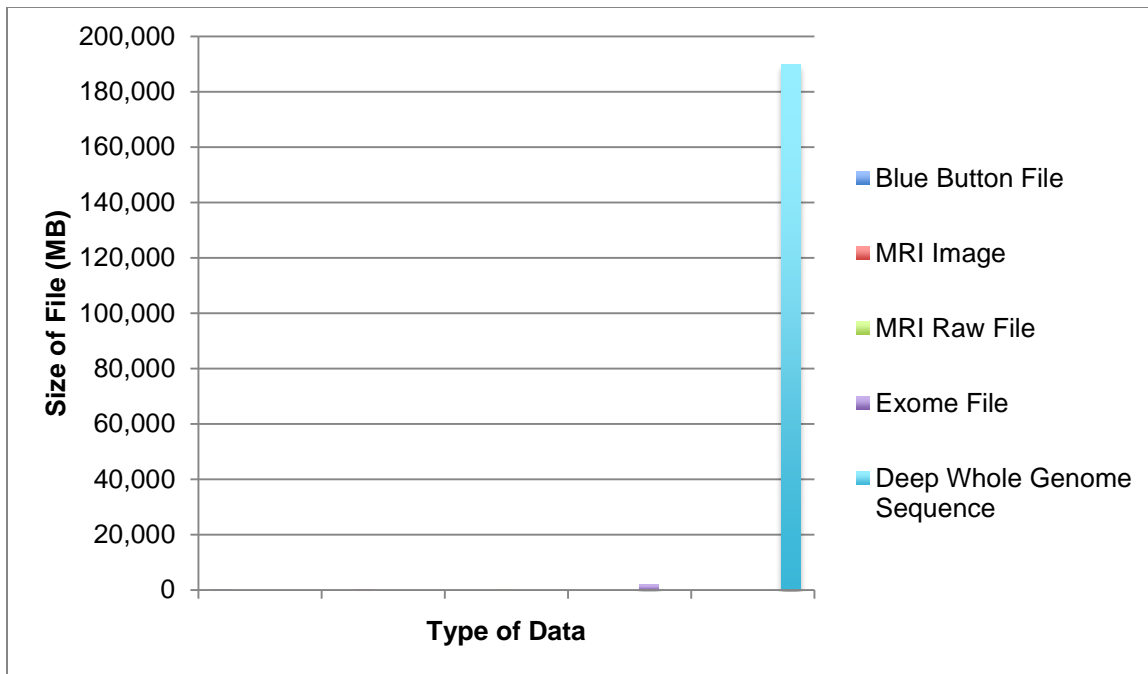
Figure 10. Size of data file for an individual

The manipulation and preparation of data for analysis is being explored by organizations interested in building collaborative research environments and scalable systems. Cloud-based computing and collaborative research environments allow for low-cost storage and processing, and enable teams of researchers to bring their own expertise into joint environments for problem solving, hypothesis generation, and analysis. These systems can be restricted to users with credentialed access, and data security can be handled by simply not enabling functions that would compromise an individual through re-identification.

In these and other environments where data is aggregated for analysis, new techniques are being brought in by the emerging field of "data science" in order to evaluate how interventions affect outcomes. These analytic techniques complement traditional, frequentist statistical methods such as linear regression or probabilistic modeling and include the use of Bayesian statistics and machine learning techniques. While novel analytic techniques can bring insight to both emerging and classical challenges in biology and medicine, the frameworks for adoption and recognition of the utility of new techniques continue to lag behind the application of these analytics to other fields. Achieving a p-value of .05 in a randomized, double-blind, placebo-controlled trial remains the "gold standard" for medical evidence of efficacy of intervention.

Although many facets of a learning health care system have been independently developed, there is a dearth of ability to create cohesion between systems. Data access barriers and proprietary data capture systems have limited the ability of organizations interested in analytics to be aware of the sort of data captured by a clinical system and how that data is stored. Inputting data into systems that may be useful from a research

perspective can be burdensome for both clinical stakeholders and individuals impaired by chronic illness. Further, few systems guarantee the quality of data inputted into any secondary repository. These sorts of "middleware" between sensors and storage systems, among types of storage systems, and between data and analysis, are both critical to the functioning of a learning health care system, and an area where significant investment has yet to be made.

