TRANSLATIONAL RESEARCH 2.0

Chris Asakiewicz, PhD.
Professor of Information Management and University Sector Executive, **Stevens Institute of Technology**

WHITE PAPER PROVIDED BY:

REMEDY INFORMATICS

INTRODUCTION

Translational Research is a means to accelerate the speed with which bench science translates into bedside therapeutics. However, research productivity is facing a crisis as pharmaceutical company pipelines decline and the complexity of healthcare delivery increases. In today's world of Big Data, medical and healthcare research organizations are seeking new and better ways to manage and leverage an increasing volume of complex data and information and translate it into new products and treatments. Translating new, basic scientific knowledge into enhanced clinical practice is a complex and challenging process that extends far beyond the development of a new drug, diagnostic test, or medical device. The process involves the integration of large heterogeneous sources of genomic and phenotypic data. The process necessitates tools and techniques to synthesize this data into usable information and the corresponding imperative to facilitate cross-disciplinary research and collaboration as well as aid in knowledge discovery. Translational Research 2.0 provides a framework upon which medical and healthcare research can accelerate the diffusion of biomedical knowledge into common clinical practice and improve healthcare outcomes.

The World Wide Web has revolutionized how researchers from various disciplines collaborate throughout the world. In the Life Sciences, interdisciplinary approaches are becoming increasingly powerful as a driver of both integration and discovery. Data access, data quality, identity, and provenance are all critical ingredients to facilitate and accelerate these collaborative enterprises, and it is in the area of Translational Research where Web 2.0 technologies promise to have a profound impact—enabling reproducibility, aiding in discovery, and accelerating and transforming medical and healthcare research across the healthcare ecosystem. However, integration and discovery require a consistent foundation upon which to operate. A foundation capable of addressing some of the critical issues associated with how research is conducted within the ecosystem today and how it should be conducted for the future.

This white paper will discuss the critical issues associated with Translational Research and their implications for future medical and healthcare research. The first set of issues concerns the enhancement of research for traditional ecosystem stakeholders, namely, research organizations and care delivery organizations, especially through the use of bio-repositories. The key questions to be addressed surrounding Translational Research 2.0 are:

- What is it and how might it aid knowledge discovery and collaboration within the medical and healthcare ecosystem?
- What benefits, challenges, and opportunities does it provide?
- How can bio-repositories enhance it?

The answers to these questions directly impact the translation of research findings from basic research, performed by research organizations, into clinical practice, provided by care delivery organizations.

Finally, this paper discusses how research can be enhanced for the broader ecosystem, through the mining and analysis of knowledge surrounding health outcomes, namely: What key challenges are associated with Translational Informatics in a world of Big Data? The answer to this question is of importance to all ecosystem stakeholders, in their collective efforts to better understand health outcomes and facilitate the biomedical discovery process.

KNOWLEDGE, DISCOVERY, AND COLLABORATION

An integrated medical and healthcare ecosystem leverages common data, analytics, and processes to accelerate research and enable collaborative discovery. By integrating data, analytics, and the processes that are key and common to the entire ecosystem, research can be strategically advanced at the intersections of these research disciplines (FIGURE 1). However, to leverage the benefits of integration requires a shift in both thinking and practice. For example, in the past data tended to be homogeneous and structured, now it is becoming heterogeneous and unstructured—as well as there being more and more of it due to the emergence of pervasive computing, social networks, and high-throughput measurement. With the increase in data volume and complexity comes the need for analytical methods and tools to synthesize it—resulting in less correlative and more predictive approaches to effecting clinical practice and health outcomes. Finally, the practice of research and medicine is becoming less individual and more collaborative in nature—with interactions taking place in near real time on a global basis. Ultimately, this shift in thinking and practice will facilitate discoveries and advances in the interconnected ecosystem areas of life science research.

Figure 1 Advances in drug discovery, drug development and clinical care promise a new generation of preventative and preemptive healthcare, but countless data and knowledge "disconnects" continue to result in delays, errors, and poor clinical outcomes. To address these challenges organizations engaged in basic research, biomedical research, health research, and medical research are beginning to look to interoperable software tools, standards, databases, and the web to accelerate the shift to personalized medicine.

An integrated ecosystem, enabled by Web 2.0 technology can facilitate:

- Knowledge Acquisition and Transformation
- Analysis, Visualization, and Sharing
- Collaboration and Publication

Along with enabling reproducibility, aiding in discovery, and accelerating and transforming medical and healthcare research.

For example (FIGURE 2), integrated research systems can enable life science organizations to better leverage discovery research opportunities with research organizations conducting basic research. Enhancing lead discovery and optimization, leveraging the results of high throughput screening, ADME and toxicology. Healthcare organizations can leverage results and findings of life science organizations in the area of comparative effectiveness, or health outcomes with care delivery organizations. Enabling better safety and pharmacovigilance, evaluating the effectiveness of therapies, enhancing patient outcomes, disease prediction, and clinical decision support. Those same care delivery organizations can facilitate the conduct of Translational Research with organizations conducting basic research. Leveraging genotyping and diagnostic testing, as well as aiding in target biomarker discovery, and facilitating protein expression. Finally, all organizations can leverage data from observational studies and clinical trials to advance their overall understanding of disease and its treatment.

