REMEDY INFORMATICS

WHITE PAPER

Enterprise Informatics:

Key to Precision Medicine, Scientific Breakthroughs, and Competitive Advantage

Gary D. Kennedy
Founder and Chief Executive Officer
Remedy Informatics, Inc.

TABLE OF CONTENTS

Introduction

Given their level of investment in data and data management systems, healthcare delivery and life sciences organizations should be deriving considerable value from their data. Yet most organizations have little to show for their effort; the capabilities of their systems are highly compromised, and the practice of precise, evidence-based medicine remains elusive. The fact that these institutions have spent many years collecting data and building infrastructure for so little return has, for many, become "the elephant in the room"—a painfully obvious and uncomfortable topic of conversation.

With two notable exceptions, every major industry has realized significant productivity gains over the past 30 years by automating their enterprise applications—utilizing enterprise resource planning (ERP), customer response management (CRM), automated trading, consolidated billing, and other systems—and then using the data generated by these applications to continuously improve their internal processes. The only industries that have been unable to implement enterprise systems, iterate around the data, and then generate incremental improvement are higher education and healthcare.

In healthcare, data should, of course, deliver value by being used first and foremost to help improve treatment effectiveness and patient outcomes. An organization could then leverage its resulting superior track record to gain market share, secure study funding, or pursue other initiatives vital to its continuing success. But in reality, data management among healthcare organizations has largely been an exercise in mere data *collection*, consuming resources more than contributing them—often to the point of detracting from an institution's ability to fulfill its core mission. This makes for a serious and unsustainable situation.

Fortunately, this problem, which has been snowballing for years, is at last solvable with advances in information technology that make it possible to break through the constraints imposed by old technology and legacy systems.

By leveraging these advances to create a sophisticated and linked set of registries, healthcare and life sciences organizations can successfully and productively conduct informatics across the enterprise. This paper presents the case for using "meta registries" to identify patterns, develop and test hypotheses, make better medical decisions, and, ultimately, achieve a competitive advantage.

Evidence-based medicine: An elusive goal

Evidence-based medicine has been the goal of many institutions for years, but has for the most part not been realized. The reality is that very few healthcare and life sciences organizations can point to tangible examples of improved outcomes directly resulting from more-informed, data-driven decisions. While there may be a few departmental exceptions, the vast majority of

> "Most healthcare executives will tell you that they can't get access or meaning out of their own data, and they don't trust anyone else's data."
>
> —**Andy Slavitt**
> Group Executive Vice President, Optum[1]

institutions struggle to correlate better data with better decisions or outcomes.

The heart of the problem is that data has been collected for decades using inconsistent methodologies and then stored in incompatible systems. Data from many sources has in essence been thrown together, often forced into whatever format makes loading convenient, and then submitted to a business intelligence (BI) engine to be analyzed. Historically, there has been no automated model to make this data consistent and to consolidate it so that physicians and researchers can use it to make informed decisions based on tangible evidence.

The result is that researchers and clinicians cannot query data on their own; they must engage IT staff to write a program for them and then produce a report based on its findings. Since it is the nature of science for one answer to lead to subsequent lines of inquiry, this process is inefficient in the extreme. Weeks can elapse between the iteration of each query and answer.

Even more limiting is the fact that it is simply not possible to obtain a composite view of all relevant data pertaining to a patient or a research subject *across time*. Having a complete, longitudinal perspective is essential for correctly identifying patterns that lead to breakthroughs in science and that guide effective treatment decisions. If the data that is used to make decisions represents only a slice of any individual patient's history—and not a comprehensive, contiguous, uninterrupted record—the patterns that emerge may be misleading. Any resulting conclusions could be invalid, perhaps with tragic consequences.

Another major limitation is that when data remains in separate silos, each created to serve a narrow purpose, that data cannot be easily linked across sources. For example, in one database, a patient may be recorded as "male," in another as "M," and in yet another as "1." If the data management system does not address these inconsistencies—which are multiplied by the thousands, and often exponentially more complex—any attempt to interrogate multiple databases will be hindered by gaps, and any conclusions drawn will be suspect. Thus, researchers are prevented from conducting the kind of cross-disease analyses that can lead to a stream of breakthroughs well beyond what can be accomplished by individual researchers working within their independent fields of study.

In an ideal world, data is used across the enterprise to improve effectiveness

When a comprehensive data set is harmonized and made accessible, truly amazing things are suddenly possible. Institutions can finally use the data they have worked so hard to collect to practice evidence-based medicine. Previously unseen patterns are discernible and lead to discoveries and advances in care. Better decisions are made. Patients benefit from higher-quality outcomes. And institutions operate more efficiently.

When a comprehensive data set is harmonized and made accessible, previously unseen patterns are discernible and lead to discoveries and advances in care.

