Changing the Model in Pharma and Healthcare – Can We Afford to Wait Any Longer?

Erich Alfred Gombocz*

IO Informatics Inc., Berkeley, California, USA
egombocz@io-informatics.com

Abstract. Innovations in healthcare delivery and Pharma require re-examination of process models at the foundation of our knowledge discovery and clinical practice. Despite real-time availability of 'big data' from ubiquitous sensors, mobile devices, 3D printing of drugs, and a mind shift in data ownership, data integration still remains one of the core challenges to innovation. Increasingly persistent, semantic data integration is gaining recognition for its dynamic data model and formalisms which make it possible to infer from and reason over interconnected contextualized data, creating actionable knowledge faster and at lower cost. While such technical advances underpin the successful strategies to drive positive patient outcomes or accelerate drug design, there are equally profound social changes towards the willingness of patients to share their own data - opening doors to new patient-centric, precision-medicine healthcare models. Adding astronomically rising costs in research and healthcare, we have arrived at a critical turning point where it is now well within our reach to change how drugs are developed, how trials are performed and how patients are treated - and we can do this with huge benefits for otherwise unsustainable industries. Examples show that not only is this possible today, but that such approaches already have traction; (i) in Pharma for assessing impact of excipient on drug stability and efficacy; for pre-clinical toxicity assessment and integral systems views on drug safety, (ii) in Government at the FDA's cross species biomarker initiative to reduce animal testing and (iii) in Health Care for organ transplant rejection assessment and COPD. Using comparative effectiveness and side effect analyses to base treatments on solid prognoses and therapy decision support, we can and must change discovery and healthcare into a data driven and patient centric paradigm. The socio-economic benefits of such a change will be enormous.

Keywords: life sciences, big data, sensors, data ownership, semantic integration, actionable knowledge, patient centric, precision medicine, decision support, use cases, socio-economics.

* Corresponding author.

C.J.O. Baker, G. Butler, and I. Jurisica (Eds.): DILS 2013, LNBI 7970, pp. 1–22, 2013.

1 Introduction

1.1 Historic Models in Life Sciences

In the past, widespread utilization of large shared spreadsheets, dedicated laboratory information management systems (LIMS), large relational data warehouses and traditional methods for extraction, translation and loading (ETL) have been used across the life sciences enterprise spectrum with more or less sophisticated approaches to interconnect in-between some of those resources [1]. Key features of LIMS include acquisition, workflow and data tracking across different modalities, data exchange interfaces, audit functions and support for their use in regulated environments. Because of rapid pace at which laboratories and their data management needs shift, the definition of LIMS has become more blurred. This is particularly due to the fact that the needs of laboratories widely vary which requires also a shift in functionality of laboratory information management systems.

Historically, LIMS and process execution have performed similar functions, building an organization's reference backbone for experimental results. More recently, assay and ELN functions have been added to extend traditional LIMS systems. However, the need to implement quality standards, the awareness of data management solutions using different architectures and the unavailability of adapted solutions for interoperability led in many cases to in-house developments instead of using commercial solutions. Particularly in large Pharma organizations the separation of data into target areas, specific projects as well as the separation of R&D chemistry, assay development and biology caused limited communication in-between groups, redundant efforts and no integral view across the data. The strict separation between pre-clinical, clinical and market data has hampered feedback within the organizations to learn from past experiences. Consequently, adverse effects got missed; clinical trial efficiency was at a low point and causing a hesitant approach in the development of new drugs.

Despite ever rising amounts of data through high throughput screening, multiplexed assays and broad use of chip technologies, the actual knowledge produced in comparison to research costs was declining rapidly [2]. Large Pharma companies were buying their libraries of new compounds from small biotech to cut costs by reducing in-house research to small focus areas. Collaboration models were restricted to consortia with narrow goals and small portions of pre-clinical, pre-competitive segments the sharing party deemed to be of no further usefulness to the organization.

1.2 Rise of New Technologies and Machines

The data landscape changed with the rise of new technologies, new developments in instrumentation, automation and exponential increase in throughput of previously labor-intense and time-consuming procedures. In the last several years, massive next generation sequencing (NGS), progress in whole genome sequencing using de novo assemblies on unimaginable scale [3], RNA sequencing and genome-wide association

studies (GWAS) have been at the forefront of genomics to be used for both, gene-based biomarker discovery and personal genomics as tool for precision medicine and the genetic selection of population cohorts for clinical trials.

