



Data Management for Precision Medicine and Genomics in 2017

Research supported by PHEMI



Data Management for Precision Medicine and Genomics in 2017

- 1 Introduction
- 2 Background
- 4 Highlights
- 5 Findings
 - 5 Investment
 - 6 Use Cases
 - 8 Data Storage Needs
 - 8 New Technologies
 - 9 Real-Time Clinical Support
 - 9 Risk, Privacy, and Security
 - 10 Data De-Identification
 - 11 Data Acquisition
 - 11 Data Management
 - 12 Unstructured and Non-Clinical Data
 - 12 Skilled Workforce
 - 13 Better Tools
 - 13 Interoperability
 - 13 Data Sharing
 - 14 Participation in Clinical Trials
 - 14 Patient Registries
 - 15 Extracting Value from Metadata
 - 16 Aging Populations
 - 16 Common Data Definitions
 - 16 From Competition to Collaboration
- 17 Looking Forward: Capabilities to Advance
- 20 Research Methodology
- 20 Participating Organizations
- 21 References

Introduction

This report provides an environmental analysis of the current precision medicine and genomics market in the United States healthcare industry. According to the National Institutes of Health (NIH), precision medicine is “an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.”¹ Precision medicine goes hand in hand with “genomics” research. Genomic information can be utilized to help diagnose and treat a person’s disease. In the last several years, many provider organizations have launched new initiatives in precision medicine and genomics in an effort to improve patient outcomes.

To better understand the current state of the market, the Foundation for eHealth Initiative conducted a series of interviews with executives at leading provider organizations across the country. The Foundation for eHealth Initiative (eHI) is a neutral, non-profit organization which convenes executives from multi-stakeholder groups to identify best practices to transform care through use of information and innovation. As such, eHI spearheaded an effort to interview and assess the state of the field. The findings shared in this report provide insight on how clinical and genetic data is used and managed, as well as the challenges providers face in genomics research and precision medicine.

1. [“What is Precision Medicine?”](#) National Institutes of Health

Background

Precision medicine is transforming the way diseases like cancer are treated by tailoring medical care to the person's genetic makeup, or the genetic profile of an individual's tumour. In precision medicine, the focus is on identifying the approaches that will be effective for patients based on genetics as well as environmental and lifestyle factors.

Genomic medicine is considered a subset of precision medicine. The National Human Genome Research Institute defines genomic medicine as “an emerging medical discipline that involves using the individual's genomic information as part of their clinical care (e.g., for diagnostic or therapeutic decision-making) and the health outcomes and policy implications of that clinical use.”²

A new federal program and funding has sparked additional interest in precision medicine. In 2015, President Obama announced the launch of the Precision Medicine Initiative (PMI) called “All of Us”, which is an innovative genomics research effort focused on developing specific treatment interventions that are tailored to the individual patient.³ This \$200 million⁴ initiative will leverage genome sequencing technology, big data, and clinical trials to provide scientific evidence to move the concept of precision medicine into clinical practice and create a single nationwide system that provides the full picture of a human life. The long-term goal of this initiative is to create a research cohort of at least one million American volunteers who will share genetic data, biological samples, and diet/lifestyle information over many years to help researchers develop individualized treatments to:

- Advance pharmacogenomics, in order to provide the right drug for the right patient at the right dose
- Identify new targets for treatment and prevention
- Test whether mobile devices can encourage healthy behaviors
- Lay a scientific foundation for precision medicine for many diseases

NIH has awarded contracts to 14 healthcare organizations to lead specific research programs related to the PMI. The data center will be run jointly by Vanderbilt University in Nashville, Verily (formerly Google Life Sciences), and the Broad Institute in Cambridge, Massachusetts. The Mayo Clinic in Rochester, Minnesota has been contracted to maintain the study's biobank

The central finding in this report is that all of the organizations interviewed want to leverage clinical and genomics data to strengthen their core competency of caring for the patient.

2. [“What is Genomic Medicine?”](#) National Institutes of Health

3. [“The Precision Medicine Initiative”](#) White House

4. [“FACT SHEET: Obama Administration Announces Key Actions to Accelerate Precision Medicine Initiative”](#) White House Briefing Room

of blood, urine, and other samples.⁵ Both Vanderbilt University and the Mayo Clinic were interviewed for this report.

In addition to federal initiatives, many providers in the private sector have also launched precision medicine and genomics research studies. For this study, we interviewed several organizations who have announced programs. As the healthcare industry has been moving towards patient-centered care for years, the precision medicine model will help providers to achieve their vision of delivering the right care to the right person at the right time. Initiatives funded by both the federal government and private sector are advancing this vision.

5. [“President Obama’s 1-million-person health study kicks off with five recruitment centers”](#) Kaiser, K., (2016)

Highlights

The central finding in this report is that all of the organizations interviewed want to leverage clinical and genomics data to strengthen their core competency of caring for the patient.