Part of the challenge in constructing a unified data framework is the heterogeneity of standards between different sorts of actors, and the lack of incentives to unify disparate standards through common architecture. An EHR vendor primarily is concerned with client needs and reimbursement frameworks. An academic or chief research officer (CRO) operating a research database follows a separate set of rules regarding collection of data, driven by ethics boards, statutes governing research, and the demands of regulatory agencies and peer review. This can be seen most distinctly in the conflicts between Meaningful Use-compliant EHR data records and what the National Institutes for Neurologic Disorders and Stroke considers to be generalizable Common Data Elements. A clinical export of all data captured in an EHR does not meet the standards for sufficiency or data labeling required by the NINDS. Any new actor entering into the health care space should be aware of this mismatch.

Table 2. Differences between data captured in EHR and sample Common Data Elements

Data Category	Meaningful Use Compliant EHR Data	National Institutes for Neurological Disorders and Stroke General Common Data Elements
Demographics	- Birth date - Gender - Ethnicity - Race	- Birth date - Gender - Ethnicity - Race - Maternal ethnicity - Maternal race - Paternal ethnicity - Paternal race
Social Status	N/A	- Education level - Marital status
Medical History	247 individual condition categories - Gestational age by category	- History taken - Any medical problems - Category of medical problems - Medical history text - Medical history SNOMED CT code - Condition start time/end time - Condition ongoing - Birth weight - Gestational age - Post-natal age - Post-conception age

Family History	• Congestive heart disease only	• Family history Diagnostic code • Family history diagnosed or not • Which relative family history
Behavioral History	• Tobacco use/non-use • Current cigarette smoker • Depression screen positive/negative • Standardized tools for assessment of pain/cognition	• Behavioral Health Assessment Date/Time • Tobacco current use/Tobacco prior use • Tobacco use start date/Stop date • Type of product used/Amount of product used • Alcohol current use/Prior use • Alcohol use start date
Physical Exam	• 29 individual diagnostic categories for cancer presence/absence • Sensory/visual exam of foot • Retinal exam • Best corrected visual acuity • Framingham 10-year risk (heart disease)	• Physical exam date/time • Physical exam conducted • Body system symptoms • Body system abnormal • Physical exam text • Exam SNOMED CT • Abnormality clinical significance • Handedness
Vital Signs	• Heart rate • Blood pressure systolic/diastolic • BMI • Height • Weight	• Vital sign date/Time • Heart rate • Respiratory rate • Blood pressure systolic/diastolic • Position during blood pressure • Temperature measurement/Unit of measure • Temperature site • Weight/Unit of measure • Height/Unit of measure • Head circumference
Laboratory Tests	• 37 distinct tests • Test values	• Specimen date/time • Panel category • Test name • LOINC code • Test performed (Yes/No)

		• Test result/Unit of measure/UCUM code
Prior and Concomitant Medications	• 145 individual medication codes	• Medication prior or concomitant use • Medication name • Medication RXNorm code • Medication text • Medication dose/Unit of measure/UCUM code • Dose frequency • Medication route type • Start date/End date • Ongoing indicator

EHR Standards and Data Governance

Part of the reason for this mismatch has to do with the method through which EHRs are certified, and research requirements. Typically, EHRs are governed by Health Level 7 International (HL7) standards and, in the United States, by the Meaningful Use incentive program. As EHR adoption has increased, additional organizations have attempted to build upon existing standards frameworks, including the CommonWell Health Alliance, an industry collaborative of organizations that have pledged to a secondary level of joint interoperability. Last, the U.S. Department of Health and Human Services has invested in core technologies to enable the health care IT marketplace, most prominently the secure messaging platforms Direct and Connect and the "Blue Button Plus" View/Download/Transmit (V/D/T) Consolidated Clinical Document Architecture (C-CDA) file format and import/export standardized code.

