How a pediatric oncologist tackles data standardization as a grassroots effort

Bringing together the entire research community to solve data interoperability

The crucial role of health informatics in bridging care and research

The monumental task of forming a national data collaborative to better treat COVID patients
Welcome to the “Health IT and Data-Sharing” issue of the Clinical Research as a Care Option (CRAACO) newsletter. This issue highlights the landscape of data-sharing, utilizing technology, exciting data commons projects and thoughts from informatics experts on bringing clinical research closer to care.

In this issue:

Teresa Zayas Cabán, PhD, National Library of Medicine at NIH, discusses formulating the National Health IT Priorities for Research and what collaborations she’d like to see in the industry to promote data-sharing.

Dipak Kalra, PhD, European Institute for Innovation through Health Data, describes incentivizing clinicians to capture clinical-grade data and increasing transparency for the general public in data-sharing.

Rebecca Kush, PhD, Elligo, provides an overview of the work Elligo has done with the FDA’s data harmonization project, as well as the challenges facing wider data interoperability today.

Amy Cramer, Pfizer, brings the audience up-to-speed on Project Vulcan and HL7 FHIR, and how collaborations are already testing data-sharing abilities in connect-a-thons.

Melissa Haendel, PhD, University of Colorado Anschutz Medical Campus, talks about her role in the National COVID Cohort Collaborative, the largest publicly available limited dataset in US history.

Samuel Volchenboum, MD, PhD, University of Chicago School of Medicine, describes how he approaches data interoperability as a grassroots effort, encouraging adoption from the bottom-up with other physicians.

Joshua Rubin, Learning Health System Initiatives at the University of Michigan, describes putting data-sharing into action for the benefit of patients with learning health systems.

The CRAACO newsletter is the official publication of the Clinical Research as a Care Option conference. The CRAACO program is scheduled virtually for April 26-27, 2021. Enjoy the Winter 2021 issue.
inSeption Group is unified in our passion to bring therapies and an improved quality of life to patients around the world. We encourage all industry professionals to join us in returning to the foundational principles of clinical research: to efficiently develop superior treatment solutions.

Every inSeption team member has an intense desire to instill hope in each patient we serve. Dedicated to making their treatment journeys as positive as possible, coupled with a determination to fight for better, longer lives, inSeption Group is reminding patients everywhere that Hope Exists.

We commend all individuals who have chosen to participate in clinical trials. Not only have these pioneers possibly benefited themselves, but their valor continues to help future generations with added effective treatments and an improved quality of life.

If you would like to learn more about how inSeption has become a refuge for clients and professionals who have been stifled by mainstream, high-volume, commoditized outsourcing models, reach out to us.
Inside the Issue

6 Teresa Zayas Cabán, PhD, Assistant Director for Policy Development at the National Library of Medicine (NLM) at NIH

8 Dipak Kalra, PhD, President of The European Institute for Innovation through Health Data (i~HD)

10 Rebecca Kush, PhD, Chief Scientific Officer for Elligo Health Research

12 Amy Cramer, MMCi, BSN, CPHQ, Director on the Global Product Development Strategic Team at Pfizer

14 Melissa Haendel, PhD, Chief Research Informatics Officer at the University of Colorado Anschutz Medical Campus

16 Samuel Volchenboum, MD, PhD, Biomedical Informaticist and Associate Professor of Pediatrics at the University of Chicago, School of Medicine

18 Joshua Rubin, JD, MBA, MPH, MPP, Program Officer for Learning Health System Initiatives at the University of Michigan Medical School Department of Learning Health Sciences

Recommended Media

Click below for media that furthers the conversation of clinical care integrating with clinical research.

podcast Overcoming Site Regulatory Issues to Optimize Clinical Quality

podcast Connecting Care and Research for the Patient with a Decentralized Trial Model

podcast How a COVID-19 Research Study was Formed with Oracle, Javara and Wake Forest

podcast Weighing Your Options in Considering Clinical Trials from a Patient Perspective

newsletter Clinical Research as a Care Option newsletter “Diversity & Community Engagement” issue

recap Full recap of the 2020 Clinical Research as a Care Option event
ONLY ELLIGO accelerates the overall research process, ensuring quality and integrity and focusing on the patient.

Only Elligo's transformative and decentralized model for clinical research, System of Accelerated Research (SOAR®), adheres to regulatory processes while also accelerating the research process to increase physician and patient involvement.

Contact us to learn how SOAR® can transform your research.
www.elligohealthresearch.com
How Health Informatics is Crucial in Bridging Care and Research

Teresa Zayas Cabán, PhD, is Assistant Director for Policy Development at the National Library of Medicine (NLM) at NIH, leading NLM’s policy development and implementation activities. Prior roles included Chief Scientist at the Office of the National Coordinator for Health Information Technology (ONC), and Chief of Health IT Research and Acting Director of the Division of Health IT at the Agency for Healthcare Research and Quality (AHRQ).

What is the role of health informatics in assisting the integration of clinical research and clinical care?

Health informatics is critical to bringing those worlds closer together and one of the linchpins in doing so. The ultimate vision is for health informatics to help us address researchers’ data and infrastructure needs to enable research that leads us to discovery more quickly, more efficiently. But also, the goal is to facilitate integration of any knowledge that comes from that research or from that discovery into clinical care effectively. The ultimate goal is to make quicker gains in health improvement. Whether that’s helping with diagnosis or treatment, new care delivery models, or understanding how we deliver care differently to be more effective. Informatics has a clear goal in helping us to achieve that vision.

One of the priorities centered on integration at the point of care. What is the process of making the information useful and contextualized?

There are many colleagues in the informatics community who have been working on clinical decision support for many, many years. There are observations that some of the ways clinical decision support has been designed and implemented could be improved upon so that it is more useful within the context of clinical care. Being able to share the data through standardized application programming interfaces will create opportunities for other developers to think about how the information is presented to clinicians, and to develop new tools to leverage those data. Those tools could sit on top of or alongside existing electronic health record systems. I think we’re moving in the direction of making that a reality.

But doing so will require a very good understanding of what clinicians are trying to do when they treat their patients. In my previous role at AHRQ, we had one funding opportunity that sought to ascertain clinical information needs to help us better inform the design of some of those solutions. I expect that we’ll need some of that research to continue to inform some of these tools moving forward.

You were part of the group at ONC that created the National Health IT Priorities for Research. In your research, what were the hurdles that you came across preventing wider data interoperability?

Advancing standards and sharing data in a standardized way will go a long way towards addressing interoperability issues, particularly interoperability of data. The opportunities that remain are ensuring that the data can be easily integrated and used across systems.

For example, if you and I are using the Fast Healthcare Interoperability Resources® (FHIR®) specification to exchange the information, and we even agree on some of the underlying data standards so that we can code the data consistently, once I send you a clinical record, where does it land in your EHR? How do you integrate that as a clinician or as a patient? How do you integrate information from across multiple provider EHR systems for use to facilitate consistent interpretation and use of the data?