Highlights of key trends and challenges are listed below. For details, see [Findings](#).

1. Investments in precision medicine programs are still evolving.
2. Use cases are diverse; pioneering providers are finding new ways to derive meaning and value from genomic data to drive clinical benefits.
3. Data storage needs are evolving.
4. Providers are adopting new technologies to handle new data demands; use of Hadoop is becoming common.
5. As a top priority, providers want to use data for real-time clinical support.
6. Risk management, privacy, and security are key considerations.
7. Providers are concerned about de-identification.
8. Acquiring data in a ready state for analysis (data acquisition) is a challenge.
9. Data management is a challenge for some organizations using big data.
10. Providers want to explore and mine unstructured and non-clinical data.
11. That there is a shortage of specialized skills, especially data science skills.
12. Providers need better tools to make the most use of big data.
13. Providers want their genomic data systems to be interoperable with their electronic medical records systems and other internal clinical systems.
14. Providers want to be able to share data, and this creates challenges for privacy and confidentiality.
15. Providers need increased participation in clinical trials.
16. Providers need support with managing patient registries.
17. Providers want to extract value from metadata.
18. Providers want to use precision medicine to meet the healthcare needs of an aging population.
19. There is a need for common data definitions.
20. Providers want to move from competing to collaborating.

Findings

Information gathered in interviews demonstrates a wide spectrum of programs and applications among providers. Some providers are ahead of the curve and are already finding ways to integrate genetic research into clinical use, whereas some providers are only now starting to consider precision medicine.

Investment

Investments in precision medicine programs are still evolving. However, the findings in this report show that all of providers interviewed to date are interested in incorporating precision medicine into their strategic plan to bring genomics to the point of care.

All of the providers interviewed are already invested and involved in genomics research. They are all seeking clinical use cases for precision medicine to realize the value of their genomic data.

The precision medicine market is still growing and providers are just starting to integrate genomics into clinical practice to inform medical decisions. Some leading providers are already engaged and invested in genomic research projects. However, the genomic market in general is in the germination phase and most providers are at the starting stages.

Interviews revealed that some executives have only now begun conversations of genomics research, whereas others have been involved with genomics for over fifteen years. For the purposes of this research, executives at organizations known to be involved in precision medicine were targeted. Therefore, our interviews focus on organizations who are early adopters.

However, it was clear from the interviews that the environment is ripe for more development in this area. The advent of modern technology, a regulatory push for better healthcare outcomes, and advancement in genomic medicine have ushered in a new era of precision medicine. Consumers want assurance that the healthcare they are receiving is the best treatment for their illness, and providers are seeking ways to improve clinical accuracy with real-time decision support.

Of the providers interviewed, just one was directly involved with the White House PMI. Most of the other providers interviewed have begun developing strategies specifically around precision medicine. All of the providers interviewed have a genomics research department. Providers understand that simply collecting genomic data is inadequate and that to derive value from it, they must turn the data into knowledge that informs clinical decisions and allows them to deliver personalized care.

With regard to the size of investment (cost), little information was shared by executives in this area. Executives were either not apprised, were not able to separate these costs from other expenses, or were not comfortable sharing.

Use Cases

Use cases are diverse. To turn aggregated data into knowledge that can inform clinical decisions requires integrating the results of analytics with human judgments of the meaning and significance of observed patterns. Synthesizing and deriving meaning from big data is a significant data science challenge. However, some pioneering providers have found a way to analyze their genomic data to derive clinical benefits.

Synthesizing and deriving meaning from big data is a significant data science challenge.

Some providers are eagerly embracing genomics research and have started aggregating terabytes of genomic data, while some are at the other end of the spectrum and have only just began the conversation of precision medicine. Providers are interested in finding ways to use technology to tap into genomics data for population health and patient engagement.

Regardless of the phase providers are in, the common theme is that providers want to specifically determine how they can use genomic data today to improve the lives of their patients. Although there have been successful use cases of the clinical use of genomic data, the value question for some is still unanswered. Providers want to be able to use this data for real-time clinical decision support and or research purposes to advance the science of genomic medicine.

Providers want to be able to use this data for real-time clinical decision support and or research purposes to advance the science of genomic medicine.

Fourteen such use cases are:

1. A leading provider has been collecting DNA, plasma, and serum samples from a population of over 40,000 patients for the past 15 years, with permission to use health records and genetic markers for research purposes. These investigators are undertaking dozens of studies to investigate and identify the genetic underpinnings of a variety of disorders. Their main challenge is that their physicians are not trained to interpret the genetic data.
2. A provider is using machine learning to look for patterns in the data, in order to make predictions about which patients should be prescribed which drugs. Since investigators know which genetic variants are associated with adverse events for certain drugs, they can test patients for the operative variants in advance, to inform downstream medication prescription decisions by clinicians.