HL7 standards are a set of proprietary standards devised to allow for interoperability of various sorts of software implemented in the clinical setting. Primarily, HL7 standards deal with communication between EHR systems and other sorts of software. As of March 2013, the primary set of HL7 standards have been published under open access. However, HL7 standards considered "foundational," including syntax implementation for clinical decision support alerts, are governed by proprietary licenses, membership in HL7, and download costs for implementation guides.

One HL7 standard, Infobutton, provides a protocol for rendering information generated in a third-party environment inside an EHR. This system can be used to provide a linked, patient-level identifier between an EHR and a secondary database, and use that linked identifier to display pertinent information such as a JPG or PDF file within the EHR. Unfortunately, Infobutton-compliant systems do not always supply dimension data necessary to render a file to the requesting database or to the EHR environment. This can result in full-page-size images being compressed into thumbnail-size boxes.

Meaningful Use is a multistage incentive program operated by the Office of the National Coordinator for Health Information Technology (ONC/HIT or, more commonly, ONC),

part of HHS. Clinicians that implement Meaningful Use-certified EHR systems in their practices are eligible for direct incentive payments. EHR programs achieve Meaningful Use certification through application to a nongovernmental certification body. The government can audit organizations that receive payment to ensure that implementation actually has occurred and that specified metrics are attained through the use of EHR systems.

Table 3. Meaningful Use criteria and goals

Stage 1 (2011–12) Meaningful Use criteria focus on:	Stage 2 (2014) Meaningful Use criteria focus on:	Stage 3 (2016) Meaningful Use criteria focus on:
Electronically capturing health information in a standardized format	More rigorous health information exchange (HIE)	Improving quality, safety, and efficiency, leading to improved health outcomes
Using that information to track key clinical conditions	Increased requirements for e-prescribing and incorporating lab results	Decision support for national high-priority conditions
Communicating that information for care coordination processes	Electronic transmission of patient care summaries across multiple settings	Patient access to self-management tools
Initiating the reporting of clinical quality measures and public health information	More patient-controlled data	Access to comprehensive patient data through patient-centered HIE
Using information to engage patients and their families in their care		Improving population health

Source: http://www.healthit.gov/providers-professionals/how-attain-meaningful-use

The majority of measures in Stage 1 dealt with ensuring that clinicians were aware of potentially harmful drug allergies and drug/drug interactions, collected demographic and basic health data in a routine fashion, implemented at least one clinical decision support rule, provided data to patients upon request, and clinicians maintained up-to-date lists on current diagnoses for the majority of their patients through the EHR system. The majority of Stage 2 goals are attained through expanding the populations covered in Stage 1, improving methods for transacting with or reporting electronic data, or providing patients more control over data.

Meaningful Use Stage 2 has a specific requirement that patients are able to view, download, and transmit (V/D/T) the data collected about them through clinical systems. The Department of Health and Human Services has supported this metric through

expanding upon the Veterans Health Administration's "Blue Button" initiative with an effort known as "Blue Button Plus." Blue Button Plus is a self-certification pledge to provide a specifically formatted Consolidated Clinical Data Architecture (C-CDA) XML file through a request facilitated by a secure messaging protocol, Direct, that was developed specifically to allow secure exchange of health information. Clinical systems can implement Blue Button Plus through code codeveloped with HHS at http://bluebuttonplus.org/. Blue Button Plus data is compliant with LOINC (lab values), RxNORM (drug names), SNOMED CT (medical terminology ontology), and ICD-10/ICD-9 (medical procedure code).

Table 4. Contents of Blue Button Plus C-CDA data

Section	Description
Header	Patient information and demographics
Allergies, Adverse Reactions, and Alerts	Includes status and severity of each.
Encounters	Surgeries, ED visits, etc.
Immunizations	Immunizations and vaccines
Medications	As prescribed by the provider
Care Plan	Planned activities and encounters
Discharge Medications	Part of hospital discharge summary
Reason for Referral	Written reason for referral
Problem List	Concerns, complaints, and observations
Procedures	History of procedures
Functional and Cognitive Status	List of impairments
Results	Includes laboratory tests
Social History	Observations like smoking, drinking, etc.
Vital Signs	Includes height, weight, blood pressure, etc.
Discharge Instructions	Written discharge instructions