The development of new sensor technologies and advances in mobile computing led to sensors being everywhere and on everything with real-time internet connectivity. Wearable medical technology is becoming a hot commodity [4]. As these devices come to market, they have great potential to help both patients and clinicians monitoring vital signs and symptoms [5]. In personal health, sensors which are always on, always with you, always tracking were changing data collection to become a continuous monitoring stream [6], providing both, individuals and physicians more accurate and more detailed data about many influence parameters on a health or disease state which previously were not available [7]. Lifestyle choices, such as exercises, habits and environments have been recorded similarly [8].

The size of all these data and the computational considerations to analyze them along with the high data dynamics require investment in High Performance Computing (HPC) and have led to tradeoffs between inexpensive highly dense storage on commodity disks and higher cost better performant NAS, SAN or Cloud services (CEPH, OpenStack, Amazon). Dependent on budgets, compromises were made, and raw data have been thrown out in favor of much smaller analyzed data sets. Algorithmic transformations to normalize in-between platforms have changed over time and metadata not always included, making review for verification in many cases impossible. While new ways of computing have been introduced which are using massive parallel computing and distributed clusters for analysis [9], management of 'big data' has become a complex, expensive and demanding task at scales beyond most forecast expectations. This development has created a new bottleneck in analysis and practical use of ever growing data repositories and made interoperability, provenance and versioning an equally important concern to plain connectivity and was instrumental in rethinking data integration in life sciences in general.

1.3 Economic Importance of Data

In a MIT Technology Review end of 2012, the question was raised publicly if personal data is the new currency [10]. The economic importance of access and utilization of vast amounts of interconnected data can no longer be denied, and the same applied equally to the life sciences. With the expansion of social networks and the drive from individuals to take care of their needs for better prognosis and treatment, the frustration about public availability of medical data has driven the movement of patients making their own data publicly accessible. Adding to this the fact, that big data analytics became a way of turning data into money [11], the assessment that Data is the new money and those who have access to it, have power became obvious with significant implications in the shift from revenue and margins driven industrial models towards customer-centric, health outcomes for patients motivated strategies in which consumers influence the drivers of healthcare systems.

This new proposition is affecting insurers, payers, service providers, drug discovery processes, drug development, repurposing of drugs for new indications alike – moving to an evidence-based, outcomes-focused, behavior-driven life sciences environment which through empowering of data will benefit all of us.

1.4 Staying Competitive

In the past, low ROI on research and development has provided little incentives to innovation or change, particularly as industry was closely watching its competitor's moves to decide about the necessity to adjust their model to new trends based on proof-of-concept (PoC) and pilot study outcomes. In a way, the unwillingness to share even pre-competitive or failed approaches within tightly controlled consortia members for collaborations has reduced the competitiveness of the industry – however, there a several positive examples on the horizon that this behavior is changing as both, Pharma and Healthcare industry have come to the realization that everyone profits from collaborative approaches to accumulative and complimentary data on common goals. Of course, any meaningful collaboration is closely tied to interoperability, and this is true equally for both, commercial and academic entities.

Crowd-sourced analysis requires interoperability, and interoperability is how big data becomes big open data, and this will assure rapid progression in scientific discovery and providing a solid foundation for Pharma and healthcare to stay competitive. Acknowledged, that crowd-sourcing as a new policy has many implications [12-13] and there is still hesitancy to collaborative data sharing, its driving force will be costs and efficiency towards new concepts which will significantly shape the future of data-driven life sciences.

2 State of the Industry

2.1 Data Generation vs. Knowledge Gain

While automation and advances in technologies have brought down costs of complex testing faster than anticipated, analysis and integration towards applicable knowledge has lagged behind. Massive data (a single run of Life Technologies sequencer produces ~900 GB raw data, 1 machine = 10 TB/day) and the sophistication required by the complexity of analysis procedures have let to the notion of the *"$1,000.- genome at $1 Mio interpretation.*"[14] – thus, in most cases, only a small fraction of information is used to build knowledge and advance scientific progress.