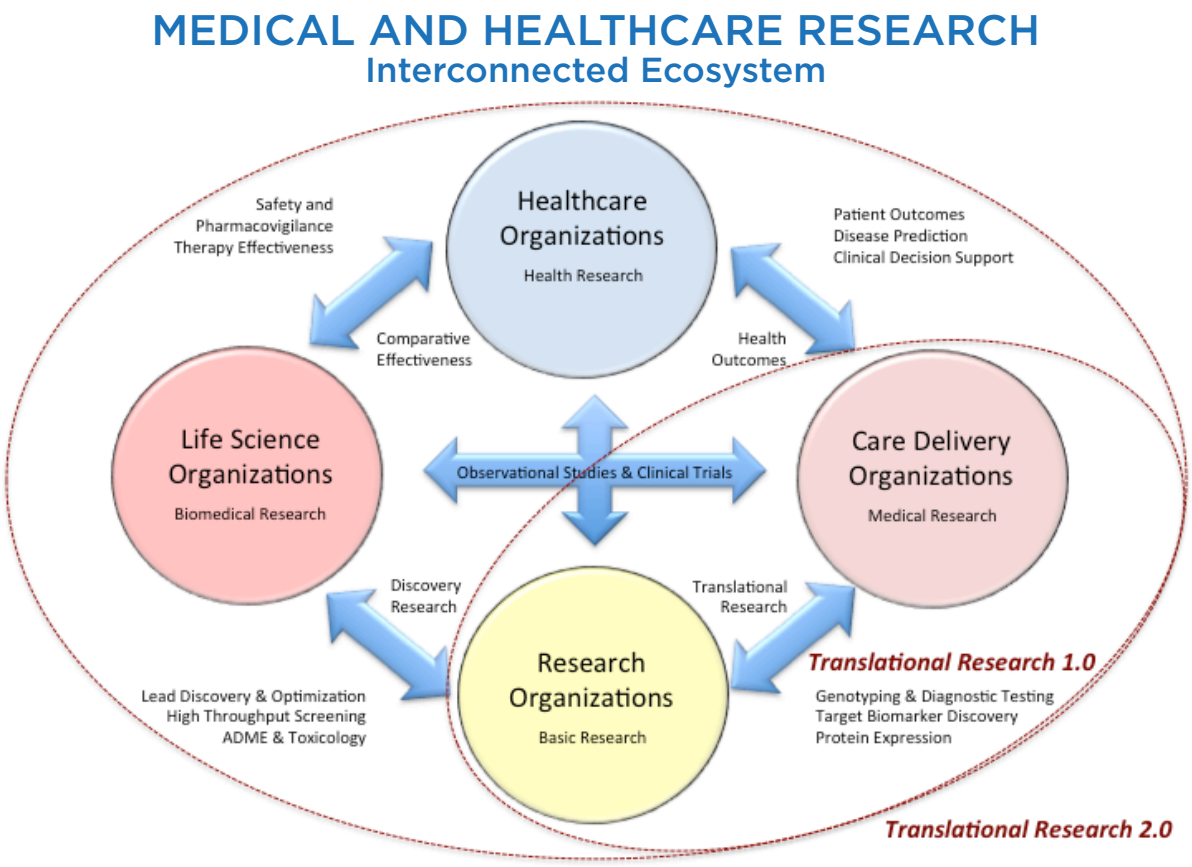


Figure 2 Translational Research 1.0 focuses on taking discovery from basic research, demonstrating their safety and effectiveness in clinical trials and than applying them as part of clinical practice in the delivery of healthcare. Translational Research 2.0 enhances the process of leveraging healthcare data from all ecosystem stakeholders in order to facilitate both the discovery and translation, as well as the health outcomes associated with new products and therapies.

Research productivity is sometimes described as though it is a concern "only" for the pharmaceutical sector and its investors. However, gaps in medical knowledge surrounding medical treatment and therapy affect healthcare consumers as well. Moreover, basic research institutions, which establish the core knowledge that underpins most pharmaceutical discovery, also face challenges and struggle to maintain research output. Increasing productivity is critical to hospitals as well, as they find themselves competing more and more for clinical research dollars and physician researchers. Various types of organizations participate in basic research, clinical trials, and the delivery of patient care. Bringing new discoveries to market requires a high degree of interdependence and knowledge sharing among care delivery organizations, life science organizations, and biomedical research organizations as well as healthcare organizations – these entities represent a "knowledge ecosystem" (TABLE 1) that, when functioning and cooperating optimally, can be incredibly productive.

KNOWLEDGE ECOSYSTEM

	Research Organizations	Care Delivery Organizations	Life Science Organizations	Healthcare Organizations
Audience	Academic, Industry, and Government	Hospitals and Academic Medical Centers	Pharmaceutical, Biotechnology, Medical Devices, and Diagnostics	Payer, Provider, and Government
Member Examples	MD Anderson, Quintiles, NCI, etc.	MD Anderson, Mayo Clinic, Sloan-Kettering, etc.	Pfizer, Genentech, J&J, Roche Diagnostics, etc.	United Healthcare, Kaiser, Veterans Health Administration, etc.
Research Focus	Basic Research	Medical Research	Biomedical Research	Health Research
Knowledge	Clinical Outcomes	Clinical Outcomes	Clinical Outcomes	Clinical Outcomes
Translational Research 1.0	Translational	Translational	Comparative Effectiveness	
	Discovery		Discovery	
Translational Research 2.0		Health Outcomes		Health Outcomes

Zerhouni's "US Biomedical Research: Basic, Translational, and Clinical Science" [1] lays the foundation for the multi-directional translation of information and knowledge to facilitate Translational Research across the entire ecosystem—with the goal of closing the time gap between a scientific discovery through basic research at "the bench" to the clinical level or the patient's "bedside"—the basis of

Translational Research 1.0. In addition, Zerhouni highlights the importance of social networking among scientists, scientific workflow management, logical data warehousing, scientific relationship management, data and specimen management, as well as tools in achieving this time gap goal—setting the stage for Translational Research 2.0.

Translational Research involves the translation of knowledge and evidence from "the bench" (e.g., laboratory-based discoveries) to "the bedside" (e.g., clinical or public health interventions informed by basic science and clinical research), and reciprocally from the bedside back to the bench (e.g., basic science studies informed by observations from the point-of-care).