Source: http://bluebuttonplus.org/healthrecords.html

In research databases, standards have been generated that pertain to governance more than data exchange. The National Institutes of Health (NIH) recently has implemented technology that allows study coordinators to generate a Global User Identifier (GUID) code for each participant in a study, allowing for data exchange between systems while removing elements of personally identifying information. Additionally, multiple disciplines have generated lists of Common Data Elements specific to their field of research, e.g., traumatic brain injury Common Data Elements put forward by a U.S. Department of Defense, Veterans Affairs, and National Institutes of Health joint working group. No universal standard or retrieval capability for Common Data Elements across disciplines currently exists. Last, the Surveillance, Epidemiology, and End Results Program registry system proliferated by the National Cancer Institute provides a population overview on cancer statistics using a common data model implemented across multiple sites.

Identifying and Controlling for Bias

Models build on observationally collected data historically have been weaker than models built on data collected during randomized controlled trials for several reasons. First, in order to aggregate retrospective data for an observational study, medical professionals would have to perform exhaustive chart review and convert narrative clinical notes into a standard data format. While narrative content is useful as a complement to data, individual practitioners seeking to extract meaning, retrospectively, from narrative are not as reliable as systems to collect data at the point of care. The process to adequately control for individual bias in converting narrative content into data requires multiple investigators and frequent statistical testing to ensure data validity. Second, prospective observational studies historically have had biases of region or health system specificity. Third, real limits in storage and computational power historically have created difficulties in aggregating and analyzing large datasets. Fourth, datasets, when compiled, generally were "fit-for-purpose" and did not include data outside the bounds of the investigation that may, in fact, have been critical to systems biology contributing to disease.
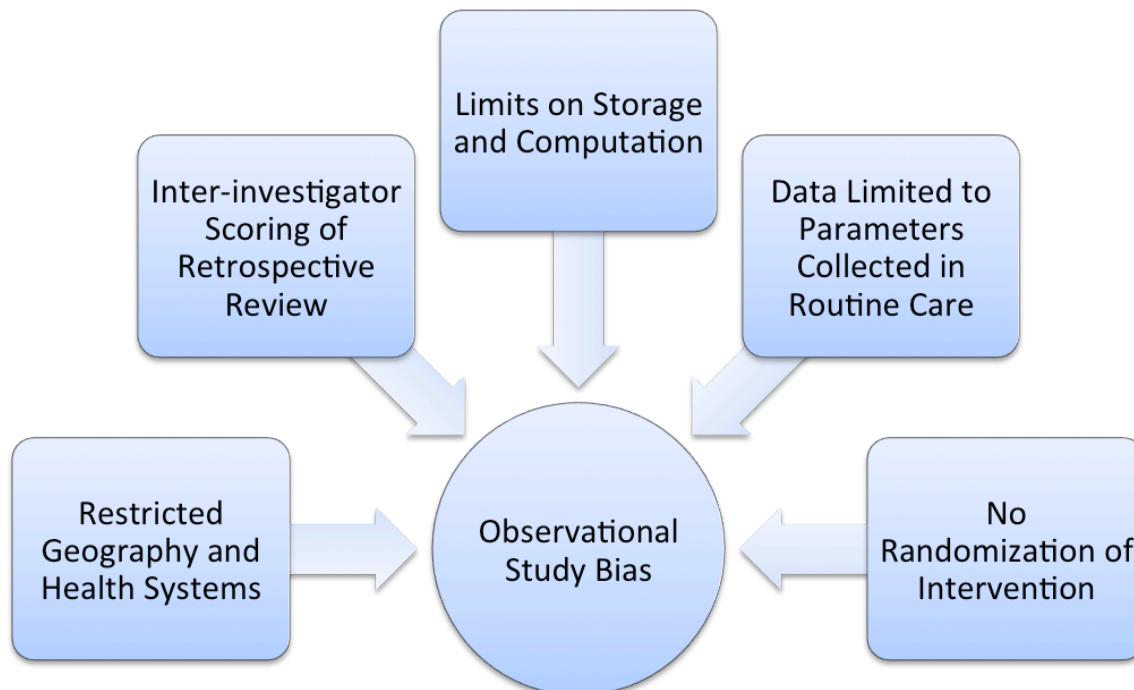
Figure 11. Common sources of bias in observational studies