The ideal model is to utilize a full complement of data to identify patterns and develop a hypothesis, and then to use a small percentage of that data to test the hypothesis. (If the hypothesis is not affirmed, then the clinician or researcher must iterate back through the comprehensive data set until a more verifiable pattern is found.) When researchers and clinicians can generate a hypothesis, query the data, and visualize the results *on their own*—in real time—they raise the level of discourse in medicine. Such agile investigation will dramatically increase the productivity of clinicians, researchers, scientists, and IT professionals.

With a 360° degree view of a disease—encompassing clinical, specimen, and molecular data—researchers will be able to see all genetic mutations, complications, and symptoms associated with it. Who knows to what breakthroughs that will lead? And when researchers have the ability to drill into data across disease-specific databases, they will be able to make untold new discoveries. Imagine, for instance, the power of having instant insight into the correlation between breast cancer and colorectal cancer, or how HgA1C levels could affect surgery outcomes.

Healthcare will take a giant leap forward when researchers and clinicians have the power to conveniently share data in a way that is harmonized. Collaboration will compress the cycle time between bench discoveries and applications at the point of care, and between outcomes in clinical practice and further scientific explorations. *The promise of translational medicine will finally become a reality.*

And, we will at last practice personalized, or precision, medicine. Its foundation will be a continuum of accurate data that affords a solid understanding of outcomes: which interventions produce the best outcome, what follow-up is required, which measures can prevent an adverse event, and so on. In the future, it will be common to see 10 or more sources of data aggregated, providing a complete picture of the patient. In this enlightened scenario, the treatment plan a physician defines will be customized to a patient based on his or her personal characteristics. A physician will identify the precise therapy for an individual that, if delivered via a specific vehicle at the most opportune time, can render the greatest benefit.

The high cost of incompatible data

By failing to optimally use already available systems and the data they generate, institutions waste millions of dollars each year collecting data that, arguably, serves no real purpose. Considering that competing healthcare providers are so intent on cutting costs that they will unite in order to save $.10 on each surgical gown, and that life sciences companies have undergone massive reorganizations to improve their margins, certainly industry organizations must be questioning their single largest expenditure that yields no concomitant tangible benefit: collecting useless data.

There is a commonly held belief that cures for many diseases already reside within existing data, if only the data could be made to give up its secrets. The key to making this happen is the ability to recognize patterns within the data.

The less apparent but far greater cost relates to lost opportunities. The pace of advancement in biomedical research has created massive amounts of data that could be used to significantly improve health outcomes. Within medical research and development, there is a commonly held belief that cures for many diseases already reside within existing data, if only the data could be made to give up its secrets. The key to making this happen is the ability to recognize patterns within the data.

Consider the number of lives lost each year because a clinician is unable to accurately identify a pattern. Too often, even though the necessary data is available, it is splintered across so many data silos that it is all but impossible for

a physician to detect a meaningful pattern in the resulting sample. Imagine the difficulty of explaining to a patient's family members that their mother died—even while there was evidence from other cases that her particular strain of triple-negative breast cancer was not responsive to the course of treatment provided.

More generally, without the ability to retrieve answers from data—confidently, quickly, and easily—we stop asking questions because it becomes an exercise in futility. People stop forming hypotheses because they cannot easily affirm or disprove a new idea, and, as a result, innovation suffers.

Another unfortunate consequence is that scientists and others in healthcare, so oriented toward basing decisions on evidence, will use whatever data is available—or worse, will use whatever data is collectible, even if it is not relevant to the problem at hand. The data they use may present a discernible pattern, but no one should assume it is the correct one. This problem is made more serious because too often it is not even recognized as being a problem.

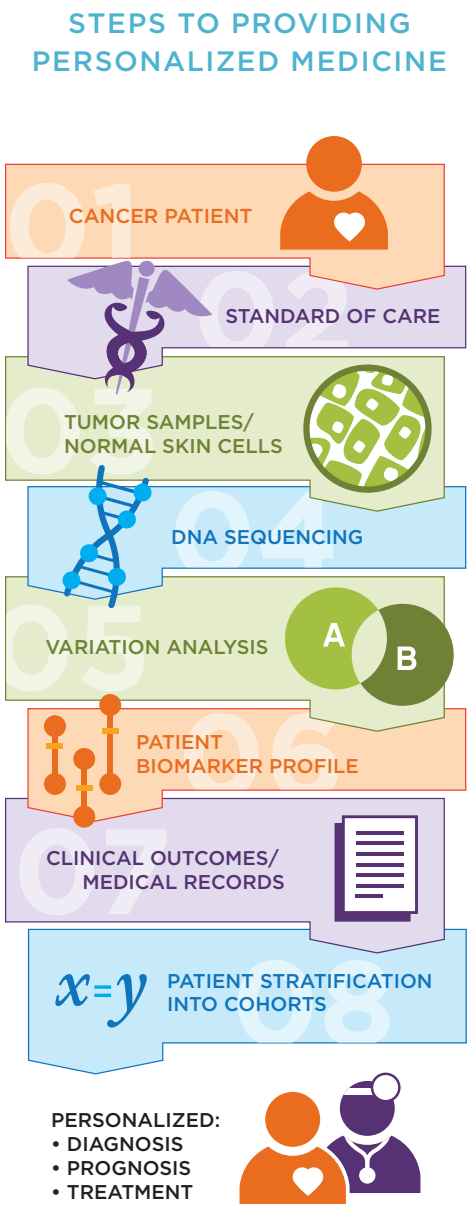
These issues are real, they are pervasive, and they are severe. They impede progress in very important initiatives such as finding cures for chronic diseases, understanding and treating rare disorders, controlling spiraling costs in healthcare, and making significantly better use of the limited funds available for medical research.