3. Some providers are beginning to collect and monitor streaming data from medical devices and monitors. This new data use is creating a challenge, as providers work to aggregate huge amounts of data that must be analyzed and translated into knowledge that can have clinical application.
4. Some providers are analyzing their data for reporting and benchmarking purposes to drive their business intelligence use case. They are still working to understand how they can meaningfully use their data, and they are interested in contracting with other third parties for predictive analytics.
5. Some providers are using their genomic data for cancer research using predictive analytic models to proactively determine patients likely to develop cancer. This allows treatment interventions and lifestyle changes to be initiated early to deter the onset of the disease.
6. One provider has discovered many of the genetics risk factors for age-related macular degeneration. These investigators can hypothesize the relationship between a disease and a drug, to predict who is likely to get age-related macular degeneration and then prescribe the drug that prevents it.
7. Some providers are invested in pharmacogenetics, and are looking for novel adverse drug events in patients so that they can infer associations among metabolic pathways, drugs, and acquired disorders. Like the previously mentioned provider, these investigators are hypothesizing the relationship between a disease and a drug to predict who is likely to get a disease and treat accordingly.
8. Some providers are using genomic data not to provide clinical support but rather to provide support to gene annotators. These investigators are providing intermediate decision support: if an annotator needs to annotate a specific gene and find a list of drugs that target the gene, the list of drugs will be available with the annotation, along with citations to the literature.
9. In research, some providers are computing very large datasets, generating, storing, and analyzing genetic sequencing data with the goal of producing a list of genetic variances that may have clinical meaning in the context of a clinical trial or a new discovery.
10. Some providers are involved in precision oncology research, providing decision support services to help oncologists interpret genomic testing results. When provided with a gene, the result is a list of drugs that target the gene.
11. One provider was selected by NIH to launch the first phase of the Precision Medicine Initiative Cohort, which will lay the foundation for a national community of one million or more U.S. volunteers, partnering with researchers, sharing data, and engaging in research to transform our understanding of health and disease through precision medicine.

12. One provider is launching a new Center for Data-Driven Discovery in Biomedicine (D3b). The Center will develop a multi-institutional biorepository infrastructure and data-analysis ecosystem that supports patient participation in collaborative tissue-based research for pediatric cancer and a shared, open-data infrastructure that will drive precision medicine discovery. These platforms will help doctors match cancer therapies to individual children and define new precision medicine approaches for pediatric cancer and childhood diseases.
13. One provider created a cancer program called the Moon Shots program, launched in 2012. The program has more than 165 immunotherapy clinical trials underway with 3,500 enrolled patients. The program is missioned to develop new drugs, discover new ways to prevent cancer, and find new ways of sharing knowledge. The new GOP Administration is committed to this Moon Shots program not only in the space of cancer research but to impact diseases of aging, including Alzheimer's.
14. In addition to genetic sequence technology, one provider is exploring immunotherapy at proteomic level. This provider is using precision medicine approaches to attack diseases with chemotherapy, based on data about how the cancer has historically behaved. The provider noted, "We know that a successful survivor can come back with cancer. We've learned that if you check that cancer later on, it may have mutated into another type of cancer altogether. This mutation is at the protein level. Hence, there is need for more precise treatment interventions for these cases."

Data Storage Needs

Data storage needs are evolving. As providers are beginning to generate datasets which are enormous in size, data storage can become an issue, and data storage requirements are growing for groups engaged in this work.

Currently, all interviewed providers store their data in-house. In addition to in-house data storage, few providers are utilizing external storage repositories such as the cloud or a supercomputing resource such as the Texas Advanced Computing Center (TACC). The general trend is that as the size of the data grows, providers are moving away from using local servers to using various cloud resources.

All providers indicated that they see future need to outsource data storage and focus on their core competency of providing clinical care. The challenge for providers, however, is to understand how to utilize available external resources to realize maximum value.

New Technologies

Providers are adopting new technologies to handle new data demands, and the use of Hadoop technologies is becoming common as data volumes soar and scalable storage and compute are required. Many of the providers we interviewed had adopted Hadoop for both clinical and

research purposes to handle the data demands of precision medicine. In one instance, Hadoop was described as a “decision support tool”

Real-Time Clinical Support

Real-time clinical support is a top priority for providers. Ultimately, providers want to understand how to translate genomic data to meaningful forms that provide real-time clinical support, and for these providers, the ability to use data for real-time clinical support is a top priority. Currently, much of the data is not used real-time

Risk, Privacy, and Security

Risk management, privacy, and security are key considerations. Data utilized for precision medicine is extremely valuable and contains sensitive patient information, but privacy and security concerns have yet to be specifically addressed for genomics and precision medicine. Currently, providers are relying on their existing internal privacy and security infrastructure to protect genomics data. Providers are interested in developing a risk management approach.