Modern observational studies can overcome many of these biases. Organizations such as the Health Maintenance Organization Research Network have created large datasets of clinically relevant data and outcomes generated from EHR data. Mini-Sentinel, an initiative of the U.S. Food and Drug Administration (FDA), collects millions of data points from health care data in order to create an early detection system for adverse drug events that is driven by observational data. Further, the use of Bayesian statistics and

applied mathematics can extract insights from multidimensional datasets and reduce biases caused by using inappropriate analyses.

Insights gained from modern analytic techniques using datasets collected in real time can be difficult to deploy and implement in secondary settings. For example, real-world observational data feasibly could demonstrate the efficacy of a generic drug used off label in modifying outcomes among a subpopulation with a genomic trait identified using a research assay. However, the FDA still would require comprehensive safety and efficacy trials. These trials likely would use standard analytic techniques to evaluate efficacy and involve significant recruitment burden, as they would have to identify subpopulations with both the disease and the relevant genomic trait.

The double-blind placebo controlled clinical trial was developed to rule out all other factors save an active agent in assessing the efficacy of an intervention. However, the Placebo Effect, the Hawthorne Effect (whereby subjects modify their behavior due simply to the presence of an observer), preexisting comorbid conditions, and challenges in studying rare or diversely presenting disease are all real challenges to modern clinical trial execution. Investigators should consider trade-offs between extracting knowledge from observational data and the quality of data gathered from trials conducted in ideal conditions.

Changing Clinical Practice: Disseminating Knowledge and Implementing Findings

Once data is compiled into a common framework, standardized, and properly governed, it can be used to inform clinical practices. Currently, there is a gap between advanced methodologies for conducting data analysis referred to as the emerging field of "data science" and the level of evidence needed by more traditional stakeholders in medical systems. Understandably, medicine historically has been conservative due to the risks involved in implementing new treatments without assessing their quality. As new techniques for demonstrating efficacy or harm emerge, however, these models should be evaluated and incorporated into care rather than forcing new forms of evidence generation to fit inappropriate thresholds.

Disparate systems exist for generating and accessing a knowledge base of prior findings, contextualizing new analysis, and disseminating research. Clinicians typically do not have the time or resources to conduct exhaustive searches of peer-reviewed literature, even when faced with a novel case. Instead, they rely on the opinions and expertise of trusted peers, and on more readily accessible services such as UpToDate. For those clinicians who do engage in search, or for individuals facing a chronic, complex illness, journal access fees create an additional barrier to reviewing the broader literature. Some health care systems, such as Kaiser Permanente, have implemented institutional knowledge bases for clinical access. Patient-facing communities like HDBuzz, a Huntington's disease information portal, attempt to curate research findings and provide high quality summaries that are relatively free of medical jargon.

Guidelines and clinical decision aids frequently are formed from an evidence base that includes clinically collected data as well as peer-reviewed research. However, the formulation of many decision aids and guidelines generally are not formalized with a repeatable methodology. Several groups of academics have begun devising separate standards for both computational guidelines (i.e., guidelines that use similar logic to computer programs and thus would be amenable to decision support in the era of EHR) and aids for "shared decision making," a process whereby the patient is given educational materials that outline options, typically among surgery procedures.

Table 5: Guideline formulation by method and type

	Not stated	Balance Sheets	Expert Consensus	Expert Consensus (Consensus Conference)	Expert Consensus (Delphi)	Expert Consensus (Nominal Group Technique)	Informal Consensus	Other
Assessment of Therapeutic Effectiveness	6	2	271	4	6	4	3	11
Counseling	12	8	580	32	17	17	63	2
Diagnosis	61	0	866	52	144	31	72	4
Evaluation	58	2	1140	60	165	44	100	10
Management	52	7	1462	97	74	49	118	11
Prevention	32	46	776	78	24	13	49	4
Rehabilitation	0	0	78	3	3	10	5	1
Risk Assessment	22	14	747	45	42	17	80	5
Screening	13	33	429	27	20	7	21	4
Technology Assessment	4	0	66	10	1	0	1	6
Treatment	49	4	1336	86	87	51	88	17