A matter of urgency

Developing trends suggest that the disparity between what is needed in healthcare and what current IT systems can provide is going to widen dramatically. The number of data sources and the overall volume of data are growing, and increasing numbers of institutions are collaborating in public and private partnerships. Moving from "small data" to "big data" and involving more players increases the complexity of the IT situation exponentially—and the longer organizations delay in overcoming these challenges and finally putting their data resources to work for them, the worse the problem becomes.

At the same time, treatment protocols are becoming more personalized and granular. In ten years, a general diagnosis of breast cancer will be insufficiently specific to suggest any kind of comprehensive treatment plan. Soon every treatment protocol will presuppose a more precise diagnosis along with the unique patient characteristics that define the subtype of cancer (see **Figure 1**). The fact that a patient's cancer occurs in the breast will be less important than

Figure 1



several other factors, including the specific genes or biomarkers involved, and data on how the previous 100 patients with the same characteristics fared over the course of a given treatment regimen.

Attempts to solve "The Data Problem"

The fact that enterprise-computing applications have, historically, been notably inadequate is no surprise to anyone involved in healthcare delivery or life sciences research. To their credit, the extremely capable people in these fields have not only recognized the issues, but they have undertaken a variety of ambitious and expensive initiatives to address the "data problem." Unfortunately, none of their attempts have been successful in resolving it, largely because they have been applied in patchwork fashion. The result is an assortment of individual, best-of-breed technologies running in parallel, all of which are used separately. Meanwhile, data remains compartmentalized in silos; this makes it impossible to obtain a composite view of a patient or research subject based on the integration of all data from every source (see **Figure 2**). As long as healthcare is unable to see a 360° view of every patient, it will continue to treat diseases, not people.

To address their need to collect and aggregate data, most healthcare providers have turned to electronic health record (EHR) applications. In the United States, the deployment of EHR applications has consumed the vast majority of providers' IT dollars, time, and human resources. EHRs have proven useful in collecting some types of transaction-oriented, structured, and coded data, especially in the area of defining and enforcing complex order sets. But it is important to recognize that EHRs were never designed for flexibility or to aggregate and harmonize all data of any type into a single system. For these reasons, EHRs cannot give clinicians or researchers all data that is pertinent to a patient or research subject, nor are EHRs able to analyze and report on data in a way that makes it possible for clinicians and researchers to recognize patterns and take the appropriate action.

Consequently, none of the top healthcare institutions in the U.S. rely exclusively on data collected in their EHR systems even though they may have cost hundreds of millions of dollars to implement; all supplement EHR data with a variety of data collected and stored in purpose-specific databases otherwise known as registries. A registry can store all the data needed to form a 360° view of an individual patient or research subject, in contrast to the narrow slice of clinical data that can be extracted from an EHR. For this reason, registries will become as important to healthcare as relational databases are to business.

Registries can be designed to measure any aspect of an organization's performance ranging from patient outcomes to the effectiveness of a particular chemical or biological compound. They can support a wide variety of needs, including disease-specific research, population health management, patient recruitment, patient behavior assessment, patient-centric outcomes research, clinical research, clinical trial or study management, ad hoc reporting, and clinical decision support. Quality and outcomes registries, meanwhile, are used to provide the

Because a registry can store all the data needed to form a 360° view of a patient or research subject—in contrast to the narrow slice of clinical data that can be extracted from an EHR—registries will become as important to healthcare as relational databases are to business.

empirical evidence needed to manage today's accountable care model.

While collecting data supplemental to the EHR via registries and aggregating clinical data from multiple EHRs are widely held best practices in healthcare, doing so nonetheless presents several challenges to enterprises intent on maximizing the value of their data:

- Unless the EHR and registry systems are linked so that the entry of certain data components in one system automatically populates the other, some data must be entered multiple times, reducing efficiency and data consistency.
- The number of registries that need to be harmonized with the EHR system—and with each other—is large and growing daily; in each system, a specific data element is often represented by multiple names and codes.
- Because these disparate technologies were not designed to work together from the outset, attempting to integrate them is an arduous undertaking and, ultimately, an unsuccessful approach that does not lead to accurate, aggregated results.