To date, bench-to-bedside efforts, as noted in Marincola's "Translational Medicine: Two Way Road" [2], have been limited because the scientific aspects are poorly understood by clinicians and the needs of the patient poorly appreciated by basic scientists. In a world of Big Data, fueled by pervasive computing, high-throughput measurement, and social networking efforts to improve and leverage healthcare knowledge across the ecosystem are becoming increasingly important.

Translational Research 1.0 has used the metaphor of bridging the gap to cross the "valley of death." To advance to the next level and fulfill the promises of personalized medicine the bridge model must be replaced by a hub model where Translational Research 2.0 is the center of the wheel with traditional life sciences verticals as nodes in the wheel. Translational Research 2.0 is the recognition that many-to-many collaborations in no predetermined order is the path to research groups connecting in novel ways and for large scale multicenter studies to manifest.

Translational Research 2.0 leverages key web technologies to facilitate the storage, retrieval, sharing, and optimal use of biomedical information, data, and knowledge for problem solving and decision-making. These technologies, when coupled with key components of the interconnected ecosystem, can be used to facilitate scientific discovery, clinical research, and the leveraging of knowledge in providing evidence-based care.

Laying the foundation for Translational Research 2.0 can be seen in the efforts of the NIH to provide funding and support for infrastructure that supports team research—acknowledging that effective scientific teams of the future require closer working relationships among basic, translational, and clinical scientists. However, the barriers to effective Translational Research must be overcome as we move from the correlative science of the past to the more predictive science of the future. Issues around research complexity and the need for greater cross-disciplinary collaboration must be addressed in order to tackle the problem of integrating biomedical knowledge across the ecosystem.

The benefits, challenges, and opportunities afforded by integrating biomedical knowledge across the ecosystem are many. At a high level, the essential information management challenges to be addressed can be classified as belonging to one or more of the following categories:

- DATA (The ability to collect and manage heterogeneous data sets with increasing levels of dimensionality): With the ever-increasing availability of high-value, patient-centric phenotypic data sources, such as electronic health records (EHRs), clinical trials management systems (CTMS), as well as bio-specimens and their corresponding bio-molecular measurements such as genotypic and proteomic expression

profiles fed by a growing suite of instrumentation platforms, the size and complexity of data sets that researchers can collect, store, and retrieve on a regular basis are growing at an exponential rate.

- METHODS (The need to employ knowledge-anchored methods to discover and test hypotheses concerning linkages between phenotypic and bio-molecular variables of interest): Given the high-throughput data sets described in the preceding challenge, a corresponding high-throughput hypothesis discovery and testing challenge also exists.

- KNOWLEDGE (The provision of systematic and extensible platforms capable of expediting data integration and analysis workflows): Integrating biomedical informatics and Translational Research requires the availability of systematic data-analytic "pipelining" tools that are capable of supporting the definition and reuse of data analysis workflows incorporating multiple source data sets, intermediate data analysis steps and products, and output types.

- COLLABORATION (Dissemination of evidence and knowledge): Integrating clinical or Translational Research and biomedical knowledge and leveraging Web 2.0 technology to disseminate the evidence and knowledge gained in a resource efficient and timely manner. The goal of scientific collaboration is to enhance scientific discovery, by having more people coordinate their work, use expensive instruments remotely, and to engage people from diverse disciplines and backgrounds. In order to accomplish this goal, the work that is done must be broken down into manageable pieces, and the community conducting it must possess a common understanding of what needs to be done.

In their paper entitled "Alzforum and SWAN" [3] Tim Clark and June Kinoshita note that knowledge acquisition in science research proceeds in a cycle—from hypothesis development; through experiment and data collection; to interpretation and drawing of conclusions; to communicating results to other scientists; to assimilating, criticizing and synthesizing the communications of colleagues—within a specific area of research and across areas of research. As a scientific web community, Alzforum contains more than 40,000 literature citations, 1,300 news articles, 4,000 comments, 10,000 antibodies, 200 research models, 350 genes from published association studies of late-onset Alzheimer's Disease (AD). The forum has more than a million visits per year—clearly a thriving scientific web community.

Bennett's "Collaboration and Team Science: Field Guide" [4] discussion of the Collaboration Continuum, re-enforces the work of Pennington, on "Cross-Disciplinary Collaboration and Learning [5], and Bos et.al.'s "Science on the Internet" [6]. Each highlights the principal barriers to successful collaboration, namely:

- Scientific Knowledge is Difficult to Aggregate
- Scientist work independently and informally
- Scientific Work Across Institutions is Complicated

They discuss how scientific information is transformed into knowledge—due to discourse (Alzforum), and how hypotheses evolve into theory (AlzSWAN). The AlzSWAN (Alzheimer's Semantic Web Applications in Neuromedicine) knowledge base builds on Alzforum's social network, to construct a semantically structured network of hypotheses, claims, dialogue, publication and digital repositories—all part of an "information-ecosystem." Bos et.al. stresses how much of the creative activity in science occurs through informal discourse as scientists

interpret and critique new findings, puzzle over discrepancies, and integrate data to formulate new ideas and hypotheses—scientific web communities benefit from greater transparency and access to valuable information (e.g., An experiment might stall or a researcher might waste months traveling down a blind alley, unaware that someone in a different field may have found the answer.)

Key to Alzforum's success is the site's ability to nurture productive discussion of ideas and hypotheses, hosting community data repositories of "publically available data only," because of the significant opportunities to add value to public data by providing a system to curate and organize it around the community's interests—AlzGene's comprehensive database of all published genetic association studies for late-onset AD. Alzforum leverages a research process model which proceeds in a cycle from: hypothesis development, through experiment and data collection, to interpretation and drawing of conclusions, to communicating results to other scientists, to assimilating, criticizing and synthesizing the communications of colleagues—as part of a Knowledge Ecosystem. AlzSWAN incorporates a large part of the biomedical research lifecycle in its ontological model, including support for personal data organization, hypothesis generation and digital pre-publication collaboration.