One major reason for this discrepancy is that the required proficiency of a whole range of experts including molecular and computational biologists, geneticists, pathologists and physicians with detailed knowledge of disease and treatment modalities, genetic counsellors and IT specialists to build analysis teams [15] is not easy to establish. Using large number of specialists was critical to complete the data analysis, for variant annotations and to interpret causative or actionable variants. Even then, clinical verification of such variants and ramifications for the treating physician and patient require even today immense efforts and make the widespread use of

clinical whole-genome sequencing for diagnosis and quality of life improvement still a distant goal. On top of this, the bioethics pros and cons of WGS of every newborn child need to be sorted out and genomics must address socio-economic disparities in healthcare.

Similar considerations should be applied to other rapidly evolving fields such as proteomics, transcriptomics and microbial implications in major diseases – rapid data generation through automated, low cost high throughput sample analyses does not match up with the possible knowledge gain from those resources. Amongst other reasons, standardization of analytical methods and algorithms and requirements for quality standards on data also has played a significant role in the usability of results across laboratories.

2.2 Traditional Data Mining

Relational data warehouses and object data bases require upfront considerations to determine which questions you want to answer, Data models (schemas) must be defined at the beginning, so such solutions, while excellent for final datasets and great performing on optimized queries for what they were built for are demanding in support due to their rigid and static structure. On the other hand, a whole host of mining solutions and visualization tools are available as relational database technology has been around for a long time and big players in information technology have embraced its use.

In Life Sciences, however, a different picture emerges as dynamic, agile solutions are required to keep pace with changing data types, formats, instrumentation as well as analytical requirements. Add to this the scale of growth, and 'big data' has demonstrated impressively, that more relational data warehouses and traditional data mining approaches cannot be the answer to today's information requirements landscape. In many cases in Life Sciences, questions to ask and potential use cases are moving targets, so any inflexible solution limits its applicability. In biological systems, the need to traverse data, to infer from other data and to search complex pattern across all your resources to find clues what kind of questions you can answer is rooted on a different set of requirements - in most cases, questions are not predefined, and the picture what and how to ask is not clear at the beginning. Include to this that in the relational world no clear connections in-between data silos exist and different proprietary schemas prevent cross-resource queries, and the limitations of such approaches become transparent.

2.3 Cost of Research vs. Outcomes

A recently published Forbes report [16] on staggering costs of new drugs, a new study comparing healthcare cost in the US with other countries [17] and the OECD Health Statistics [19] provide insights into costs of research versus outcomes which are stunning, but well known within the industry. A representative of Eli Lilly estimated the average cost of bringing a new drug to market at $1.3 billion, a price that would buy 371 Super Bowl ads on television [16]. On average, a drug developed by a major pharmaceutical company costs at least $4 billion in R&D (see Table 1 below)

Table 1. Number of approved drugs and drug development costs of major Pharma companies (2012)

Company	Approved Drugs	R&D Costs/Drug [$ Mio]
AstraZeneca	5	11,790.93
GlaxoSmithKline	10	8,170.81
Sanofi	8	7,909.26
Roche AG	11	7,803.77
Pfizer Inc.	14	7,727.03
Johnson & Johnson	15	5,885.65
Eli Lilly & Co.	11	4,577.04
Abbott Laboratories	8	4,496.21
Merck & Co Inc	16	4,209.99
Bristol-Myers Squibb Co.	11	4,152.26
Novartis AG	21	3,983.13
Amgen Inc.	9	3,692.14

Source: InnoThink Center For Research In Biomedical Innovation; Thomson Reuters Fundamentals via FactSet Research Systems (adapted from [16])

Looking at the quality of healthcare and its costs between countries gives interesting insights into the state of global healthcare. The US spends $8,233 per person/year [17] which is 2.5-times more than most of developed nations and uses 17.6% of GDP for healthcare [17]. At the same time, the US had 2.4 practicing physicians per 1,000 people comparing to an average of 3.1 among OECD countries. In hospital beds per 1000 people, the US ranges with 2.6 well under the OECD average of 3.4.

Life expectancy in the US was increased by 9 years between 1960 and 2010; Japan's by 15 years and in OECD countries on average by 11 years [17]. In the drug development arena, per patient clinical trial costs have risen on average by 70 percent across all development phases since 2008 [18].