Of course, it has always been possible to link and harmonize data across multiple databases if it is done manually, but doing so on an organization-wide basis would require armies of subject matter experts to go through the data piece by piece in order to turn it into a single, coherent whole. Indeed, some institutions characterize this as their "million man problem," because they feel as if it would take one million people working together a full year to accomplish—or, just as unreasonably, that it would take only one person a million years.

Without exaggeration, it is an excruciatingly manual process that is typically so protracted, so costly, and such a wildly inefficient use of resources that it is rarely done, and then only on a very limited basis for a select, relatively small set of data for use in a particular project rather than on a large scale. It is, in fact, such a complex and costly undertaking that some organizations that have been collecting data for 30+ years are choosing to start over rather than map all their legacy data to a new system.

The prohibitive cost of manual data linking and harmonization underscores the need for organizations to be able to build registries, perhaps hundreds of them, in a way that supports automated, timely, and cost-effective integration.

The impediments

So why has it taken healthcare and the life sciences so long to address the same problems that other industries have already put behind them? What, exactly, is so difficult about collecting data via EHRs and other sources, aggregating it in a way that is meaningful and consistent, and then analyzing it to make informed decisions? Why has the healthcare industry been unable to take advantage of the transformative nature of

Figure 2

aggregated data as other industries have successfully done? The reasons are many:

The extremely complex nature of aggregating, mapping, and harmonizing healthcare data. The primary reason it has taken healthcare so long to bring all of its data together and then use that data for improvement is because it is extremely difficult to do correctly—and there is no margin for error. The inferences made have to be right. New sources and types of data, which are becoming available on a monthly basis, need to be integrated fully and immediately. So many "standards" exist that each data element is commonly represented in many different ways, too often depending on context and the background of the person entering the data. It has taken so long because the industry has lacked a technological answer for dealing with these, and many other, daunting complexities—until now.

The primary reason it has taken healthcare so long to bring all of its data together and then use that data for improvement is because it's extremely difficult to do correctly—and there is no margin for error.

Inflexible, antiquated technology architecture. More than half of all U.S. inpatient encounters are recorded on EHR systems built on an architectural structure developed in the 1960s: the Massachusetts General Hospital Utility Multi-Programming System (MUMPS). Leading EHRs that are based on a MUMPS architecture include Epic, Meditech, VistA (a system widely used by the Veterans Health Administration), and a variety of homegrown applications that are based on Caché, a popular middleware that is a derivative of MUMPS. Technologically advanced when it was introduced 47 years ago, the MUMPS platform is inherently inflexible and thus inadequate for the needs of clinicians or scientists conducting any sort of research—an obstacle that cannot be surmounted by the software vendors using it. Renaming MUMPS to Caché or branding it as a "post-relational" platform does not alter its underlying rigidity and lack of adaptability.

Most of the EHRs on the market that are not based on MUMPS architecture, such as Oracle, have a relational database foundation. An Oracle-based system is more adaptable than a MUMPS-based system, but the fundamental problem remains that EHRs are designed for transactional, structured, and coded data. They will never be able to collect or analyze *all* data on any patient or subject. They do not track results throughout a heterogeneous patient population. It is simply not possible to combine all types of genotypic data with phenotypic data through an EHR and then use the resulting data sets for pattern recognition.

While the current generation of EHR systems may improve incrementally to enable more information to be collected and stored in the EHR, there is no way to make existing EHR architectures more flexible. In fact, enterprise applications have traditionally become more restrictive over time instead of more flexible. One either has a modern architecture that easily accommodates change, or one does not.

Flawed underlying assumptions about data linking. Conventional "wisdom" says that depositing all data from all relevant sources together in a data warehouse will somehow magically provide a comprehensive picture of a patient and his or her condition, a notion quickly disproven by taking a closer look at the practice.

The longstanding approach to loading different types of data into the same database is to apply a process known as extract, transform, and load (ETL). This practice is followed throughout the healthcare industry as standard operating procedure, even though ETL was designed for use with very simple transactional data, not deeply complex medical data. A further, and final, strike against ETL is that the process fails to account for how various data points interrelate. Simply put, ETL is incapable of linking clinical

data generated by an EHR with research data from clinical trials, genomic panels, and many other types of specialized data collection devices; even patient-centric outcomes and quality data are largely eschewed by clinical systems such as EHRs because the necessary linkages cannot be established.

And that is because today's ETL solutions were designed for warehousing data, not integrating it. So as data sets are placed in or retrieved from storage via standard ETL procedure, and moved from one place to another, they convey in isolation as a series of self-contained heterogeneous systems with no, or very little, integration. Lacking integration—which results from the absence of a master ontology and its governing semantics—each data set tends to lose meaning every time it is moved from one source to its new destination, not unlike the way a simple sentence or phrase is unwittingly modified every time it is passed from one child to another during a game of "telephone."