At its core, Translational Research 2.0 facilitates collaboration across multidisciplinary, team-based clinical and Translational Research projects by assisting in:

- Identifying major categories of information to be collected, managed, and disseminated throughout the clinical or Translational Research process and the ways in which they relate to one another, thus enabling the development of integrative platforms capable of addressing such needs in a systematic manner.
- Providing individual researchers with the ability to understand how their unique activities contribute to a broader goal of generating new knowledge or evidence that spans multiple domains or sub-domains, thus increasing awareness of the needs to exchange or disseminate such information in an easily and readily consumable manner.

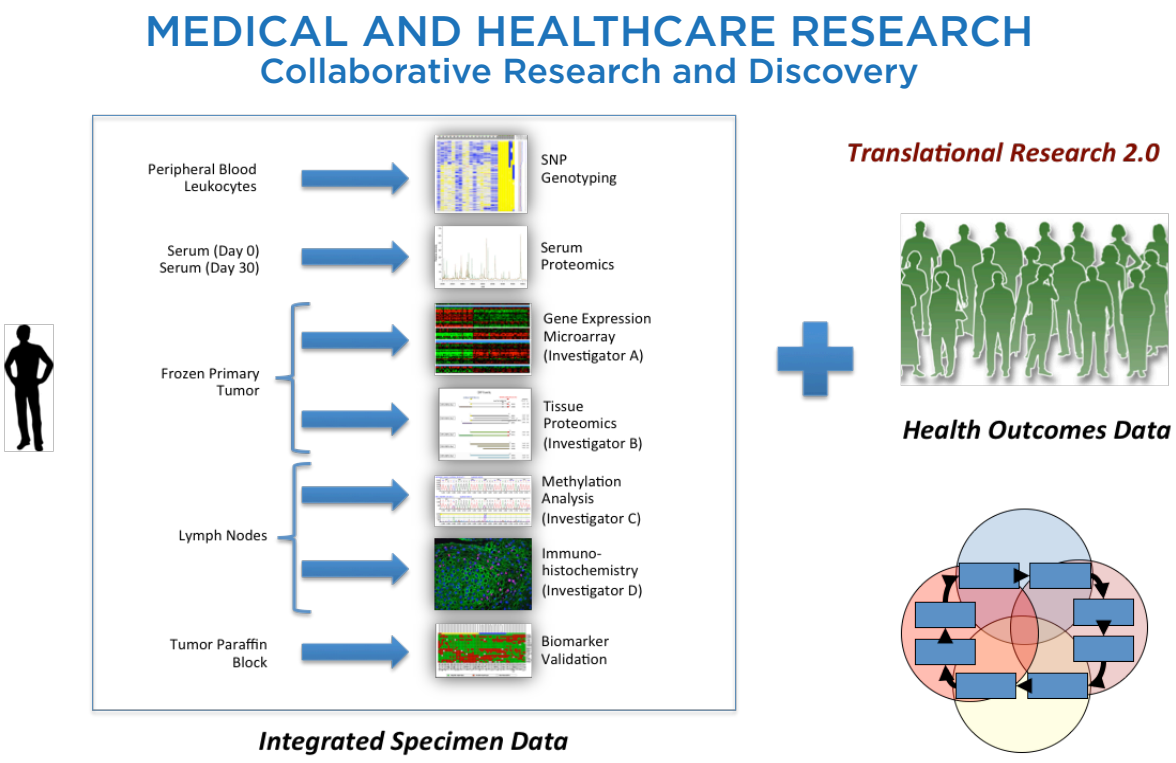
Pennington's "Cross-Disciplinary Collaboration and Learning" highlights the concept of the Knowledge Ecosystem as well as the need for ecosystem schemas to mesh to facilitate cross-disciplinary collaboration. This meshing facilitates an exchange of knowledge in ways that are conducive to making sense of a subject without requiring depth of understanding. Pennington suggested that the development of collaborative solutions to a complex problem requires a two-phased approach. Phase 1 is an idea generation phase with leverages collective thinking. Phase 2 is an idea implementation phase that requires a combination of convergent and divergent thinking. Finally, cross-disciplinary learning in organizations requires a form of "organized learning," where all of the stakeholders in the collaboration are the learning organization.

An example of this type of cross-disciplinary collaboration and learning with respect to disease modeling can be seen at the Center for the Development of a Virtual Tumor (CViT), part of the National Cancer Institute's Integrative Cancer Biology Program. Here CViT is building an ever-growing community of researchers around the world dedicated to computational and mathematical cancer modeling. Its online outlet, CViT.org, currently provides participants with all the tools of a community-driven website: wikis, blogs, forums, member profiles, and RSS-based news updates. Deisbeck's "CViT" [7] discusses how integrative cancer biology research relies on a variety of data-driven computational modeling and simulation methods and techniques geared towards gaining new insights into the complexity of biological processes of critical importance to cancer research—Like Alzforum, CViT provides the enabling technologies to foster multi-scale cancer modeling and simulation.

Another example, Nielsen and Gower's "Massively Collaborative Mathematics" [8], highlights the potential of "Open Source" biology and mathematics, specifically as a model for collaboration and aggregation of insights from people with diverse backgrounds. The "Open Source" model (also seen in Alzforum) allows ideas to be explored from many different perspectives and allows for unanticipated connections to be made—an open source approach leverages a broader organization around issues. Leveraging an aggregation of insight from people with diverse expertise—and open data sharing allows for open data analysis extending the limits of human problem solving ability.