This numbers are clear indicators that the cost vs. outcome ratio needs to be improved [19] and the current models require adjustments

2.4 Data Ownership: Closed Data vs. Patient-Shared Access

Social media has arrived in healthcare. Patients are sharing publicly their own data, in which case no restrictions on scientific use apply. While many impediments by HIPAA compliance requirements to provide only selected, de-identified subsets to certain authorized individuals have been circumvented by such developments, new questions arise on the consequences from changes in data ownership and the shifts from hospital and providers to patient, and how this may impact integrated research.

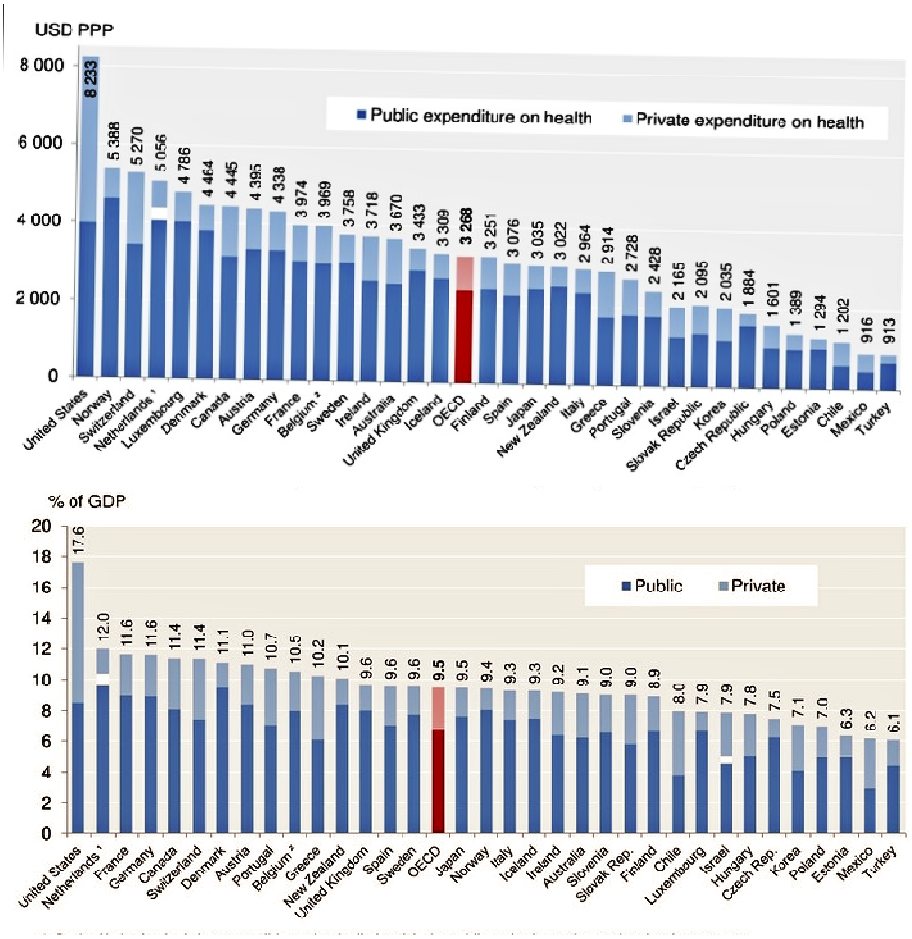

Fig. 1. OECD Statistics 2012: Healthcare cost comparisons across countries (Source: OECD Health Data 2012 [2])

Private and public expenditures per capita (upper panel) and in percentage of the GDP (lower pane)

An article in 2011 in the Harvard Journal of Law and Technology opened the discussion about privacy concerns and why data ownership alone cannot resolve data access problems [20]. Nevertheless, with more and more internet connected personal health devices in use and a strong movement towards prevention and wellness, the aspect of data ownership on real-time, near-continuous monitoring of vital functions and their sharing among private individuals and physicians is a good indicator that closed data will lose ground against patient-shared access in the future.