Because at the end of the day, you want to be able to gain insight and knowledge. Making the data available is one step, but ensuring they can be integrated and used is critical.

What were the considerations for security on the different levels and different types of usage?

What we outlined was the need to ensure privacy and security of the data. We outlined some issues around consent and data sharing in the agenda. Specific to that, there are a couple of different issues to parse out. One is developing and implementing clear privacy policies and practices, and being transparent about those.

Transparency is particularly important, not just with clinicians, but with patients and their caregivers, so they can understand how data are being shared and used. And then implement security practices consistent with those policies.
We talked about the need to innovate around how data are shared, and how we parse which data are shared for what purpose, so that it’s less of a burden on a clinician to make sure that their patient’s data isn’t going where it shouldn’t, but the system has features and functions that allows data to be shared. For example, in the case of COVID-19, to facilitate sharing data in whatever deidentified way is appropriate for public health reporting and tracking, we need additional innovation in that regard to make that a reality.

What needs to happen to ensure that good quality and compliant data is captured and useful for all relevant parties?

We need to move away from thinking about what needs to change at the point of capture and think about how we enable capture of data that are needed, in the level of quality that we hope for through other means. And we need to do that in ways that do not require additional work from patients and caregivers and the clinical team. Unfortunately, sometimes what ends up happening is that the burden falls to patients or their caregivers to capture or report on some of those data. For example, ONC has been exploring the use of automation. There might be opportunities to add instrumentation or leverage data that comes from devices or even from other systems – Human Services, for example. Census tract data can be used to predict certain outcomes.

Identifying what other data sources might be available that we can share in a more automated way might be one way to go. The other is finding ways to address those quality issues that don’t impose an additional burden on clinicians as they’re capturing data. Whether that’s a redesign of the interface, or implementing some quality checks in the backend, and figuring out how to imppute or analyze data.

While presenting on a panel last year, a presenter on the same panel argued that data not of quality for research are arguably not of quality for clinical care, either. So, sometimes we shift the conversation towards thinking about quality issues, and how that impacts research. But the reality is that clinicians and care providers need high-quality data too. If we think about addressing some of these needs together, that would also go a long way in both advancing data needs and quality of those data, and reducing burden on clinicians and patients.

You have spoken about FHIR many times over the years. How have you seen the broader uptake, understanding and usage of that evolve?

The development and adoption of FHIR has actually moved fairly quickly in healthcare compared to the development and adoption of other health data standards. There’s a lot of support and enthusiasm behind that, not just from the health IT developer community and the healthcare provider community. You have insurers, researchers, and the pharmaceutical industry – NIH is encouraging the use of FHIR and issued a notice doing so in 2019. All of that has helped with interest, but also adoption.

There is interest in making sure the specification works as well as possible for as many — as there are different needs from different healthcare stakeholders. Its use is now required under HHS regulation. ONC and the Centers for Medicare & Medicaid Services (CMS) published regulations last year that will require the use of FHIR APIs. ONC, in particular, advanced the use of FHIR APIs because they recognized that FHIR was well-received and adopted within the healthcare community. Use of a standardized API will go a long way towards advancing exchange of health data, not just for clinical care, but for other purposes as appropriate.

The other thing that ONC finalized in its Cures Act Final Rule is the United States Core Data for Interoperability. This is a set of data classes and elements that will be standardized with a set of specified vocabularies. It’s something that will further enable interoperability; we will use FHIR to exchange data in a standardized way, but the data underneath will also be standardized according to specific vocabularies. That’ll help not just for clinical care, but for researchers to be able to bank on those data elements to be in standard form within an EHR.

Are there collaborations or partnerships that you think would be good for creating infrastructure?

A more explicit collaboration between healthcare, health IT developers, the informatics community, and the research community to ascertain common needs, and to see if we can carve a path forward across those needs. Even within healthcare, there are different use cases that might have some common needs. Providing care, issues in population health, public health, research, and post-marketing surveillance, and even regulatory decision-making might have some common needs that we can come together as a community to identify and address.

One of the things we mentioned in the ONC National Health IT Priorities for Research agenda was also the need to do more to engage the patient and participant community, and the need to be more inclusive. If we want healthcare to work for everybody, then we need research that is inclusive of everybody.

That will require engaging diverse populations in research and engaging the institutions that serve them as well. We need to either fund partnerships with better-resourced organizations, provide resources or tools that they might be able to use and leverage, and understand what their needs might be. That way the threshold for participating in research for them is not out of reach.

The ultimate vision is for health informatics to help us address researchers’ data and infrastructure needs to enable research that leads us to discovery more quickly, more efficiently.
Tackling the Next Big Hurdle in the Wider Adoption of Mass Data Sharing and Interoperability

Dipak Kalra, PhD, is President of The European Institute for Innovation through Health Data (i~HD). He plays a leading international role in research and development of Electronic Health Records, including the development of ISO standards on EHR interoperability, personal health records, EHR requirements, security and data protection.

What is the work you lead in utilizing health data?

I’m the President and founder of the European Institute for Innovation through Health Data. We are involved in a combination of projects and activities that help to tackle some of the challenges that prevent everyone collectively from making better use of health data.

Those issues include data protection, complying with data protection practices, upholding good quality data protection, data quality assessment and improvement, the promotion of interoperability standards, the assessment of clinical research software – to make sure that systems uphold data protection practices – and also promoting the value of data across the ecosystem, helping stakeholders to come together.

What is the holy grail to enable data capture at source to be high-quality and usable?

Every time you try to impose a correction for poor data quality, even with the best quality statistics, you are slightly guessing what the original might have been. And the further removed you are from the original context, the less precise that correction usually is.

The holy grail is to enable data capture at source to be high-quality and usable, that means structured and coded, or easily extracted from narrative, accurate for the patient, and mapped into standardized formats, so that it can easily be reused.

The challenge is how to make that happen, because data quality can be time-consuming. So the real challenge is how do we incentivize good data capture at source? And how do we facilitate it through the systems and practices and education so that we get good data? Additionally, you don’t know where you stand unless you measure it, so we also promote the measurement of data quality; that way an organization can tell where it is.

Bring patients in. Let them be part of that transparency.

Do we have the technology to truly be able to achieve the level of interoperability needed?

Absolutely. There are always innovations. And it’s exciting, when I look at research papers, how people are cutting the edge in all sorts of ways, especially with sensors, voice recognition, touchscreens and other devices.

However, there is so much that is already industry-standard out there that healthcare doesn’t use well enough. So even if we caught up with yesterday’s technology and used that everywhere, we would be doing incredibly well, rather than saying let’s wait until tomorrow’s technology is ready. If we do that, we won’t make a start.

The other thing is that we have to look at the incentives. If busy clinicians have to perform the extra data entry effort to code their entries more completely, do they get the value from that data? Are they able to get automated decision support? Do they still have to do clerical duties? Do they have to type in something that the computer should be able to show them already?

We’ve got to help them. Because if they find themselves in a situation where they are working harder, and the computer’s not working harder for them, they will resent it. And then you’re not going to get their buy-in.