As the analysis of widely distributed data becomes more commonplace, integrated research systems and Web 2.0 technologies will be needed to facilitate the process of data and knowledge collection, management, and use. In today's environment (FIGURE 3) there is increasing need for integration. Data and the experimental apparatus for collecting it typically belong to organizations with different objectives as to the data's use.

Within these organizations, various individuals are responsible for different aspects of data acquisition, processing, and analysis. Sometimes, entire projects involve collaborations across organizational boundaries. Translational Research 2.0 requires a foundation upon which large diverse heterogeneous data sources (containing genomic, phenotypic, and outcomes data) can be integrated and analyzed so as to accelerate the pace of scientific discovery, the conduct of clinical research, and the leveraging of knowledge in providing evidence-based care.



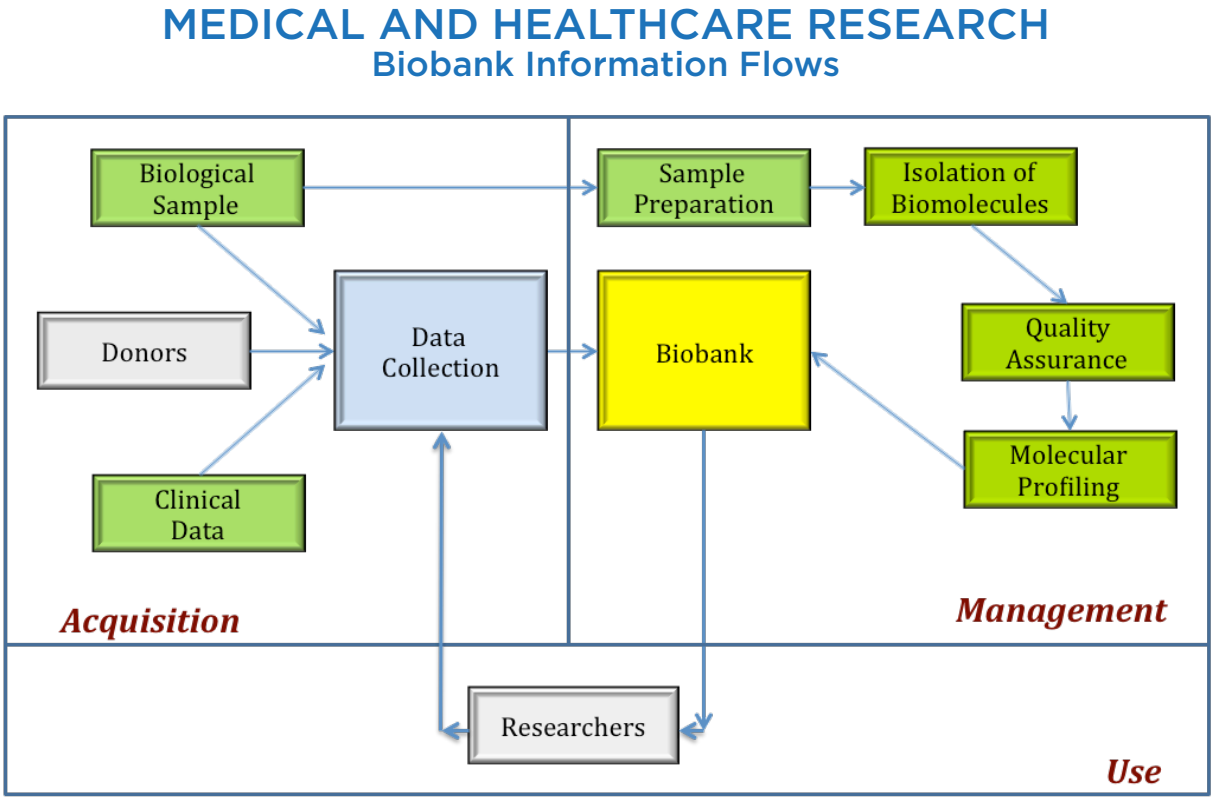
Price Waterhouse Coopers (2008) Research Rewired: Merging Care and Research Information to Improve Knowledge Discovery. Price Waterhouse Coopers Health Research Institute. Available at: http://www.pwc.com/us/en/healthcare/publications/research-rewired.jhtml

FIGURE 3 One of the primary challenges facing scientists today is the need to synthesize the high volume information pouring in from many sources, disciplines, and specialties. The speed of discovery is greater than ever, and the ability to access, interpret, and synthesize information is an important component in the conduct of research. Web 2.0 applications can provide that critical bridge to that access. These tools allow users to find their affinity groups and potential collaborators.

BIOBANKS AND TRANSLATIONAL RESEARCH

With integrated access to clinical and research data, investigators can explore research questions or plan studies that minimize duplication, maximize the potential for valuable results and encourage cross-institutional collaboration. None of this is possible, however, without biobanks that provide high quality biospecimens. Analysis of 125 biomarker discovery papers published between 2004 and 2009 found that more than half included no information about how specimens had been obtained, stored or processed. A survey conducted in 2011 of more than 700 cancer researchers found that 47% had trouble finding quality samples, 81% limited scope of research, and 60% question the findings of their studies.

Momentum in building biobanking resources with well-annotated clinical specimens is growing, with the emerging recognition that biobanks are a critical pillar in facilitating and accelerating Translational Research discoveries.



Z. Zimmerman, et.al. (2004) BIOBANKS: Accelerating Molecular Medicine Challenges Facing the Global Biobanking Community. IDC Special Study. IBM Life Sciences. November 2004. at http://www03.ibm.com/industries/global/files/Biobanks_Accelerating_Molecular_Medicine.pdf

FIGURE 4 Large multi-center clinical studies often involve the collection and analysis of biological samples. This analysis is dependent on timely, complete, and accurate recording of analytical results and associated phenotypic and clinical information. Research to improve healthcare is increasingly supported by advances in genomics, proteomics, and metabolomics. To allow statistically meaningful analyses, all of these methodologies demand large numbers of adequately collected and annotated biospecimens from both diseased and non-diseased individuals. Well-managed biorepositories provide for receipt, storage, processing and/or distribution of bio-specimens through standardized operating procedures, along with management of their associated data.