Over the years, hundreds of standards have been implemented with the goal of improving data compatibility. Historically, however, no system has proven compatible with all the standards, and standards often overlap or are inconsistent, incomplete, or subject to interpretation. The ETL process, applied to data that has been formatted in any number of ways and is drawn from a variety of different sources, does nothing to correct this, and rather than making disparate data more compatible actually contributes to the production of data silos.

Unwillingness to learn from advances in other industries. There is very little argument that healthcare is the least advanced of any major industry as it pertains to information technology. There are many lessons that those in healthcare delivery and life sciences research can learn from other sectors. Unfortunately, best practices from other industries are rarely adopted, either because the "it won't work in healthcare" excuse is invoked or because healthcare industry practitioners believe they must make, and learn from, their own mistakes.

As the noted information security researcher and innovator Dr. Peter Tippett has observed, "If only healthcare could do with data what finance does with data . . . three things would happen: we'd have dramatic reductions of costs of healthcare, we'd have dramatic improvements of quality of care, and we'd have an entirely new kind of science."[2]

An imperfect delivery model has created crutches and disincentives to act. Just as a rising tide lifts all boats, a competitive environment sharpens all swords. Within industries where pricing and value (cost factored by quality) are transparent, competition forces a "how can we do it better than the other guy?" mentality—and the resulting innovation benefits the entire industry.

Healthcare has not faced as much pressure to change itself, as have other industries, due to a dearth of competition. For instance, a clinician might not be strongly motivated to implement a new procedure, even one promising to improve outcomes, if its use was expected to reduce the

"If only healthcare could do with data what finance does with data . . . three things would happen: we'd have dramatic reductions of costs of healthcare, we'd have dramatic improvements of quality of care, and we'd have an entirely new kind of science."

—Peter Tippett, MD, PhD
Chief Medical Officer and Vice President, Enterprise Solutions, Verizon

number of patients he or she would be treating in the future. However unlikely this scenario may seem, the facts are that systemic incentives exist which run counter to physicians' innate desire to do the right thing. Much of the effort to replace the fee-for-service model with a more accountable, results-based model is aimed specifically at this counterincentive in today's system.

Policy makers also forget that a value-based system is dependent on a competitive landscape, and that competition is based on transparency of price and value—neither of which is the norm in healthcare today.

Using the Hippocratic Oath as an excuse for inaction. Some organizations interpret the oath taken by all physicians as "first do *nothing different.*" Rectifying the problems in life sciences research and healthcare delivery requires a different way of thinking than the mentality that brought on the problems we already face. Trial—and even error—must be tolerated and embraced

Figure 3

for innovation to take place. Obviously this needs to be carefully managed to protect every patient, but the industry has reached the point where maintaining the status quo and playing it safe is the riskiest of all alternatives.

Now, a solution is at hand

To implement evidence-based medicine, the industry needs a way to tightly integrate all available data from every relevant source and then use the data to generate a composite view of a patient, research subject, or disease. Such a view would include a patient's medical history, pedigree charts, genomic data, and even environmental data. Also needed is an automated way to harmonize all the data that currently exists in clinical data sets, financial management, and clinical trial systems, as well as data from biobanks and research labs—because only then, after the data has been harmonized, can it be made computable.

Technological advances perfected over the past decade make this possible. A new technology and methodology provide the necessary structure to link information in a way that a computer can handle, and is consistent with the way physicians understand and relate medical concepts, properties, and terminologies. This methodology, or *ontology*, provides context for each data element and relates it to other data elements according to formal and unambiguous rules. This step is often referred to as *harmonization*.

Figure 3 illustrates how a master ontology can establish the relevancy and validity of each data element, rendering it semantically consistent, explicit, and computable. No matter how many different ways a query is posed, data that has been effectively harmonized will deliver the same fact-based answer—an answer that can be applied with confidence.

When an ontology harmonizes all data entering a registry, it is possible to adopt a computing model that is source-agnostic, enabling it to accept and make use of data from multiple sources, no matter what its original purpose. Picture a registry of registries or multidimensional database akin to a cube (see **Figure 4**): into this cube is poured all the data that is available and relevant to a particular healthcare or life sciences entity.

Next, the harmonized data within each registry is linked across registries in a consistent manner. This *mapping* process is what provides dimensionality to the cube. In **Figure 4**, for example, the horizontal dimension is represented by a variety of data sources pertinent to each patient. Sources such as EHRs would provide part of the Signs & Symptoms dimension. A genomic panel of tests would provide some of the Genomic data layer that could be enhanced by other sources such as tissue samples for cardiovascular data and blood assays for diabetes. Epigenetic, microbiomic, and metabolomics data would all be generated by tests specific to each disease state.

Disease registries specific to various types of conditions provide the vertical structure in the cube. Obviously, each patient could appear in multiple disease registries, and each registry will contain multiple patients.

The industry has reached the point where maintaining the status quo and playing it safe is the riskiest of all alternatives.