Although all providers acknowledged the privacy and security risks of storing, sharing and using genomic data, not much additional investment has been put towards this component of data management. One provider stated that “the privacy and security of our clinical genetics is handled like a lot of our health data, so there are very strict rules about accessing it; it’s not necessarily treated differently. With pharmacogenetics, it doesn’t have any special sensitivities for privacy and security; however, on the research side there is a lot of concern about privacy and security, but so far researchers nationally have been careful to maintain that. The research side seems to have avoided some of the breaches that have happened on the clinical side. From my point of view we’re doing okay, although there is clearly risk there.”

Another provider agreed, saying that “the information security problem is a lot of blocking and tackling of the right controls and processes. I think the challenge for all healthcare organizations is we need additional best practices and frameworks that lead to better security posture and reduce risks.”

A third provider has adopted an enterprise risk management approach to addressing privacy and security concerns. This provider see the problem as a governance issue and personnel are spending a lot of time to change their governance practice to include security. The provider stated, “It goes beyond a legal question to become a business risk decision or an enterprise risk decision.”

Providers want to be proactive in safeguarding genomic datasets and invest in the adequate security infrastructure to protect this sensitive information. Most providers are relying on their existing security measures that protect patient information, but given the recent peak in security

breaches, this may not be adequate. Genomic data needs special privacy standards and security measures to protect personal information and the integrity of research initiatives.

Data De-Identification

Related to privacy and confidentiality is de-identification, providers are concerned about de-identification. De-identification is the process of removing or obscuring any personally identifiable information from patient records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them.⁶

Personally identifiable information (PII) must be adequately protected at all times, and de-identification is a primary mechanism for doing so. Providers are concerned that current systems do not adequately de-identify genomic data when used for research purposes, and they also worry about the future control and use of the data.

Providers acknowledge that, in practice, the acquired datasets are not always de-identified. Most providers expressed skepticism when asked if their genomic data are properly de-identified to the point where the risk of unintended disclosure of the identity of individuals is negligible. It also became apparent that providers are not requesting de-identified datasets for research purposes. On the contrary, one provider noted that researchers are usually able to simply pull data from the Electronic Health Record, and in fact are not always limited to records with prior obtained consent.

In discussing the challenges with de-identification, providers revealed that part of the problem with mechanisms like pseudo-identifiers and other common methods of de-identification some research projects require investigators to follow up with the patient and assess how they are doing over time.

For genomic data in particular, providers acknowledge that they are not taking appropriate steps to de-identify information to protect the confidentiality of the individuals. This is because genomics data is commonly used for research purposes and in these cases patients generally grant consent to the use of their bioinformation. Even so, these practices create confidentiality liabilities and increase the vulnerability to potential privacy breaches and the indications are that providers need additional resources and support to adequately de-identify their genomic datasets.

While it may not be possible to remove the disclosure risk completely, de-identification is considered successful when there is no reasonable basis to believe that the remaining information in the records can be used to identify an individual. To ensure successful data protection, it is essential that techniques are appropriate for the intended purpose and that their application follows the best practices.

6. "[Data De-identification: An Overview of Basic Terms](#)" Privacy Technical Assistance Center

Data Acquisition

Acquiring data in a ready state for analysis (data acquisition) is a challenge. Providers are challenged to gather, filter, clean, and generally move data into a ready state for analysis before the data is put in a data warehouse or any other storage solution.⁷ The interviews revealed that providers feel that they are not acquiring data that is in a ready state for analysis.

One provider observed that “if you have a well-established and well-structured dataset you don’t necessarily need a lot of big data technologies.” The challenge providers are having is to acquire data “in a way that is easy to access in a data-science perspective. This is not just about normalizing to common clinical definitions; rather, it is about being able to handle very high transaction volumes of data and structure it in a way that can be accessed by the end-user for clinical support.”

The acquisition of big data used to be commonly characterized by “the three Vs”: volume, velocity, and variety. While these three remain demonstrably important, providers are starting to focus on a fourth “V”: value. Providers, therefore, want adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value forms of the data are actually processed and analyzed. They also want to be able to structure their data in a way that can be easily translated for clinical use.

To improve acquisition of high-value data, providers need to first assess the current requirements for data acquisition, and second, to unveil the current approaches used for data acquisition and third, to evaluate future developments in data acquisition. Solutions that help providers optimize the value of their genomic data will be beneficial.

Data Management

Data management is a challenge for some organizations using big data. In addition to challenges around data acquisition, providers have unmet needs around processes for validating, storing, protecting, and processing the data.

Providers want processed data that is accessible, reliable, and timely, to be used at the point of care. They are looking for data management systems that allow them to manage and scale with commercially robust levels of infrastructure. One provider specifically mentioned data management needs for documents and clinical notes to extract or search those documents efficiently for real-time point-of-care use cases.