Research on disease-associated, as well as healthy samples is being used to elucidate the relationship between genotype and phenotype and help identify the genetic cause of disease and health. Molecular profiling requires high quality biological samples with associated clinical information in order to form the basis for genetically informed medicine. Biobanking is vital to closing the loop in this model of research, feeding back into research relevant biological samples, associated pathology reports, molecular characterization, and the associated treatments and clinical outcomes. Translational Research 2.0 facilitates the leveraging of research, patient, and clinical data across the eco-system to drive the development of targeted interventions and to facilitate the practice of genetically informed medicine at the point of care (FIGURE 5).

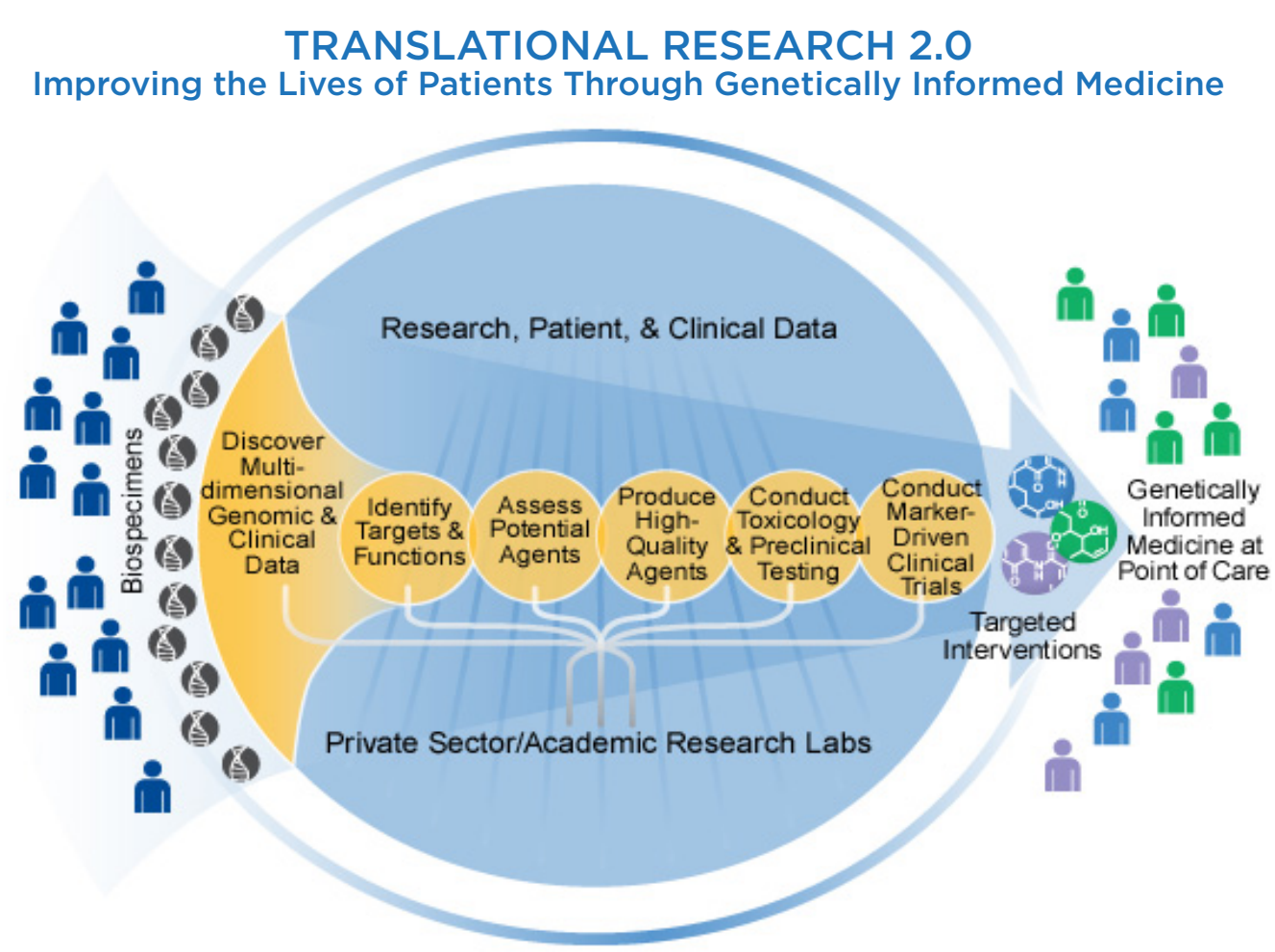


FIGURE 5 A proposed model of medicine that proposes the customization of healthcare, with decisions and practices being tailored to the individual patient by use of genetic or other information. Source http://www.cancer.gov/aboutnci/servingpeople/cancer-research-progress/therapeutic-platform

The combination of biobanking, high throughput research together with bioinformatics and genomics has proven to be a powerful approach in biomedical research. It opens up many opportunities in the field of diagnostics, predictive factors, biomarkers, pharmacogenomics, drug discovery, and understanding of disease mechanisms, thereby giving the potential to have a major impact on prevention, diagnosis, prognosis, prediction, and treatment of human diseases. To fully exploit this revolution in the combat of human disease it is vital to have access to large quantities of human tissue samples of high quality, ideally both diseased and unaffected fresh unfixed tissue samples from patients and unaffected tissues from healthy control individuals. The Translational Research process requires effective acquisition, management, and use of biospecimens, and their corresponding genomic and phenotypic information—Translational Research 2.0 enhances the way these specimens can be synergistically leveraged across the ecosystem. Translational Informatics enhances how the resulting data and its integration with other sources of information can be leveraged to advance ecosystem knowledge and aid in discovery.