Is there hesitation amongst clinicians that capturing clinical-grade data would be too overwhelming to their already busy schedules?

That is a reasonable fear, and it’s understandable, so you’ve got to look at enabling the incentives. The first one is that if you look at all the stakeholders who’d like to use data, there are many of them. But who has to pay the price of collecting good data? The answer is the healthcare organization. The healthcare provider carries the can for enabling good data, and everybody else gets the benefit. These other stakeholders need to contribute to the solution.
They need to be looking at how they can incentivize data capture and how they can subsidize the costs. This needs to consider that the clinical coalface has become, over the years, very good at reading free texts such as previous clinical letters, so why should they make the effort, especially if their clinical systems are clumsy?

The other big issue we can push for is much better clinician-patient centric utility of the systems that are procured: the user interfaces, the ability to code, the ability to create templates and structures. Some of those things are quite clumsy and clunky; they’re slow to use. You actually can’t expect clinicians to do fast and furious data entry, even in a well designed system, if the computing power is slow. It takes a total quality perspective to really improve the quality game. And our Institute’s efforts, and those of other colleagues around us, is to really push for that holistic perspective.

What guidelines or best practices would you encourage to make sure there is trustworthy handling of data?

At the coalface, in the clinical environment, healthcare organizations have an obligation, and the general data protection regulation does require this, for transparency information. If the information is being used in a number of ways that the patient might not anticipate, even if it’s internal use for quality assessments and safety monitoring, etc., it’s good practice that it is declared.

Then if we look at the downstream uses of the data, it’s often possible to aggregate the data and anonymize the data and therefore make it extremely difficult to identify an individual. The principle is that the further you are removed from the patient, the less you need to know who the individual is. And you can adopt information security safeguards and data protection safeguards to actually make sure the individual isn’t recognizable (or at least that it is very very hard).

An important corollary is that even if you could robustly anonymize the data, a number of surveys and groups that I’ve interacted with have said that individuals still care about how their data is used. They don’t want it used in unscrupulous ways. The bigger organizational structures that are brokering population-level data sharing need to have transparency too.

Involve patients and citizens in those decision access rules. Don’t just assume that you can sell a nice, glossy explanation to the public about why they should trust decision-making bodies. Bring them in. Let them be part of that transparency. And then they can also be your voice to explain to their colleagues, and in the public space, that these decisions are being made in sound, robust and ethical ways.

How do we incentivize good data capture at source and facilitate it through the systems and practices and education so that we get good data?

In your presentation at CRAACO 2020, you highlighted one of the bigger challenges of demonstrating value to convince reluctant stakeholders. Can you expand on that?

It’s all very well to say, “We’d like to use your data for other purposes.” But that’s a kind of dark hole into which your data goes, and you never see anything back. So when I talk about demonstrating value, what I mean is that when insights are gained from data, those insights should be shared.

I don’t mean sharing confidential-level information, but high-level summaries; I find that companies, academic groups, etc., are very open to sharing a summary of what’s been done with data and what has been learned. They’re not revealing commercial secrets. It’s not giving away the mathematical formula, but it is enabling the public to recognize: “When data like ours is used by these kinds of organizations, this is what happens; this is the good that comes from it.”

The other area of value demonstration is closer to organizational level. So, for example, we are in the middle of a project with hospitals, looking at their data to see if it is of a good enough quality that it can be learned from, and if it is, then my plea is: get learning! And shout about it: tell other hospitals about the benefit of using health data to improve.

What do you see as the next big hurdle in the wider adoption of mass data sharing and interoperability for all stakeholders?

There are a few big categories. The biggest category, in my opinion, is getting public support for the scaling up of their data being used, and establishing a sufficiently accepted portfolio of security safeguards that data protection authorities and everybody else is comfortable with, and patient organizations can endorse.

The second is meeting the standards about good quality and interoperable data, including the suitable use of terminology systems and other semantic structures. Any one standards body doesn’t hold the kingpin on this topic, it’s really the multiple standards landscape that helps with that challenge. We not only have to get adoption of standards, we must get better alignment and concurrent use of a multi-standards ecosystem. That means the standards bodies need to be working closer together than they do, even though they’ve done a lot in recent years to collaborate.

The third one is making sure that the data quality is good enough, which I’ve described earlier. And fourth is to actually embrace the patient or the citizen. We are shifting towards the patient being a capturer of data, through wearables, smartphones, apps that collect data at home. That’s part of the health data ecosystem. It’s not replacing the value of clinically captured data, but it’s adding value through additional data.

This is a fragmented multi-vendor, multi-app ecosystem. There is limited standards adoption of personal health systems in partnership with healthcare. We need the ability to glue together citizen data and clinical data into a common, usable environment for learning.
How Data Standardization is the Key to Clinical Innovation

Rebecca Kush, PhD, is the Chief Scientific Officer for Elligo Health Research, president of Catalysis Research, and founder and president emeritus of Clinical Data Interchange Standards Consortium (CDISC).

A huge benefit of data standardization is starting trials faster, providing better care for patients and not duplicating work. With that great potential, how can we convince people and demonstrate value?

It’s very hard to streamline processes if everybody wants their data in a special format for them and you can’t do any kind of standardization. If you can standardize things upfront, like case report forms and edit checks, you can start a study significantly faster. A key goal at CDISC was that standardized data would make it easier for the FDA to review the data they received so they could do higher quality reviews, get into the reviews quicker, and not have to become data managers.

People often think, “We have to do it our way, and this way is better.” They think standards block innovation. It’s actually the reverse. If you have standards so that you don’t have to worry about the data exchange, you can start innovating around the tools, user interfaces, and other things.

We started looking at the case report form standards. Janet Woodcock and Mark McClellan, when he was the FDA Commissioner, wrote the Innovation vs. Stagnation document and referenced the need to make it easier for sites so that every sponsor doesn’t provide a different case report form, forcing them to redo training and relearn how to do the data collection.

You can still make the case report form look and feel exactly how you want. But on the back end, if the data is formatted the same, it can be exchanged easily and retain the meaning. So, there’s still a concept out there that this blocks innovation. Once people see what standards can do for them, they see how much faster things can go and that innovation isn’t blocked.

Eighty percent of the data collected on most trials is common data: it’s concomitant meds, medication history, physical exams, demographics, etc. If you can standardize all of that, with edit checks around that data, some have shown you can start up studies 70% to 90% faster.

Can you describe Elligo’s involvement in the FDA common data model harmonization (CDMH) project?

The FDA is very interested in using real-world data to augment clinical trial data. There were some initiatives like PCORI and the 21st Century Cures Act, which are pushing FDA to ask how you can use healthcare data, in addition to what you get in with randomized clinical trials, to help make decisions and augment the study data.