In addition to challenges around data acquisition, providers have unmet needs around processes for validating, storing, protecting, and processing the data.

7. “[Big Data Acquisition](#)” Lyko, K., Nitzschke, M., & Ngomo, A. C. N. (2016)

Unstructured and Non-Clinical Data

Providers want to explore and mine unstructured and non-clinical data. Physicians noted that mainly discrete data are hosted in their warehouses and they are not yet exploring tools that can mine unstructured data such as physician notes, images, etc. Unstructured data presents a new set of opportunities; for example, sophisticated natural language processing algorithms can be used to mine call transcripts can now “read” physicians’ and nurses’ notes.⁸

Providers also understood that big data includes collecting non-clinical and unstructured data to get a complete picture of one’s health. They are interested in capturing patient-generated health data from wearables and connected devices to get a more holistic view of the patient. Recent trends in the industry show that there has been significant growth in the form of unstructured data such as video files, audio clips, and pictures captured on wearables and mobile health applications.

Providers have realized that they can no longer ignore the vast amounts of unstructured data that is being accumulated, within, and outside their firewalls. As technologies and data management techniques have evolved, unstructured data is an insightful informational source that can help find more use cases and clinical applications for precision medicine.

Unstructured data presents a new set of opportunities; for example, sophisticated natural language processing algorithms can be used to mine call transcripts can now “read” physicians’ and nurses’ notes. Recent trends in the industry show that there has been significant growth in the form of unstructured data such as video files, audio clips, and pictures captured on wearables and mobile health applications. Providers have realized that they can no longer ignore the vast amounts of unstructured data that is being accumulated, within, and outside their firewalls.

Skilled Workforce

Providers note that there is a shortage of specialized skills, especially data science skills. They say that finding and keeping people with the right skills, especially data science skills, is a challenge.

Referring to the dogma of the 3Ps: People, Process and Product, one third of the issues providers are having in the space of genomics and precision medicine is a result of not having staff with the needed qualifications. In this study, the majority of providers stated that they are finding it difficult to recruit and keep the right talent to fill data scientist and analyst positions. They are looking for analysts who have a multidisciplinary background in information systems and technology, science, management and engineering, thus this mixture of skill is not easy to find.

8. [“How to Mine Unstructured Data”](#) Rupanagunta, K., Zakkam, D., and Rao, H.

Furthermore, most physicians have not been trained on how to interpret and use genetic data. Hence, even with existing personnel, there is a need for more education and training in the specific areas of genomics and precision medicine.

Better Tools

Providers report needing better tools to make the most use of big data. Providers are struggling with ingesting the data output from their many heterogeneous applications, all of which they hope to feed into their big data platform.

Providers are searching for software solutions that can not only help them to store and manage their data, but can leverage the value of big data for practical applications.

Providers are searching for software solutions that can not only help them to store and manage their data, but can leverage the value of big data for practical applications. Some providers that are just beginning to get involved with research are interested in partnering with data management vendors that support transaction processing and analytics.

Interoperability

Providers want their genomic data systems to be interoperable with their electronic medical records systems and other internal clinical systems. In particular, providers want clinical research management software that will integrate with the EHR.

From a safety perspective, it is paramount that clinical information from research trials be able to flow to the electronic medical records system for providers to access when necessary. For instance, if a clinical trial patient shows up in the Emergency Department, the patient's EMR should flag him or her as a trial patient, to quickly give information to the caregiver to help them know whether one of the drugs in the trials is causing the adverse event. Achieving this level of interoperability is pivotal for providers to be able to use data from clinical trials to inform decision making at the point.

Data Sharing

Providers want to be able to exchange data with external systems, for example for research purposes, but also because of a regulatory push to share data. This trend raises issues around data sharing. There is a slight conflict between the push to share data and HIPAA rules, which reward for risk-averse behavior.

Although sharing data will benefit the enterprise and the patient, providers are treading cautiously with sharing genomic data. One provider commented, "I see a lot of risk and I share the concern that this data is very personal and we want to know where that data lives, who is accessing it and how it is being used. I also see that, clinically, it can help patients have the

best treatment by sharing. So there is a fine line between risk and reward. Even people like me, in the security role, we can sometimes forget the value of this to the patient.”

Data sharing requires additional oversight or governance processes to analyze the cases where personally identifiable information (PII) is shared outside the organization and risk profile those different requests. Based upon the risk profiling, providers can create a set of standards and expectations around data controls. Accomplishing this requires alignment of stakeholder interests, increased financial and corporate priority investments, and better industry standards, policies, and reimbursement structures.

“I see a lot of risk and I share the concern that this data is very personal and we want to know where that data lives, who is accessing it and how it is being used. I also see that, clinically, it can help patients have the best treatment by sharing. So there is a fine line between risk and reward. Even people like me, in the security role, we can sometimes forget the value of this to the patient.”