TRANSLATIONAL INFORMATICS AND BIG DATA

Translational Informatics is the application of informatics theory to Translational Research, which seeks to translate scientific discovery into practical applications to improve human health. Carrying out this objective requires both facilitating the process of discovery (hypothesis generation and testing) in biomedical research and the identification and adoption of best practice in healthcare (prevention and treatment).

Technological developments over the past two decades have increased, by orders of magnitude, the capability for molecular data generation. The ability to produce large quantities of clinical and biological data promises unprecedented advances for diagnosis and treatment of human disease, but appropriate utilization of large-volume molecular data sets mandates changes in traditional paradigms of data generation, analysis, and interpretation. The discipline of translational medicine has been developed, in part, because of the recognition that resources and efforts must be focused on producing clinically relevant research in a timely fashion. An important challenge for Translational Researchers in today's Big Data environment of comprehensive and noisy molecular data sets is developing analytic strategies (TABLE 2) capable of constructing clinically useful tools from this wealth of data.

CASE STUDY

Researchers at the BloodCenter of Wisconsin's Blood Research Institute needed a single system to manage an NIH-funded research project spanning more than 20 globally distributed research sites and disparate data sources, such as patient registries, sample data, biorepositories, lab test results, and clinical findings.

Remedy delivered Investigate™, the world's first and most powerful Integrated Research Management System to connect clinical, laboratory, and biobanking data types, and streamlines the capture, querying, and reporting of complex, multicenter research activities so that researchers can rapidly correlate results, discover patterns, and act on findings.

With Investigate, researchers at BRI have unprecedented analytical power and have been able to manage and interact with their data in real time. Implementing a Translational Research 2.0 tool like Investigate will researchers at BRI an edge as they continue their search for answers and subsequent scientific discoveries.

To read more, visit Remedyinformatics.com/case-studies

Traditional biomedical research conducted under conditions in which the number of samples that make up the study population (n) exceeds the number of variables (p) being investigated. For example, the well-known Framingham study has tracked several hundred variables and potential cardiac risk factors in a population of more than 15,000 patients (n>>p). Recent technological developments in molecular biology allow collection of thousands of biological data points per sample thereby exceeding the total experimental population size by several orders of magnitude (p>>n). Nevertheless, statistical power remains a barrier to research studies with insufficient access to samples and research subjects.

TRANSLATIONAL INFORMATICS AND BIG DATA

Translational Research 1.0	Translational Research 2.0
Categorical Variables	**Continuous Variables**
Clinical researchers often perceive categorical (discrete) variables to be less complex than continuous variables, in theory and application – which leads to the common practice of "discretizing" continuous variables.	Most modern approaches to biomedical data modeling are designed for use with continuous variables.
Discrete Modeling	**Probabilistic Modeling**
Clinical models designed to predict discrete diagnostic or therapeutic classes traditionally have been considered more practical than probabilistic models – neither require modeling based on probability distributions nor mandate consideration of the potential effects of randomness.	Involves shifting from deterministic strategies using discrete classes to modeling approaches based on probability distributions – cancer is a prototypic stochastic system, in which accumulations of random mutations are associated with differential frequencies of disease, prognosis, and therapeutic response.
Linear Modeling	**Nonlinear Modeling**
Linear modeling strategies are commonly used in modern translational analyses. The underlying assumption – regardless of complexity is that the fundamental relationships among the data being modeled are linear.	Nonlinear modeling strategies are beginning to be developed and tested for use with large-volume biological data sets. Nonparametric Bayesian approaches are one example of nonlinear modeling strategies that have been successfully applied to diagnosis and survival prediction.
Homogeneous Data Sets	**Multimodal Data Sets**
As the type and volume of biomedical data increase, databases integrating multiple data modalities must be constructed and formatted prospectively for use in translational investigations	Constructing hybrid data sets containing both clinical and biological data require strategies to ensure semantic compatibility across heterogeneous data sources, as well as, integration of molecular and clinical data for translational hypothesis generation.
Finite Statistical Analysis	**Computational Learning**
Most classical models of scientific investigation are finite in their design and conclusions: A question is asked and a hypothesis is constructed; an experiment is designed; data are generated and analyzed; and an optimal answer for the question is produced based on discrete analysis of the finite data set.	Computational learning includes both artificial intelligence and machine learning – an approach well suited to translational medical application because of their ability to identify patterns within large data sets, to make predictions from new samples and to improve their predictive accuracy over time.

TABLE 2 Marko, N.F. and Weil, R.J. (2010) Mathematical Modeling of Molecular Data in Translational Medicine: Theoretical Considerations. Science of Translational Medicine. 3 November 2010. Vol. 2 Issue 56. Pages 1-6. Schadt, E.E., et.al. (2010) Computational Solutions to Large-Scale Data Management and Analysis. Nature Reviews: Genetics. September 2010. Volume 11. Pages 647-657.

Modern definitions of "Translational Research" may be as simple as the "bench-to-bedside" concept, as complex as multipart definitions highlighting the discipline's focus on developing practical therapeutics—Translational Research 1.0, or as broad as definitions including public health and regulatory implications—Translational Research 2.0.

Translational Informatics in the context of Translational Research 2.0 leverages new tools and techniques appropriate for the types of Big Data found in the healthcare ecosystem (TABLE 3). For example, organizations that interact in the ecosystem share several common objectives as they relate to Translational Research 2.0, namely an ability to leverage translational informatics and to better predict disease, support medical decision making, ensure patient safety, as well as improve health outcomes.