Mitra Rocca, an informatician at FDA, wrote a grant proposal and got it funded through PCOR Trust Fund, which funds federal agencies to work together. That was Phase I. They needed somebody to help find data partners who would sign these data agreements and run a query to do a use case for the FDA to see if the methodology would work to take the various models that were being used within research networks—PCORnet, i2b2, OMOP, and Sentinel—to the BRIDG model. The use case query was run by data partners to obtain data in one of these four data models. They said, “Let’s bring in the data across the finish line to FDA and see if FDA can use that real-world data to inform decisions.” Elligo received a small grant through FDA to be essentially the “middle person,” to find these data partners and to work with FDA and NCATS on this project.

There was a lot of work done on mappings with many federal agencies involved. When we finished Phase I, there were a few lessons learned. They said, “We need to do a Phase II leveraging FHIR,” which we’re in the middle of doing now. We’re trying to find data partners who can run a query for the FDA, provide data in FHIR or one of the common data models (and the NIH NCATS will map that into CDISC), and then map that into CDISC, which goes to the FDA. We’re trying to make it easier to use real-world data to help in regulatory decisions.

If you can have standards, you don’t have to worry about the data exchange; you can start innovating.
You were a part of a paper that describes why a systems approach is needed that enables sharing of data and lessons learned at scale, to combat COVID. In what ways has COVID changed the conversation around data-sharing, and what infrastructure is needed?

The infrastructure needed is complicated, and that’s what this paper was about. The data comes from a lot of different sources; we then have to be able to understand and share that data. It needs to be high-quality, standard data. It takes quite a coordinated effort. When we looked at the whole system, it felt like boiling the ocean. We asked, “What could we try to do that would make an impact now and has a reasonable scope?”

We’ve been working with CDISC and trying to find out what’s going on out in the world to come up with a way to prove that you’ve been vaccinated. It’s not a very big data set. So, all we’re trying to do is say, “We’ve got 10 data elements; let’s exchange these so that they are understandable.” If I say to you that I got my vaccine 01/11/2021, you might think I got it on January 11, 2021. But that date in the U.S. would be interpreted as November 1, 2021, in Europe. Let’s agree on a date format; let’s agree on what we want to have as the data. And then all these technology apps out there that are being developed to document vaccinations will be better able to use that data.

What are the conversations you believe are needed to convince people to make this a priority in their work?

There are two main conversations. One is that people don’t really understand that when we talk about a data standard, we’re really talking about the format. You can collect any data you want. But if I tell you I’m going to collect information on sex or gender and the choices are “male, female, undifferentiated, or unknown,” and one person gives choices of “M” and “F” and another codes these as “1” or “2,” there is confusion. It doesn’t really matter what we decide to use as long as we come to a consensus agreement on the data standard.

Exchanging the data along with the meaning depends on including the appropriate metadata along with the data itself. We’re really talking about standardizing the metadata that explain what the data mean. Using the prior example of dates, if you receive a bunch of numbers, and you don’t know that this number is the month and that number is the day, you’re not necessarily going to interpret the date correctly—it may depend on where you live. We often use the example of the Mars space orbiter that crashed because scientists in Europe and the US were using different units of measurement. It is like showing your units when you do a math problem. That’s all we’re trying to do. We’re not trying to tell you how to format your case report forms, we’re not trying to tell you what data to collect, or how to build your app, we’re just saying to make sure that when you exchange the data, the people who receive it understand the meaning of it.

The other conversation is around value sets. It’s basically the picklist that you give people to go with a data element, for example, the severity of an adverse event.

If people use a different picklist, then one cannot combine, aggregate, or compare that data because you’ve collected it a whole different way. HL7, at one point, had 15 different ways to express sex or gender. You can have 15 ways to say it, but do we need that many? Usually four or five options is enough to handle what we need for that field.

Can you speak to the concerns around privacy and security?

That’s where you need to understand who owns the data and who is allowed to use the data. When the National COVID Cohort Collaborative formed, everybody signed a data use agreement, and that took several months to actually develop. When we started the CDMH project, we wanted to work with one unnamed EHR vendor. They spent so long being worried about what we were going to do with this data (which was going to the FDA, who has a right to see it anyway) that, after a year, we didn’t have an agreement.

One of the problems we ran into with the first CDMH project was when the data partners did the query, out of ~7000 patients, they came out with less than 10 patients. To be exact, they had one patient that met those criteria. So, they couldn’t share that data because it could not be deidentified. The reason things get a little easier in the field of research is because we just don’t want the data that identifies patients. We give pseudonyms and patient numbers, and only the investigator can connect that up with who the patient is.

You are involved with the Learning Health Community. What do you hope it accomplishes for improving care and research?

The whole idea of a Learning Health System is to take the data that we can from healthcare, learn from it, and then more quickly get that knowledge back into ways that people benefit from what was learned.

There was a prior paper that I helped write about the work being done by Dr Hidehisa Soejima at the Saiseikai Kumamoto Hospital in Japan. For 20 years, he’s been working on what he refers to as clinical pathways: healthcare protocols that outline, “This is the path you’re supposed to follow when you get a patient like this,” and then ask, “Did it work?” Every other month, he has a meeting where different teams working in his hospital share what they have learned from a clinical pathway and discuss whether the pathway should be changed based upon what has been learned. He’s done so much work that they’re actually collecting healthcare data in standard templates in the EHR and the NIH of Japan, AMED, is funding an expansion of this work. ◆

For more information, visit elligohealthresearch.com
What are the latest developments with Project Vulcan and FHIR?

The goal was to bring together the whole research community: large pharma companies, academic medical centers, government agencies, patient perspectives, and large tech companies, to work on problems. We start with why we need Vulcan; that’s because we fully need to integrate research into the delivery of healthcare by streamlining data collection and exchange into a singular process, which means that at the point of time that the clinician is collecting data, it should be collected in a way that could be reused for other purposes, including research. It’s not an additional thing for them to fit into their workflow, but it actually becomes part of a workflow. We need to have that bidirectional flow of data to and from the researchers and the clinicians, but also bidirectional to and from the patient.

In order to reach that vision, we’re collaborating with the international research community to align clinical data and clinical research data at the point of collection. We’re going to be doing this by developing out the HL7 FHIR standard to support this bidirectional flow of data.

The guiding principles that we use: Bridge existing gaps to connect industry collaboration. One of the key components of Vulcan is that we do not want to replicate anyone’s work. We’d rather partner with individuals’ work. If we strategically connect and we work together, then we can obviously maximize collective resources, because the workflow of research goes through many different hands and many different perspectives. We need to be able to hear all those voices at the table, so that whatever process is built out end to end, is really going to be a very streamlined process and not create a bottleneck in another location.

It becomes really important to have the international community. So what’s really exciting is that we have clinicians from Japan who are engaged; we have folks from Europe who are engaged and we’re getting other geographical areas becoming engaged.

This is really important because research is a global entity. We want to make sure that we all work together to come up with the best process that we can, while maintaining that compliance, regulatory guidance, and the geographical culture of any specific area.

And then the last thing is that if we all just talked about it, that would be nice but it wouldn’t really be helpful. And that’s where we need to deliver integrated tools and solutions. In January, we had our first workgroup meeting. And at that meeting, there was a connect-a-thon, a process that you go through, in order to iteratively try out the standard, find the gaps for it, and then take it back to the standards development process. You’ll keep doing this circular process until the standard is built out in a way that we know it can be used. It’s a live testing ground where people come together, and start exchanging data using that standard.