Participation in Clinical Trials

Providers need increased participation in clinical trials, and they voiced challenges with identifying participants for clinical trials. There are significant sex, race and age-based disparities in clinical trial enrolment, and many phenotypes are not included.⁹ In this respect, providers noted concern that the data they are generating from genomics research may not reflect the broader population.

For example, the people who participate in clinical trials tend to be more homogeneous and are usually wealthy, Caucasian males in 50s and 60s. When drugs are developed based on small datasets from a relatively homogeneous population, they turn out to be not as effective when applied to a more heterogenous population in communities.

Providers want to get a wider representation of patients to participate in clinical trials to increase the accuracy and precision of their research data and applicability of results generated, so that drugs will work for a more heterogenous population. Data solutions that present an opportunity for providers to improve on their recruitment strategies for trials, and to upgrade their methods of identifying more eligible participants for clinical trials would be beneficial.

Patient Registries

Providers need support with managing patient registries. The Precision Medicine Initiative and other precision medicine projects nationwide will require tools that can automate and simplify the management of the patient information.

9. “[Dialogue on Diversifying Clinical Trials](#)” Society for Women’s Health Research and the Food and Drug Administration, Office of Women’s Health

A registry is a collection of information about individuals, usually focused around a specific diagnosis or condition. Registries may collect information about people who have a specific disease or condition, while others seek voluntary participants who may be willing to participate in research about a disease.¹⁰

With genomics, maintaining a patient registry is crucial to uphold the integrity of the research data. The Precision Medicine Initiative led by NIH (“All of Us”) can be viewed as a national registry of one million or more volunteers for the purpose of conducting research on precision medicine for all diseases. Providers are struggling to do this as it is not widely automated and requires significant administrative resources.

Maintaining the registry not just for the PMI, but for any other research project, will be a challenge. Meaningful Use (MU) Stage 2 regulations under the Affordable Care Act stipulates the need for reporting to registries to improve population and public health.¹¹ To do this requires special tools and interfaces. Providers indicated that managing registries is expensive and is currently done by different vendors for each service line. They also revealed that it can take up a significant portion of administrative resources as it requires a full-time employee to manually enter the data to be fed into the national registry.

Providers need automated solutions that can simplify tasks and reduced the administrative burden necessary to maintain these registries and manage the flow of data to and from the national registries. It is worth noting that we are not yet privy to the intentions of the current GOP Administration it relates to amending many healthcare laws and policies. Hence, it is not known if the requirements under Meaningful Use will be changed by this new administration.

REDCap is a research/data capture open source software developed by Vanderbilt University available for all Universities and Academic Healthcare organizations. It is specifically geared to support online or offline data capture for research studies and operations and some providers are using it to maintain their registry related data. A gap that the providers identified was that they need to figure out what it will take to interface data from their warehouse to be automatically updated onto web applications like REDCap to significantly reduce workload and burden for managing the registry.

Extracting Value from Metadata

Providers want to understand how to extract value from metadata. Providers recognized that the amount of metadata associated with data is growing and they are interested in understanding if there is value that can be derived from it.

10. [List of Registries](#), National Institutes of Health

11. [Specialized Registry](#), Centers for Disease Control and Prevention

Metadata is data that describes other data. It summarizes basic information about data, which can make finding and working with instances of data easier. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.¹² Providers are interested in how metadata is created, how it can be standardized, and whether there are standard tools for extracting value from it. This opportunity of big data is yet to be fully explored.

Providers are interested in how metadata is created, how it can be standardized, and whether there are standard tools for extracting value from it. This opportunity of big data is yet to be fully explored.

Aging Populations

Providers want to use precision medicine to meet the healthcare needs of an aging population. The population is also expected to become much older. By 2030, more than 20 percent of U.S. residents are projected to be aged 65 and over, compared with 13 percent in 2010 and 9.8 percent in 1970.¹³ This presents a significant opportunity for increased research and treatment discoveries that target degenerative brain diseases like Parkinson's and Alzheimer's.

Common Data Definitions

There is a need for common data definitions. A significant challenge noted by providers is the need to be able to apply common data definitions to different variables as there are multiple variations of biological or clinical terms and there is no dominant industrywide standard for data definitions. For true interoperability to occur among providers, it must be possible to normalize data to common clinical definitions across the industry.

From Competition to Collaboration

Providers want to move from competing to collaborating. They recognize that the industry's inclination to compete rather than collaborate is a challenge to progress.

One provider observed, "It's easier for us to partner with those abroad than with our neighbors here in the US. You make a lot of money when you have a patent, especially in cancer, and that's a challenge." Providers are looking for a culture shift from competing to collaborating. This change in culture will only come about when you have alignment in leadership vision, incentives, policies, and reimbursement models.