TRANSLATIONAL INFORMATICS AND HEALTH ANALYTICS

Healthcare Providers	Life Science Organizations	Academic Medical Centers	Public Health Organizations
Disease Prediction	**Trial Planning, Monitoring, and Review**	**Disease Modeling**	**Safety and Pharmacovigilance**
Prediction of disease based on phenotypic and genotypic risk factors.	Facilitate decisions regarding clinical trial conduct.	Patient/disease risk modeling and analysis.	Mining and predictive modeling of compound safety both pre and post approval.
Clinical Decision Support	**Trial Operations**	**Target and Biomarker Discovery**	**Health Outcomes**
Support for medical decision making at the point of care.	Facilitate decisions regarding enrollment patterns and key performance indicator monitoring for sites.	Biosimulation and predictive modeling of the effect of drug candidates on disease targets and biomarkers.	Evaluation of therapy effectiveness in target populations.
Patient Outcomes	**Safety and Pharmacovigilance**	**Protein Expression**	**Comparative Effectiveness**
Identify patient risk patterns and effectiveness of associated prevention programs.	Mining and predictive modeling of compound safety both pre and post approval.	Facilitate analysis of protein expression data and identification of new protein with similar expressions.	Evaluation of comparative effectiveness of target compound with other compounds in its class.
	Health Outcomes	**Genotyping and Diagnostic Testing**	
	Evaluation of therapy effectiveness in target populations.	Identification of genotypes of interest based on experimental data.	

TABLE 3

While the healthcare industry has historically been slow to turn to analytics, leading practitioners now realize data is at the center of informed and precise decision-making that can ensure their organizations' future. Healthcare delivery is complex—so is medical research. Today's ad hoc methods of managing research information are beginning to strain under increasing demands for new drugs, more personalized medicine, better diagnostic tools, and post-market safety monitoring. A range of organizational and technical barriers must be overcome to enable the frictionless information exchange among all appropriate participants that is necessary for progress in providing "smart" healthcare.

The twin challenges in Translational Informatics:

1. Aggregate, map, and harmonize disparate data types ranging from the clinical (patient demographics, medical, surgical, drug histories) to the laboratory (assays, genomics, proteomics, metabolomics), to the biobank (samples, handling, processing, storage, quality).
2. Separate signal from noise by converting data (raw bytes) to information (analyzed) to knowledge (interpreted) to wisdom (clinically actionable).

Big Data is orthogonal to the twin challenges and manifests in data levels (TABLE 4). The challenge in Big Data is to get past Level 1 as quickly as possible to reach higher levels of data sophistication where information is compressed and knowledge and wisdom can be derived from raw data.

REMEDY INFORMATICS

TRANSLATIONAL INFORMATICS
Data Levels

Data Level	Description	Example
Level 1 Raw	Low-level data for a single sample, not normalized across samples, not interpreted for molecular abnormalities	Sequence trace file; Affymetrix CEL file
Level 2 Processed	Data for a single sample that has been normalized and interpreted for the presence or absence of specific molecular abnormalities	Mutation call for a single sample; amplification/deletion for a gene in a sample; expression of a variant
Level 3 Segmented Interpreted	Processed data for a single sample further analyzed to aggregate individual loci into larger contiguous regions	Amplification/deletion for a region in a sample
Level 4 Summary Finding	Association across samples, among molecular abnormalities, sample characteristics, or clinical variables	A finding that a particular genomic region is amplified in 10% of all samples

TABLE 4 Source: https://tcga-data.nci.nih.gov/tcga/tcgaDataType.jsp

Translational Research 2.0 seeks to translate new basic scientific knowledge—such as developments in genomics—into enhanced clinical practice. At its foundation is the need to find new ways of enabling collaboration, facilitating discovery, and accelerating the diffusion of new knowledge throughout the healthcare ecosystem and into common clinical practice.

Many leading-edge organizations in the ecosystem have evolved to make this a reality, others are in the process of navigating this shift, and the majority will recognize how valuable an integrated solution can be in advancing medical and healthcare research not long after their predecessors. The goal of this white paper is to help facilitate and support this evolution in Translational Research to the next level.

For information about how Remedy Informatics can help your institution develop a highly customized registry in a fraction of the time and for a fraction of the cost of in- house development, speak with one of our consultants today for a free, no obligation consultation by calling (801) 733-3300 or emailing info@remedyinformatics.com.

REFERENCES

1. Zerhouni, E. (2005) US Biomedical Research: Basic, Translational, and Clinical Sciences. JAMA, Vol. 294, No. 11, Pages 1352-1358.
2. Marincola, F. (2003) Translational Medicine: A two-way road. Journal of Translational Medicine Editorial, downloaded from: http://www.translational-medicine.com/content/1/1/1.
3. Clark, T. and Kinoshita, J. (2007) Alzforum and SWAN: The Present and the Future of Scientific Web Communities. Briefings in Bioinformatics Vol. 8, No. 3, Pages 163-171.
4. Benett, L.M., et.al. (2010) Collaboration and Team Science: A Field Guide, National Institute of Health, April 2010, Available at: https://ccrod.cancer.gov/confluence/download/attachments/47284665/TeamScience_FieldGuide.pdf.
5. Pennington, D. D. (2008) Cross-Disciplinary Collaboration and Learning, Ecology and Society, Vol. 13, No. 2, Available On-line at: http://www.ecologyandsociety.org/vol13/iss2/art8/.
6. Bos, N., Olson, G.M., and Zimmerman, A. (Eds.) (2008) Science on the Internet. Cambridge, MA: MIT Press.
7. Deisboeck, T., et.al., (2007) Advancing Cancer Systems Biology: Introducing the Center for the Development of a Virtual Tumor, CViT. Cancer Informatics. Special Issue. Pages 1-8.
8. Nielsen, M and Gowers, T. (2009) Massively Collaborative Mathematics. Nature, Vol. 461, No. 15, Pages 879-881.