The data should move seamlessly across systems as appropriate.

What does Vulcan’s end goal look like on the ground?

In my opinion, it should be a bidirectional flow of data. So the data documented in the EHR for clinical care should be documented in a way that we can use it for clinical research. The other thing is clinically relevant data that’s being used in a clinical research setting: that should be accessible to the clinician in near real time, because it is data around the patient’s health.

And then there also has to have that third piece in there: it has to have bidirectional flow of data back to the patient. That includes the data we’re collecting, but also the end results of the study. Even if the study wasn’t successful, I would want to know how it turned out. For a lot of people, who have been in studies that have gone on to make huge changes in the way we’ve provided care. And they were part of the reason why.

Bringing Together the Entire Research Community to Solve Data Integration and Interoperability

Amy Cramer, MMCi, BSN, CPHQ, is a Director on the Global Product Development Strategic Team at Pfizer, Co-Chair of HL7 FHIR® Accelerator Vulcan, dedicated to translational and clinical research and a TransCelerate member. Her current focus is on clinical trials solutions, which includes interoperability of data and the bridging clinical care and clinical research.
Letting people who are participants know what the outcome of their research, in the most compliant way, is really important. That’s what we’re talking about when we talk about interoperable exchange of data. The data should move seamlessly across systems as appropriate.

We still have to keep all of the required safety, patient privacy, compliance, regulatory, data sharing and data usage agreements. It’s just that it shouldn’t be so onerous to, when appropriate, exchange it.

**What is the infrastructure that you think is needed to be able to pass data back and forth?**

In Vulcan, we talk about standard development. And part of development is how data is captured at the point of care. But mostly Vulcan is around how is the data exchange: terminologies and how to build out the resources to fit our use case, the domain analysis models, who the data has to exchange with, etc.

The infrastructure is more how to implement in the end. That will be part of Vulcan, and that’s where the connect-a-thons come into play. In the connect-a-thon, vendors actually come and they actually start exchanging the data. It’s all mock data, but they start exchanging following that standard.

So you find out, “If I only use those resources, it doesn’t really quite give me everything I need. I also need to know the patient’s XYZ. So I need to add that piece of information in for the data to make sense.”

This is the DIKW pyramid. At the bottom of it is data. I could say to you “21.” You don’t know what that is; you’re going to assume it’s probably “2021.” But what I was talking about was somebody’s respiratory rate. So the next layer of it is information. What we need to do is, once we exchange the data, make sure it’s coming across as information. And then the next layer is how to make knowledge out of that information. And then the final one is wisdom.

In my mind, that’s really what research is doing. It’s taking the data to information to knowledge so that we can get to wisdom and really create.

Part of Vulcan as we move forward will be these implementation projects. But we’ve started doing that work already through the connect-a-thons.

The vision is a process that’s developed out by all those individuals, because what other otherwise what will happen is we may have somebody who will say, “This is the way we want to exchange our data outside of our site firewall,” and then pharma will say, “That’s great, but we can’t ingest that data that way; our systems aren’t set up for it.” Or they say, “We need to accept the data into this system, because it is built to translate the data.” What we really need to work on is to come up with that, but without getting supporting any one way or commercialization.

The final project will be an implementation guide, but that has to come at the consensus of the entire community. It takes a little bit of time.

**Where are we in this journey of reaching what Vulcan sees as an endpoint?**

We’re always going to be finding innovative ways to do things better, right? But where can we get to a point where we can really start to interoperably exchange data? I think we’ve already started that.

We have several different organizations that are using FHIR to exchange data for research right now. Those are our trailblazers, and they’re on the cutting edge. And now we have this huge interest; there’s many organizations that are coming together really sending out this message of collaboration. Now we have so many other voices here with these early adopters, and we can start to look at how we can scale and adopt this.

And that’s where the challenge is going to be. At this point, the technology basics are there; they need to get developed out further, but there’s a process to do that. The next thing is consensus building and adoption. That’s going to take a little bit more time.

Some people say it takes about 5-10 years to develop things and get certain regulatory agencies to move forward. But people may not be aware of how long regulatory agencies have been investigating and actually funding projects that are dedicated to the use of HL7 FHIR for research.

My goal would be that, by this time next year, we have the capability to exchange data from at minimum, EHR to a pharma company, that is scalable and reproducible and that we have more than one vendor or way to do this. We need to have either one that’s completely free or we need to have different companies competing in the market.

**Is there anything that you’d be curious to know from another person about how they’re doing this?**

Yes: What are their plans on scaling this and how are they overcoming the learning curve and encouraging adoption without getting themselves into a situation where they’re only working with an individual company? Because we need to have that open market in order to be successful in this.

And scaling into different companies, vendors, partners, countries, but also additional data domains. So most people who are currently using FHIR for research are able to do the same data domain. But what data domains are they scaling to?

My goal would be that, by this time next year, we have the capability to exchange data from at minimum, EHR to a pharma company, that is scalable and reproducible and that we have more than one vendor or way to do this.
Taking on the Monumental Task of Forming a National Data Collaborative to Better Treat COVID Patients

Melissa Haendel, PhD, is the Chief Research Informatics Officer at the University of Colorado Anschutz Medical Campus. Dr Haendel worked with Kenneth Gersing, MD, NIH, and Christopher Chute, MD, DrPH, Johns Hopkins University, on the National COVID Cohort Collaborative, a data collaborative formed to use COVID-19 clinical data to answer critical research questions and address the pandemic. One of the things that we’ve noted is that even within a given model, different sites are populating those models differently. We found a lot of different data quality issues that were revealed by aggregating the data, that you didn’t really know was really such a problem in the distributed context because you never see the data as a whole like this. In that way, everybody’s data has been getting better.

The other piece of this in terms of regulatory oversight, we partnered strongly with NIH to figure out how to get the regulatory oversight to be able to have the institutions transfer their data, as well as to provide broad access with institutional data use agreements to anyone who has completed human subjects training and security training, and can be vouched for by their institution.

That goal was to support anybody who might be an expert in machine learning, but knows nothing about clinical data, or clinical experts working together. It takes a village to analyze these data. You need clinical expertise, statistics, machine learning, data modeling, clinical data informatics, etc. We really worked hard on the regulatory components to make sure that we could provide broad access.

The third part was that NIH had an already instantiated FedRAMP-certified secure data enclave where we could push the data in the early part of the pandemic. It takes almost two years to get FedRAMP certification for the security review. If we hadn’t had the right analytical platform already in place with FedRAMP certification, we would never have been able to create the enclave. We’re up to 82 institutions now that are pushing data. It’s broader and bigger than the CTSA community in which it started, and includes IDeA-CTR organizations and other local and regional health care organizations such as OCHIN. This is wonderful as it ensures that the cohort is demographically diverse and representative of our nation.