12. "Metadata" Tech Terms

13. "An Aging Nation: The Older Population in the United States" Ortman. M. J., Velkoff, A. V., and Hogan, H.

Looking Forward: Capabilities to Advance

Expanding Options for Data Storage

Providers realize that they need better understanding of how to utilize and leverage external data storage resources like the cloud and state repositories such as the Texas Advanced Computing Center (TACC). Outsourcing storage space will become an inevitable option in the future as the size of their data grows to exceed their in-house capacity.

Specifically, providers are beginning to see the benefit of focusing on their core competency of delivering clinical care and outsourcing the adjacent functions such as data storage and management. From this study, it is clear that providers are open to exploring external options for data storage and want to understand how they can leverage the full benefits of joining state repositories such as TACC. They want access to information that can consult them on the advantages and disadvantages of using external storage platforms to guide in their business strategies.

Integrating Genomics with Clinical Practice

Amassing genomic data only provides value if it can be used to improve the quality of life for those who have shared their bioinformation or to advance the knowledge of genomic medicine. Providers are motivated to find use cases for their genomic data that can increase the value to their patients and improve health outcomes.

Most providers in this study are already involved in genomics research, but few are currently using the data to develop precision medicine use cases. All providers expressed interest in developing their precision medicine capabilities, so there is an opportunity to help providers realize the clinical value of their genomic data by developing their research competencies to be able to offer specific and personalized clinical treatment solutions for their patients.

Merging Genomic Datasets with the EMR and Other Clinical Systems

In general, genomic datasets for research purposes are independent and separate from the electronic medical records. Providers in this report are beginning to realize the need to integrate these two systems to form a complete picture of the patient's medical profile. Specifically, they are asking for query functions in the EMR to pull results and analyses from the genomic databank for real-time decision support for clinicians. Use cases include determining the best drug to prescribe based on the presence of genetic variants that can cause adverse drug effects, proactively identifying oncogenes and taking preventative action to deter or prevent the onset of cancer and discovering new genetic links to diseases.

Overcoming Interoperability and Privacy Issues

Providers see that they need to overcome interoperability issues, privacy and security concerns, and data sharing limitations if they want to leverage the power of big data in identifying health trends and improving the health of populations.

The providers interviewed are optimistic about the future and they all share goals that are aligned with achieving interoperability and mutually beneficial data sharing. The White House Precision Medicine Initiative has sparked nationwide interest in precision medicine and it is hoped that this initiative will pave the way for the development of policies and payment structures that support true interoperability and data sharing needs. As for privacy and data security, although it is true that genomics data has not yet been the target of ransomware (as have medical records), it is still a vulnerable information source and there is an urgent need for heightened safeguards to protect this sensitive information.

As for privacy and data security, although it is true that genomics data has not yet been the target of ransomware (as have medical records), it is still a vulnerable information source and there is an urgent need for heightened safeguards to protect this sensitive information.

Alignment of both effort and investment in these key areas among all stakeholders is required in order for industry-wide progress to be made. Providers need to assess their current capabilities and determine what is missing for them to achieve true interoperability, improve privacy and security infrastructures and establish mutually beneficial data sharing relationships.

Managing Patient Registries

The growth of the precision medicine market will require enhanced technology tools that can streamline the processes of data capture, data sharing, and data standardization. The providers interviewed shared that they are looking for solutions that can reduce the administrative burden of maintaining their local registry, interface with other internal clinical systems, and exchange data with national registries. This is an area of opportunity for vendors in the space of precision medicine and genomics.

The growth of the precision medicine market will require enhanced technology tools that can streamline the processes of data capture, data sharing, and data standardization.

Role of Marketing, Social Media, and PR

The success of the precision medicine research projects and initiatives rely on the voluntary involvement of people. Hence, to capture the interest of a wide representation of participants, there is need for marketing to create a positive buzz about clinical trials and to help people realize the benefits of getting involved. Organizations will need to design an effective marketing strategy to identify potential participants and tailor the

marketing to the needs of specific target populations. Social media is a powerful platform that can be leveraged to transmit these messages and reach a wide audience.

Using Precision Medicine to meet the Healthcare needs of an Aging Population

The GOP platform mentions Alzheimer's disease in the context of advancing research: "Federal and private investment in basic and applied biomedical research holds enormous promise, especially with diseases and disorders like autism, Alzheimer's, and Parkinson's."¹⁴ This presents a tremendous opportunity in the space of precision medicine for public-private partnerships in research to find cures for patients suffering from degenerative brain disorders. As the US population is aging, this is an area where innovation in technology, information systems, and clinical techniques is necessary.

Vendors can play a critical role in providing the data management infrastructure to support these research programs and offer tailored solutions to the needs of individual organizations.