Source: http://www.guideline.gov/matrix.aspx

Technology creation and assessment: An obvious use for systems of health data is in the creation of new technologies. Systems for the transfer and aggregation of health data, securely and with proper governance, between organizations should enable collaboration between patients, innovators, clinicians, and researchers. Additionally, current gaps in the learning health care system and suboptimal system implementation require continued improvement. Of particular interest are technologies that collect data with minimal burden on the part of the patient or the investigator; technologies that enable complex analysis; technologies that disseminate and integrate new findings to traditional stakeholders, patients, and caregivers; and technologies that assist clinicians, payers, and individuals in making complex decisions. During the creation of these technologies, it is important to identify the roles, risks, and rewards of the various groups convening to create them.

In addition to developing new technologies, organizations can use health care data to evaluate and improve upon existing technologies. Advantageously, many of these activities can fall under the auspices of quality improvement rather than research and are governed by a less strict set of rules. Unlike technology development, organizations come to the table with significantly more intellectual property in technological assessment and improvement. Roles, risks, and rewards in technology assessment should reflect these differences.

The Case for Openness

Ideally, by this point in the guide, members of a new collaboration will have reached agreement on initial steps. Goals have been agreed upon, IP has been identified and negotiated around, and the intermediary technologies and sources of data necessary to achieve a short-term and long-term goal have been identified. However, even the best-formed plans rarely survive reality, and accomplishing a path to sustainability requires recognizing when trade-offs need to be made and how stakeholders can declare victory in the absence of achieving the goals they agree upon at the outset.

One model for allowing collaborations to achieve maximum success is by conducting all activities in an open, transparent, and shareable manner, not only with the immediate members of a collaboration but also with the broader community. This has the benefit of ensuring that anyone who can leverage resources developed by a collaborative is freely able to do so, but restricts the ability of members of the collaborative to maximize outputs of their IP. In the short term, an open collaborative has "first-mover" advantage, and can exploit their new knowledge before any other organization. In the long term, an open collaborative builds expertise that they can utilize to seek profits or fees in the market.

A fully closed collaborative that seeks to change the broader market in a more formal manner encounters other risks. It is harder, for example, for external stakeholders to verify whether a new technology performs as advertised. The peer review process ameliorates some of this, but barriers to publication and dissemination still exist, and the research-to-practice gap remains a major concern.

The methods of control form a gradient from fully open to fully closed, and large portions of the gradient can be described using a small set of conditions. This pattern applies to property rights in a gradient from the public domain to "all rights reserved," with intermediate stages forming around requirements like attribution, noncommercial use restrictions, "share alike," and so forth. The pattern also applies to privacy restrictions, including requirements like do not re-identify, do not redistribute, do not recontact, do no harm, and so forth. And the pattern also emerges in technology, from sites that use technology to make it easy to harvest data to sites that enforce registration, and on to sites that apply digital restrictions measures to enforce zero data harvesting.

These patterns very often intersect and reinforce one another. Taken together, the three form a significant chilling effect on legal data reuse, although just as often, user fatigue

from reading lengthy terms of use on consumer websites can mean that the policies are never read, much less understood. These patterns also drive the implicit business models of collaborations built around data—and if the patterns don't fit the model, the collaborations are unlikely to achieve their goals.

Instead, we need non–zero-sum ways to think about control on data. The greatest innovation might think about how designed systems can result in automated flows from closed to open—as a dataset loses economic value, for example, its controls might automatically decay as well. A dataset might carry a "deposit now" mandate but start behind a firewall to give a scientist a chance to mine it for new insights, then auto-post at six months to a collaboration's public site with a noncommercial restriction for all to see, and then the noncommercial restriction itself might decay after six more months, leaving a dataset in the public domain and subject only to citation.