Once the data is harmonized and mapped, it truly can be *aggregated* across each database within the "registry cube," and analyses can be conducted across three dimensions:

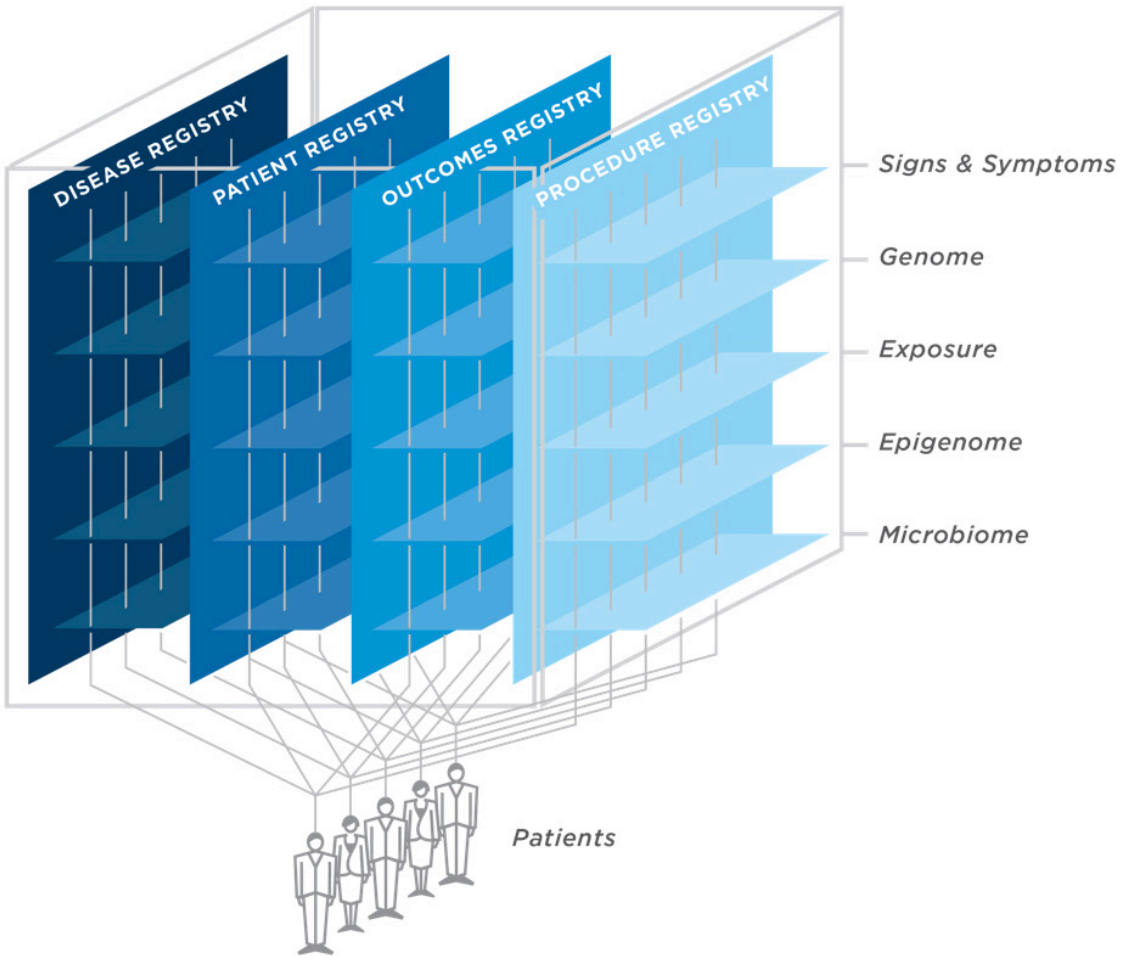
- First, the cube can be utilized to deliver a 360˚ view of an individual patient or research subject.
- Second, it can be employed to focus on population-specific patterns.
- Third, it can render a look at all relevant attributes of a specific disease regardless of whether the condition has ever been manifest in individual patients.

From data such as this a clinician would be able to precisely define a more comprehensive definition of a specific cancer subtype based on the characteristics of its underlying molecular structure, environmental causes, or responsiveness to certain treatments. This level of precision would prove much more productive than characterizing a tumor merely by its location or by its underlying physical manifestations.

Ideally, a registry cube would be used in conjunction with the ongoing collection of clinical and research data to create an "ever-learning" ecosystem of knowledge acquisition and dissemination. Such a system is modeled after the concept of an "information commons," a data repository that links different types of data—such as layers of molecular data, medical histories that include information on social and physical environments, and health outcomes—to individual patients. Research data would be continuously contributed as a result of research initiatives and collaboration; clinical data would be gleaned from the medical records of participating patients.

Figure 4

A REGISTRY CUBE ENABLES MULTIDIMENSIONAL ANALYSIS



A bright future

An institution that adopts this new computing model will be able to employ sophisticated informatics across its enterprise, ushering in a new era of medicine and research productivity.