Predictive Modelling and Consolidated Data

In the space of cancer research, providers divulged that the next progress step is to develop predictive models to identify cofactors or biomarker that will better predict whether a person will develop a disease—for example, to help predict whether a person who survived cancer will develop a different type of cancer. This kind of prediction requires obtaining data on the social determinants of health and chronic diseases that may be amplifiers for cancer. This analytic is most important not only for cancer but for health in general.

In addition, providers want a data bank that has the querying to allow them to click a button and pull specific information, such as the number of cancer patients that are being treated for cancer in the left side of their lung.

14. ["Where Trump And Clinton Stand On Health Care And Medicare"](#) Gurnon, E.

Research Methodology

In this study providers were identified based on criteria that assesses their involvement in genomics research and use. Research was conducted to identify organizations involved in genomics and precision medicine. A select group of provider executives were selected to participate in a formal phone interview with eHI to share insights on how they are using big data, their involvement with genomics and precision medicine, and what challenges they are facing in managing, storing, using and analyzing the data. The interviews were structured and involved 10 questions with the following objectives:

- To understand how “real” and “accessible” the precision medicine market is
- To understand the state of data management in precision medicine from research to clinical implementation
- To understand if organizations are adopting newer technologies, like Hadoop, in order to handle the data demands of precision medicine, and the budget to accommodate the different technologies

Participating Organizations

Thank you to the following industry pioneers who contributed their time and expertise to this report:

CancerLinq / American Society of Clinical Oncology

Children’s Hospital Philadelphia

Johns Hopkins Hospital

Indiana University Health

Marshfield Clinic

Mayo Clinic

Vanderbilt University Medical Center

among others

References

Centers for Disease Control and Prevention, (2015). Specialized Registry. Office of Public Health Scientific Services (OPHSS). Retrieved at: https://www.cdc.gov/ehrmeaningfuluse/specialized_registry.html

Gurnon, E., (2016). "Where Trump and Clinton Stand On Health Care And Medicare." *Forbes*, Retrieved at: <https://www.forbes.com/sites/nextavenue/2016/08/12/where-trump-and-clinton-stand-on-health-care-and-medicare/#782d59853784>

Kaiser, K., (2016). "President Obama's 1-million-person health study kicks off with five recruitment centers." *Sciencemag.org*. Retrieved at: <http://www.sciencemag.org/news/2016/07/president-obama-s-1-million-person-health-study-kicks-five-recruitment-centers>

Lyko, K., Nitzschke, M., & Ngomo, A. C. N. (2016). "Big Data Acquisition." In *New Horizons for a Data-Driven Economy* (pp. 39-61). Springer International Publishing. Retrieved at: https://link.springer.com/chapter/10.1007/978-3-319-21569-3_4

National Human Genome Research Institute, (2016). "What is Genomic Medicine?" Retrieved at: <https://www.genome.gov/27552451/what-is-genomic-medicine/>

National Institutes of Health, (2016). List of Registries. Retrieved at: <https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries>

National Institutes of Health, (2016). "What is precision medicine?" Retrieved at: <https://ghr.nlm.nih.gov/primer/precisionmedicine/definition>

Ortman, M. J., Velkoff, A. V., and Hogan, H., (2014). "An Aging Nation: The Older Population in the United States." Population Estimates and Projections Current Population Reports. Retrieved at: <https://www.census.gov/prod/2014pubs/p25-1140.pdf>

Privacy Technical Assistance Center, (2013). "Data De-identification: An Overview of Basic Terms." Retrieved at: http://ptac.ed.gov/sites/default/files/data_deidentification_terms.pdf

Rupanagunta, K., Zakkam, D., and Rao, H., (2012). "How to Mine Unstructured Data." Retrieved at: <https://www.information-management.com/news/how-to-mine-unstructured-data>

SWHR, FDA, (2011). "Dialogues on Diversifying Clinical Trials: Successful Strategies for Engaging Women and Minorities in Clinical Trials." The Society for Women's Health Research United States Food and Drug Administration Office of Women's Health. Retrieved at: <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM334959.pdf>

Tech Terms, (2017). "Metadata." Retrieved at: <https://techterms.com/definition/metadata>

White House Briefing Room, (2016). "FACT SHEET: Obama Administration Announces Key Actions to Accelerate Precision Medicine Initiative." Archived at: <https://obamawhitehouse.archives.gov/the-press-office/2016/02/25/fact-sheet-obama-administration-announces-key-actions-accelerate>

White House, (2016). "The Precision Medicine Initiative" Archived at: <https://obamawhitehouse.archives.gov/precision-medicine>



Copyright © 2017
Foundation for eHealth Initiative
All rights reserved. March 2017

www.ehdc.org

Research supported by PHEMI

 **PHEMI**[®] Big Data. Shared. Simply. Securely.

www.phemi.com info@phemi.com www.twitter.com/PHEMIsystems www.linkedin.com/company/phemi