This concept of mixed controls, and of assigning a "half-life" to controls that guides datasets toward a more fully open lifespan, has another potential benefit. If attached to a common framework technically—standard data formats, pragmatic annotation, and curation—then applications and discoveries that are built in the open can move back into the closed portion. A mobile application built on open data can just as easily be used on closed data earlier in its decay cycle, because the underlying formats and architectures are the same. Users who are collaborating on the public data will be likely to look for, and request, access to data earlier in the life cycle, increasing collaboration on private data as well. Mixed controls also means that data providers can and should be able to mark data or parts of datasets with different permissions, and to transform datasets so that much of the content can pass through.

Operating as a Trust

A similar mechanism for control, drawn from existing property law, is the land trust. There are two kinds of land trust, each of which resonate with the organizational designs already explored for health collaborations:

1. A private, not-for-profit organization that, as all or part of its mission, actively works to conserve land by undertaking or assisting in land or conservation easement acquisition, or by its stewardship of such land or easements.
2. An agreement whereby one party (the trustee) agrees to hold ownership of a piece of real property for the benefit of another party (the beneficiary).

The trust as a business model has several attractive elements. It is inherently a collective mechanism for negotiating access to shared goods; it allows groups with widely divergent interests to work together; and it provides guidance on how to govern a collaboration as parties enter and exit. Trusts also form a gradient from strictly controlled (as in a conservation trust to protect a wetlands property) to very liberal (as in a real estate investment trust designed to maximize the construction of condominiums). It is likely that health IT collaborations will need to account for a similar range of desires

and preferences among patients, providers, payers, and communities, and the trust model may well provide the needed flexibility.

The trust model (or similar models such as data "banks") implicitly assumes that the collaboration is creating "club goods," which could be either software or data. Club goods are excludable, meaning that parties can be left out of the club—for payment, in traditional excludable goods; or, in a data estuary, it also could mean for not signing onto the legal, technical, and social "norms" of the club . However, digital club goods also are nonrivalrous, which means that one party's use of the data or software does not prevent its simultaneous use by another party.

The trust could be an intermediate stage between public and private, where users sign up for constraints that are something between open and closed, but where discrimination against classes of users is not allowed. Thus, anyone willing to accept the terms could access the data, but anyone violating the terms would be liable to suit by the managers of the club on behalf of the collective. Some data simply would pass through the trust "stage" on its way to fully open, and data that for regulatory reasons can never be fully open (human identifiable, for example) might sit in a trusted space whose sole reason for existence is privacy protection but imposes zero additional controls.

CONCLUSION

Collaboration in health projects is increasingly necessary. There are simply too many institutions, too many devices, and too many communities of individuals to corral into silos or monopolies—and doing so would lose the potential for innovation that comes from collaborative work. But any successful collaboration eventually will rub up against a variety of systems that will punish any failure to address legal, technical, and organizational details up front.

Those systems include reimbursement systems, regulatory approval processes, grant applications, institutional review boards, and investment diligence. The intent of this playbook is to help potential collaborators prepare for contact with these systems, and to prevent "catastrophic success" from overwhelming these collaborations after they begin to create value. It is far easier to address these issues earlier than later.

The ideas in this playbook are a jumping-off point. The text here will be mirrored at [URL] so that collaborations who implement the ideas will be able to post case studies, ask questions, add annotations or comments, and form support communities. As implementations emerge, we will identify reusable assets and add them to the text: these may include template contracts and other legal documents, key software components, filing documents, and more. The goal is a living resource, not simply a document frozen in time.

As part of achieving this goal, the playbook will be rolled out in a series of events in Boston, Houston, Pittsburgh, Penn., and Kansas City, Mo., in 2014. Please contact the

authors if you wish to host an event in your city, or if you have reusable open assets related to implementations that you wish to add.

DISCLAIMER

This document provides general information about legal topics but it does not provide individual legal advice. The authors are not a law firm and do not provide legal services. Using this playbook or sending us email does not create an attorney-client relationship. The authors provide this general legal information on an 'as-is' basis, make no warranties regarding the general legal information provided, and disclaim liability for damages resulting from its use. We strongly recommend you consult an attorney before entering into contracts, documents of incorporation, or other legal relationships.