Part of your recent work was taking on the monumental project of forming the National COVID Cohort Collaborative. Can you describe the process of pulling those sites into a common infrastructure?

I am the contact principal investigator for a coordinating center called the Center for Data to Health, which is the coordinating center for informatics across the 60 Clinical and Translational Science Award sites. The goal of that coordinating center is to try to advance data sharing and interoperability and collaboration. It’s no small task. We had been marching along on our various paths to trying to achieve some of those goals. But when the pandemic hit, we asked ourselves what we could do that would really help the nation address the pandemic that is different and complementary to what already exists. We didn’t want to duplicate effort; we were trying to make the most of all of our person power across the nation as best we could.

There are a number of distributed research networks that use different common data models such as PCORnet and OHDSI. They each have helped institutions put their data into a data model structure locally, behind their clinical firewalls. In a distributed fashion, you can ask questions across them all. That way, the data stays behind the firewall, but you can still ask questions across a large group of sites. Some of these groups have been very successful at doing that, so we didn’t want to duplicate that. The issue though is in that scenario, you know what questions you want to ask and that you can answer. So it’s much less able to do discovery-oriented questions. If I want to ask about how many patients with condition “Y” are on drug “X”, I can do that. But if I want to find positive and negative correlations between all the drugs and all the conditions, it’s much more challenging to do in a distributed fashion.

Therefore, we partnered with the four primary research networks. In that way, we aligned the different common data models and allowed push of data from any site that has their data in any of those models. It democratized the models, so that any institution could participate if they had their data in one of those models.

Doing better research studies is one of the key ways that data can be used effectively.
You describe your vision as weaving together healthcare systems, basic science research and patient-generated data. How does informatics bring clinical research closer to clinical care?

A lot of my work is in the rare disease space where we really have a requirement to engage patients in their own care and in the research. They’re very enthusiastic participants. That’s one of the one most wonderful things about the rare disease community. A big component is the way that rare disease patients are really engaged in the research both on the clinical side and the basic research side.

In the context of the N3C, we are working with a patient group that has just been a really great emerging partnership. They have created us a survey and advocacy for “long COVID” patients. And we’re going to be working with them to pull in patient-reported information alongside the EHR data, as well as other kinds of data types, such as wearables, that would come directly from the patients. We’re including them in our analyses of the data and thinking about what their needs are in terms of having information come back to them.

One of the hopes of the N3C is to say, “We pulled our data and we found these results or analyses,” or “We’ve made these tools or predictive algorithms. Let’s deliver that back to the clinical site so they can use it.” If it means that “X” drug exacerbates COVID outcomes, stop using that drug.

The same thing is true for the patients; what the patients get out of it. We want to make sure as they’re reporting their symptoms and interacting with systems that help support them, that they get information back that helps them think about how they fit into the context of the greater data that’s being collected as well.

For example, when you fill out a survey, and it says, “75% of users answered the same way as you.” There could be something like that, which helps the patients understand how common or how unusual their progression is versus the other patients in the system. Especially if they have an unusual presentation, to understand what to do about it, and how to make sure that’s logged, that people get back to you and that there’s a strategy for understanding that.

We believe these types of interactions can lead to a much richer dataset for us to integrate with the clinical data, to provide a much more complete picture of the patient that we can analyze to help better classify patients into categories. That can help us better decide what their care and treatment should be.

Could it be used in a post-COVID context?

That’s something we’ve been working on trying to figure that out. We do believe that the structure of doing this has been revolutionary for informatics across the US. It’s the largest publicly available limited dataset in US history. It’s a phenomenal degree of partnership and sharing across different communities and institutions. The infrastructure for the harmonization has been built in rapid fashion with the partnership of those common data model research communities. That could readily scale to doing all EHR data and all disease areas, but the regulatory requirements for that would be different. This data was transferred with the sole exclusive purpose of being used for COVID. As long as you’re studying COVID in some way, the access is pretty straightforward.

We do believe that this could scale for many different things. There are groups applying for grants to see if they can use the infrastructure that has been built, but for different disease areas and with different regulatory agreements. Simultaneously, conversations are ongoing across NIH and across the institutions about how wanting to grow this capacity for the ability to be used by the community for all diseases.

That’s where we would like to head but there are a few regulatory barriers. In addition to that, will people feel the same about data-sharing when the pandemic is over? We hope that this has changed the culture, and that we demonstrated security and safety of sharing data in this way, so that people will feel more confident in doing that more broadly for future research.

What guidelines would you like to see in order to make full use of the data we are generating, including from newer sources of data?

Two comments. One has to do with how we design trials and studies. For COVID research, or any new disease, we often start without a good definition of what constitutes somebody with that disease. “Who has long COVID?” is our question of the day, for example. This is where that patient-matching classification comes in. There isn’t just one long COVID patient, and they’re not all necessarily suitable for the same research study.

Research studies are expensive and time-consuming; we want to make the most use of those resources. We need to be using observational data from the EHR, but also from the patients, whether it’s wearables or patient surveys, to better inform how to classify patients into different subgroups that can then be assigned to the right research study. Doing better research studies is one of the key ways that data can be used effectively.

The other thing is a long-term strategy of team science around translational analytics or collaborative analytics. You cannot find results from these data as one person. You have to have very diverse expertise working together in order to ask the hard questions, find answers and make sure that they’re robust, clinically and in terms of data science. We’ve been working on coming up with “domain teams” to help support people, projects and teams to make sure that they have the expertise they need to do good science and to do efficient science. Make science go faster, better.

You cannot find results from these data as one person. You have to have very diverse expertise working together in order to ask the hard questions, find answers and make sure that they’re robust.
How a Pediatric Oncologist is Tackling Data Standardization as a Grassroots Phenomenon

Samuel Volchenboum, MD, PhD, is a pediatric oncologist and biomedical informaticist at the University of Chicago. The Volchenboum Lab is housed within the University of Chicago Department of Pediatrics. Their flagship project, the Pediatric Cancer Data Commons (PCDC), began as an initiative of the University of Chicago Center for Research Informatics.

What is the Pediatric Cancer Data Commons and what does it hope to accomplish?

I trained as a pediatric oncologist, and after that, pursued a fellowship in clinical informatics and obtained my Masters in Biomedical Informatics. Early on in my career, I started to note all the inefficiencies inherent to the treatment and clinical trials processes. As I kept looking at the system, it became more and more clear that we were really just suffering from this pervasive lack of interoperability.

One of the main causes of this lack of interoperability is a lack of data standardization. We have a system where if you go from one hospital to another hospital across the street, it’s still very likely that they are going to have trouble using your data. A lot of medical data are still transferred is on CDs and via fax machine.

It really drove my professional direction towards trying to build systems that allowed us to share data better and to really tackle the problems at the root. We didn’t want to do the easy thing, which is just take data and then hire people to standardize and then share it. We wanted to tackle it as a grassroots phenomenon: teach people what standards are, empower them to build a data dictionary that’s validated with international input, and then let people harmonize the data into that standard and, ultimately, use that standard to collect subsequent data.