By unlocking data and giving researchers and clinicians the ability to explore it on their own, organizations can make decisions with confidence that enable the delivery of measurable and permanent improvements in clinical outcomes. More specifically, practitioners will be able to:

- **Identify trends and valid patterns in data, supporting innovations and evidence-based decisions**
- **Foster collaboration between scientists, potentially leading to breakthroughs that would be impossible for any researcher working in isolation**
- **Advance cross-disease research by performing powerful analytics on large data sets liberated from silos for each source or type of data**
- **Practice translational medicine, accelerating discoveries from bench to bedside and back again**
- **Give researchers the tools to make new discoveries faster and more cost-effectively**
- **Improve patient outcomes**

These advances will lead directly to a competitive advantage via an enhanced academic or clinical reputation, which attracts funding and talent, and an improved cost structure. Complex research projects can be undertaken quickly and cost-effectively; researchers can concentrate on analyzing data for patterns rather than worrying about how to pull together the data they need and waiting on IT resources.

The way forward

Aggregating, mapping, harmonizing, and then analyzing phenotypic data integrated with genotypic data is a long way from where the healthcare and life sciences sectors are today. Moving toward this goal will require a number of well-considered steps:

Adopt a registry-centric computing model. Institutions motivated to take advantage of a plethora of analyses, tests, assays, and novel decision-support systems will need a framework on which to hang this new data. Categorizing and mapping the data into a variety of registries linked together into a multidimensional structure such as a cube makes this all possible.

Seamlessly link registries to each other. Registries, or purpose-specific databases, rapidly become indispensable to recognizing patterns. Any institution that appreciates the value of registries will quickly grasp that a registry of registries (i.e., meta-registry) is also essential. The most insightful patterns emerge from composite research that spans the boundaries of several diseases.

Instantiate bidirectional translational research. The NCATS funding mechanism put in place by the NIH has focused academic medical centers on the importance of translating basic science breakthroughs into tangible benefit at the point of care. Translating evidence generated at the point of care back to guide new research has not gained traction quite as quickly but will eventually become just as important.

Identify and implement "Big Win" opportunities first. The cost model in the life sciences industry is such that initiatives promising only slightly more favorable results rarely gain traction. With the power to analyze outcomes as never before, organizations can prioritize initiatives with an understanding of the payoff.

Spend the time to get the mappings right. Data that is unceremoniously dumped in a warehouse without a flexible and comprehensive way to ensure that its location is accurately catalogued is likely to contribute little practical value.

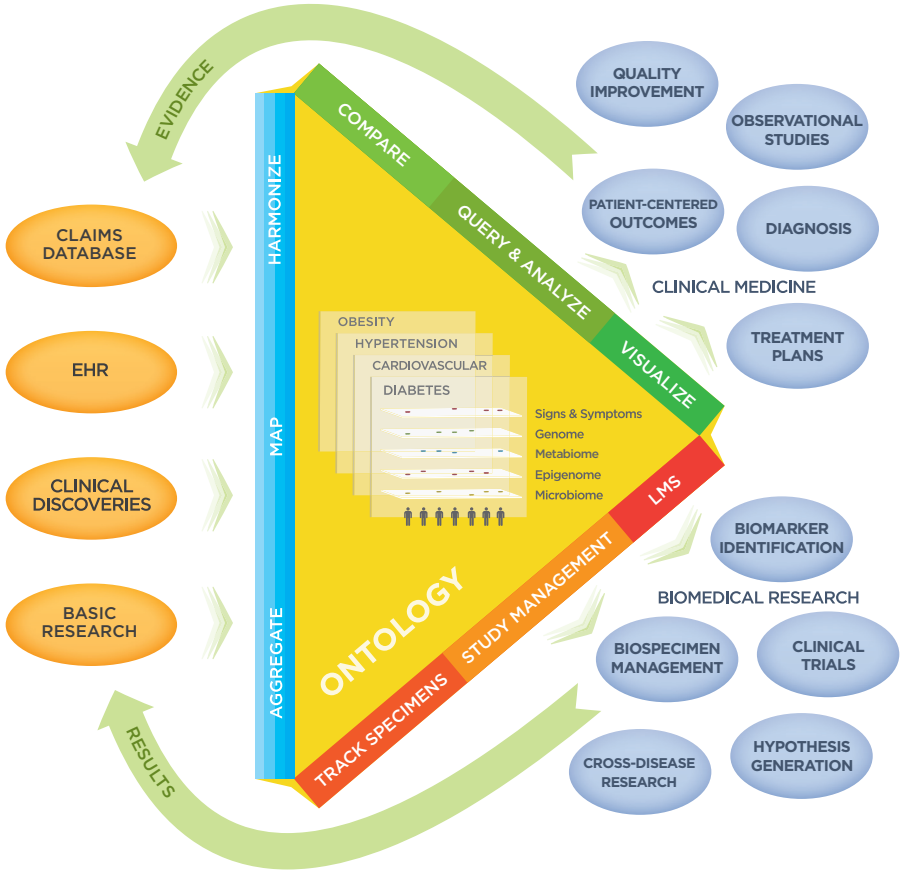
Enhance industry-standard ontology concepts with institution-specific concepts. There are a variety of related, but not complimentary, ontologies in healthcare, a situation that is unlikely to change in the next 20 years. An organization that wants to store harmonized data and collaborate with other institutions needs to commit to using the standards that are available. If it is conducting research in a yet-to-be-standardized discipline such as microbiomics, then it must implement its own ontology, and it needs a formal mechanism to adapt to new standards as they become available.