Around 2014, Dr Susan Cohn, a neuroblastoma expert at UChicago Medicine, came to me and said, “We have this collection of about 10,000 patients of neuroblastoma, which is a very rare pediatric tumor. It’s in a big Excel sheet. Can you help us figure out how to share it?”

We set up an infrastructure to take these data that was somewhat standardized, and started working with this neuroblastoma group to develop rules around how we would share it and how researchers could use it. And based on that success, the rhabdomyosarcoma group came to us to build their data commons. We started from the ground with them; we went to get collaborators in Europe and in the US. We created a big consortium, and set up rules of engagement. It took us a year and a half to build a data dictionary, painstakingly negotiating every single value and element in the data dictionary to come to a consensus.

And then after that, we started to receive more funding to support our work with more disease groups. Now we’re covering just about every pediatric cancer and are in various stages of building data dictionaries and harmonizing the data. We are tackling the problem at its source and providing a great way to share data and lower barriers to research. We hope to move beyond pediatric cancer into other areas, and we hope to empower people to do the right things with data.

We wanted to tackle it as a grassroots phenomenon: teach people what standards are, empower them to build a data dictionary that the world agrees upon and then let people harmonize the data into that standard and then collect it.

What has this work elucidated for you about current data capture?

People are always going to put data in the way that’s the easiest possible. If you offer people a blank textbox, they are going to use it. If you have to enter 10 things about a patient - their performance status, height, weight, blood pressure, etc., and you’re faced with 10 checkboxes, or a big empty text box, you’re probably going to just fill that box with text that will then be very hard to study.

The current methods of data capture could cater very well to structured capture, but they allow easy outs. That may serve the purpose of billing, but it ignores the really great things you can accomplish with the EHR - serving the research needs and performing rules and evidence-based medicine. You lose all those things when you don’t capture the data in a granular, structured form.
What are the challenges that you’re coming across when creating a common data language?

The biggest problem is that, up until now, most groups have been left to their own devices in creating their own data collection forms. The data elements we see are not clearly harmonized. For example, in the USA, one group might have, for site of disease, “lips, cheek or face.” But the group in Europe might just have “head.” You have to sit there and have a scientific discussion, and they have to come to an agreement. So that’s one of the biggest challenges: trying to figure out how to let science drive the building of the dictionary. Then harmonizing the data is just a technical issue. There are cultural issues, language issues, issues around which coding systems do we use, etc. We’ve been able to tackle these by getting buy in from the right kinds of clinical leaders across the different countries.

How does someone enter into the PCDC and start contributing data?

We’ve taken a very UN approach to this. The different disease commons are run by disease consortia that we’ve helped set up for each disease group. So, there’s a soft tissue sarcoma consortium, an AML consortium, a neuroblastoma consortium. For example, there’s a group in South America that wants to join our retinoblastoma group. We invite them to meetings, and then the executive committee will have to work with them to decide if they want to come in as members and contribute data and what those data look like.

One of our big missions over the next couple of years is to move very decidedly into less developed areas of the world where most pediatric cancer occurs and try to bring this same idea of standardized data collection to help improve care. To do that, we’re obviously going to need to try to bring in many more groups. Right now, we’ve mainly engaged groups in North America and Western Europe, but we need to move beyond that to get all these other groups in. Defining the governance and rules of engagement are going to be really important.

How does this impact or affect the type of care that clinicians are able to offer their patients?

The goal of our work is to make research data more available to clinicians to try to lower the barriers to research. If you look at the neuroblastoma group, they’ve used the data for all sorts of projects that have changed the way we diagnosed and treat children with neuroblastoma. At one point, the cutoff for age for neuroblastoma risk was one year. Based on the data in the data commons, they did a study and changed it 18 months. That immediately affects a large number of patients and the kinds of chemotherapy they will get.

One of more exciting ways we’re working now is to create better ways to match patients to precision medicine trials. We are working with The Leukemia & Lymphoma Society to create a decision support tool that is going to allow clinicians to go to the tool, enter information about their patient – clinical, genomic, immuno-phenotype – and match them to clinical trials. The goal of the Pediatric Acute Leukemia (PedAL) project is to have a child with relapsed leukemia go onto a clinical trial within 72 hours. We’re hoping to release the first generation of the tool later this year.

One of the big frontiers we’re going to cross next is how to get data right out of the EHR in real time to help with patient care. Because right now, by the time you collect data into a research commons, it’s old; the patient is diagnosed or already being treated. But how can we collect data in real time and use it for real time decision making for patients?

What is coming down the pipeline for the Pediatric Cancer Data Commons and your work?

We have made the determination only to work with de-identified data for our data commons. We remove the dates and instead have the age in days of the patient at the time of whatever event happened. We may not know that a kid had a CT scan on July 1, but we’ll know that they were 56 days old at the time. But just because the data are de-identified, that does not change at all the fact that we operate in a fully HIPAA-compliant infrastructure. We have been successful in implementing GDPR regulation throughout our data commons. But that means that for every single site that we’re sharing data with, we are negotiating a data-sharing agreement.

Even though the data are de-identified, the GDPR is extremely strict about holding data. There is the concept of “the right to forget,” meaning that if a patient in the EU decides they don’t want their data as part of the commons anymore, even if it’s de-identified, we have to be able to show that we can remove their data from the commons.

What are the security aspects of developing an international data commons platform?

We have made the determination only to work with de-identified data for our data commons. We remove the dates and instead have the age in days of the patient at the time of whatever event happened. We may not know that a kid had a CT scan on July 1, but we’ll know that they were 56 days old at the time. But just because the data are de-identified, that does not change at all the fact that we operate in a fully HIPAA-compliant infrastructure. We have been successful in implementing GDPR regulation throughout our data commons. But that means that for every single site that we’re sharing data with, we are negotiating a data-sharing agreement.

Even though the data are de-identified, the GDPR is extremely strict about holding data. There is the concept of “the right to forget,” meaning that if a patient in the EU decides they don’t want their data as part of the commons anymore, even if it’s de-identified, we have to be able to show that we can remove their data from the commons.
The Power of Learning Health Systems for Sharing Data to Connect Care and Research

Joshua C Rubin, JD, MBA, MPH, MPP, is Program Officer for Learning Health System Initiatives at the University of Michigan Medical School Department of Learning Health Sciences, and Executive Director of the Joseph H. Kanter Family Foundation, the only philanthropic foundation founded by a patient whose overarching mission is to realize a patient empowering Learning Health System vision. He also serves pro bono as Founding President and CEO of the Learning Health Community.

Can you describe the learning health system model and your work with it?

These are health systems that have a cultural commitment to continuous improvement as a result of every interaction. So as opposed to telling every patient “Take two and call me in the morning,” it’s much more about how to keep learning from every experience of every patient. The fundamental idea is, “How do you learn from every care experience and every other relevant health experience and use that to improve health?” It becomes an iterative cycle of trying to improve. Among other things, it can empower clinicians – in collaboration with patients and in consideration of their values and preferences – to really tailor treatments and even dosages to what is likely to work best for each patient. No more “one size fits all”.