Insist on a data collection strategy at the point of interaction. Genetic characterization data is becoming viable from a cost standpoint; the next challenge is that there are only informal mechanisms at the point of care to collect the data. Even phenotypic data such as behavioral data is not readily usable unless it is part of the EHR, which would exclude most patient-reported data.

Address structural problems now before Big Data exacerbates them. Big data will expose poorly designed architectures that cannot scale. It is advisable to put together a flexible and salable infrastructure today that is sufficient for dealing with the data currently available. Then, as the volume and complexity of the problem grow, the answer will be to purchase more hardware or license more capacity on the cloud.

Figure 5

AN INFORMATION COMMONS EMPOWERS INSTITUTIONAL ADVANCEMENT



Prepare for a new paradigm in life sciences and healthcare outcomes data. Such a system (as shown in **Figure 5**) is centered around an "information commons," a repository that enables all registries, data sets, applications, and analysis to be available to anyone who has access rights through an informatics portal. The heavy lifting, linking, provisioning, and harmonizing of the data would be done in the background. Data to populate this repository would be continuously contributed by the research community and harmonized from the medical records of participating patients.

Conclusion

Healthcare institutions and life sciences organizations need a way to benefit from their data; it is their only untapped asset. Toward this end, electronic health records are receiving most of the attention and funding—yet EHRs do nothing to address the research side of the informatics problem, and they also fail to address departmental needs for clinical data management or flexible reporting and analysis. Consequently, purpose-specific databases, also called registries, have begun to fill the void left by the inadequacies of EHR systems.

Healthcare and life sciences institutions need a way to benefit from their data; it is their only untapped asset. By applying enterprise-wide informatics, organizations can improve patient outcomes and ultimately gain a strategic advantage over other entities.

The raison d'être of a registry is to aggregate, harmonize, and analyze disparate data from many sources in order to reveal a pattern or trend in a patient's care. To yield consistent and accurate results, data that is loaded into registries must be harmonized through a comprehensive master ontology. These registries are most useful when they are mapped to each other in a way that enables cross-registry analysis and reporting.

Multiple dimensions of data are easily represented by a registry cube that can be analyzed from any perspective semantically relevant to the master ontology. Institutions utilizing a sophisticated set of registries logically linked in a multidimensional cube are able to advance their research and decision-making far beyond what can be achieved when data remains in silos. By applying enterprise-wide informatics, organizations can improve patient outcomes and ultimately gain a strategic advantage over other institutions.

About the author

Gary Kennedy, Founder and CEO of Remedy Informatics, is an established thought leader in the area of enterprise informatics and the chief visionary responsible for his company's strategic focus on translational research and game-changing research informatics solutions.

Kennedy began his career as a Marketing Manager for Intel before becoming one of the first few employees hired at Oracle Corporation, where he proved instrumental in that company's early efforts to shape the relational database market in the 1980s and early 1990s. While at Oracle, Kennedy held a variety of leadership positions—including National Sales Manager, Senior Vice President of Oracle Corporation, and President of Oracle USA—and helped multiple Fortune 50 companies implement enterprise-wide relational databases and software applications designed to break down data silos and give executives a comprehensive view of their data.

The author, who holds a BA in finance from the University of Utah and an MBA from the Kellogg School of Management at Northwestern University, was previously named Utah Entrepreneur of the Year by Ernst and Young.

About Remedy Informatics

Founded on the premise that advances in healthcare and the life sciences result from the discovery of previously unseen patterns in data, Remedy Informatics delivers solutions that properly aggregate, map, and harmonize data from both clinical and research settings to give analysts a 360° view of their patients or study subjects in real time.

By breaking down information system data silos to provide truly longitudinal views of patient care and subject behavior, the company is accelerating translational and clinical research discoveries across the medical spectrum, and enabling the tracking of practice patterns and treatments that improve outcomes and increase efficiencies.

Contact us

855.736.3396, option 3
info@remedyinformatics.com
remedyinformatics.com

1 Andy Slavitt, "Using Big Data to Make Patients Better and Help Guide Doctors" (panel discussion at the Forbes Healthcare Summit, New York, NY, December 7, 2012).

2 Peter Tippett, ibid.