Netflix will try to keep getting better about figuring out what movies you like, based on what it learns about you and when it learns about other people. Can we do something like that in a much bigger way in healthcare and do it in a way that is not so much of a power asymmetry? Can we do it by having the patients who want it onboard and having the systems all doing it together as a movement? And can we do so in ways that – beyond merely respecting patients’ privacy – actually is driven by patients for patients?

Part of my role now is running the Joseph H. Kanter Family Foundation, the mission of which was to create a world in which every decision affecting health is informed by actionable knowledge of what works best. When Joe Kanter was diagnosed with prostate cancer, he wanted to know not just the doctor’s experiences, but for every other patient like him, what treatment did they have and what were the outcomes?

I also work at the University of Michigan Medical School Department of Learning Health Sciences, which is built around bringing together scientists from seemingly divergent disciplines – social science, biomedical science, data science, computational sciences, implementation sciences – and get them to all work together to solve this problem.

I also serve pro bono as the Founding President and CEO of the Learning Health Community, the goal of which is to build a movement around this change, and not view it as a technology project or as a healthcare project, per se. The whole movement is anchored in a shared set of consensus core values that embody a shared vision for a future of health we can only realize together.

There are a lot of folks that think about how to manage, store and distribute lots and lots of data; and how to share data and make it accessible. I think the question gets more interesting when you ask how to also manage the knowledge that was generated from all this data. How do you get the knowledge back to where it needs to go to inform the decision? And how do we get it there in ways that are up to date, that are tailored for the person or people receiving it, and that are readily actionable by them?

How do you learn from every care experience and every other relevant health experience and use that to improve health?

At CRAACO 2020, you discussed the need for permanent infrastructure. Can you expand on that?

It’s everyone in the same ecosystem, and all working together. Part of the infrastructure becomes the experience of communicating and working with one another. These people are experts in their own disciplines; they have their own jargon, their own ways of measuring success and their own way of thinking about things; in essence not even speaking a common language. How do you begin getting everyone to work together? Start with a shared passion for improving things.

This all just becomes a cultural shift. All these people were trained at institutions where the medical school, research, public health, and nursing are all different buildings with different walls.
The silos persisted after that. It’s a real paradigm transformation into thinking about research and clinical care being one in the same instead of functions of different units.

You have to look at how many times the same problem keeps getting solved, or not solved, as the case may be. The infrastructure you build once and use it many times. If you’re trying to get a bunch of people across back and forth across a river, eventually you build a bridge. But that often isn’t the way things are funded or thought about. You’ll put together a research project to learn one thing; you’ll build the pieces of that infrastructure as you go and dismantle it when you finish. That has to change.

**Where are we now in the pursuit of maturing learning health systems?**

In the early 2010s, this was an idea that a handful of people were talking about it. A lot of work began in 2007, the Institute of Medicine, now the National Academy of Medicine, launched one report about it that eventually became a series of reports and brought some very bright people together to think about it over and over again. On the technology side, with the Great Recession of 2008, and the stock market and the housing bubble, in the Recovery Act there was a massive investment in taking health records that were on paper and digitizing them through the HITECH Act and incentivizing what was once termed “meaningful use” of health IT.

That was definitely building a piece of the infrastructure, and it led to the conversation about, “We’re spending all this money on electronic health records, what comes next? How do we build on all this?” This has evolved from an idea to something that is really starting to take shape. Part of what’s next, and this is what the Learning Health Community has been working on is trying to think through, is what would a maturity model look like? So if Cleveland Clinic or Mayo Clinic says, “We’re a learning health system,” can you rate them on a scale of one to five on how much of a learning health system they are? What are some of the metrics of maturity in this?

And if the need for infrastructure and the need for learning should have been apparent, it’s from what we’ve been through with COVID-19. How do we build the infrastructure if something like the COVID-19 pandemic happens again? How do we identify that it’s happening and how do we stop it quicker? If patients in this hospital are getting this treatment and patients in another hospital are getting a different treatment, how do we start to figure out what’s working (and for whom) and have enough data on the patients to compare?

**We need to build this future in ways that protect patient privacy and autonomy, and empower patients to be active participants in enabling continuous learning.**

**What next steps would you like to see being taken?**

I’m really into the idea of democratizing the knowledge that’s out there. A lot of knowledge gets produced and goes into places that are not accessible – or at least not accessible to all. It’s really about making the knowledge accessible and usable by people, and by machines and tools that can help people get to it and to use it effectively.

There also needs to be a culture shift. If knowledge produced in Michigan gets used somewhere in California, it’s not going to harm Michigan’s bottom line. In fact, collaboration would expand access to data and to people who could contribute their experiences and knowledge to shared learning on a broader scale and scope; everyone could get a better return on what they invest into the collaborative learning effort.

It’s really about finding ways that we’re all in this together and being much more open about sharing. I don’t believe it has to be idealistic; there are mechanisms that can be developed that will allow the producer of the knowledge to get credit for or keep track of it, while also getting it into more hands and more usable forms.

In the patient community, there’s a motto about letting patients help. There should also be a motto of letting the machines help. Machine helping is not a replacement for expertise, but it’s a good complement to it. It is not about machines replacing humans; it’s about empowering people to do even more than they could.

**What would you like to see more of in regards to community-building and data-sharing in the future?**

I’ll build on my democratization point and say that I do feel like we’re at a crossroads where things can end up either open and accessible in lots of different hands, different places, and different opportunities, or things can end up much more monolithic. I don’t think we want one or two companies being dominant players; I don’t think we want knowledge that we want to create (or perpetuate) a “haves” and “have nots” vis-à-vis knowledge that can save lives. I think we want this to be very open.

We have this great opportunity to do something. Hopefully, we can and will. But there is this risk of actually going the other way, of all the great things we’re building ending up in proprietary systems that are accessible to the wealthiest health systems and therefore end up in the hands of the few privileged patients. We have to be very cognizant of that and make sure we don’t build more walls as we’re building this transformative future of health together.

There is little question that the future of research and care will be informed by lots of data from lots of people and likely exponential growth of biomedical knowledge. Key principles of learning health systems will shape this future of health. We need to build this future in ways that protect patient privacy and autonomy, and empower patients to be active participants in enabling continuous learning for the sake of improving their own health, the health of their loved ones, and the public’s health. ◆
Upcoming Conferences

8th Annual **Patients as Partners**
April 7-9 | Virtual
3-Day Festival of Ideas from Patients, Pharma and FDA on How Patient Engagement and Involvement Gets Done.

PatientsAsPartners.org

6th Annual **CRAACO**
Clinical Research As A Care Option
April 26-27 | Virtual
Integrating clinical care and clinical research so more patients can participate for better outcomes for all.

CRAACOevent.com

11th Annual **DPHARM**
Disruptive Innovations
September 28-29
Reporting on decentralized clinical trials and disruptive thinking to reduce the burden of participation in clinical trials.

DPHARMconference.com