3 Methodology for Change

3.1 Semantic Approach to Data Integration – Meaning, Inference, Reasoning

Resource description framework (RDF)-based integration (W3C standard) [21] opens new avenue and possibilities for rapid and efficient data integration. It has been around for quite some time, and its benefits as an agile, extensible environment built for interoperability have been widely demonstrated. Semantic data is much easier to connect, to visualize and extend as it does not rely on isolated RDBMS schemas, a contested standard data description, or a proprietary middleware translation layer. Dynamically built application ontologies can be adjusted and data remapped as needed; meaningful, not arbitrary schema-based connections drive its framework of triples, providing capabilities for inference and reasoning and pattern-based queries across the network graphs. The built-in basis for interoperability of true 5-star compliant RDF resources is not only a needed convenience, but a must for today's life sciences needs to utilize a fast array of publicly available linked data resources. RDF and its web ontology language, OWL [22] providing an excellent way to represent data with changing needs, with ability to reuse, repurpose in an easy to adopt and maintain fashion – allowing for disambiguous queries, pattern discovery and graph traversal across multiple RDF-represented resources.

In addition to being a globally standardized framework which links data based on their meaning, emergent properties include network visualization, visual query, faceted browsing, machine inference, and pattern recognition. Recent advances in provenance and versioning [23-24], in the development of public formal ontologies [25] and their direct accessibility through tools [26] have shown increasing interest in life sciences as foundation for larger project. Examples of such ongoing efforts are the development of the Translational Medical Ontology and Knowledgebase [27] driven by both, industry and academia, and the connex between medical informatics and bioinformatics in knowledge building in the clinic [28-29].

3.2 Linked Life Data, Linked Open Data – Consequences

The significant increase in the quality of Linked Data (LLD, LOD) [30-32] brings promising add-ons to qualify experimental findings early on through enrichment with external resources – but interoperability and different provenance remain still impediments for broader applicability as well as changes in licensing for previously 'open' public resources. Legal restrictions on use without modification prevent certain data resources from becoming interoperable as mapping and harmonization functions to other data cannot be applied.

As government funding for some linked open data cloud resources is unsure due to austerity and budget restraints in the US, Japan and Europe we will have to ensure to establish new business models between data provider and consumer to warrant continuous availability of such resources; either through private/academic/government partnerships or new concepts based on resource value for organizations. As the socio-economic benefits of maintaining these resources by far outweigh contributions towards their sustainability, such models will benefit all participants greatly.

3.3 Complexity and Change Require Dynamic, Adaptable Models

New and better scientific methods and analysis tools require adaptation for changes. As outlined before, the complexity of functional biology calls for network analysis of interconnected data in their relationships to each other. This leads logically to semantic integration approaches and network-driven systems to establish better understanding of complex biological intertwined reactions.

The healthcare industry is now about three years into '*meaningful use'*, an ambitious incentive program to convince hospitals and private practices to use electronic health record (EHR) software. Regulators are also bringing HIPAA into the 21st century, and similar efforts are underway for telemedicine. Above all, it looks as if healthcare finally seems ready to benefit from big data, cloud services and other disruptive technologies that have dramatically changed other vertical industries [33]. As healthcare costs have tripled within the last decade and despite over $10 Billion payments in healthcare incentives, we cannot afford to have EHR systems which are not interoperable and CRO's which are disconnected from their customers. A good example about possibilities in progressing with success in complex diseases like atherosclerosis to assess life threatening risk of plaque rupture via biomarkers leading to discovery of previously unknown pathway involvement, using such approaches to take advantage of integrated knowledge can be found in [34].

3.4 Understanding Biology: Shifting towards Interoperable, Integral Systems

The need to contextualize experimental findings with pathway involvement and mechanisms is apparent as pharmacogenomics correlations not necessarily always match biological systems responses. Only when utilizing as much as we possibly can know, we will succeed in comparative effectiveness to select the best treatment at the right dose based on a patient's profile, lifestyle, disease stage and individual drug response.

The shift towards an integral view is the key to improving effectiveness of therapies and better understanding of the impact of a disease stage and a patient's profile on response, prognosis and outcome. Indications are that this is happening now. This year's Health Information and Managements Systems Society's (HIMSS13) conference brought the announcement, that five leading electronic health record (EHR) vendors were forming the '*CommonWell Health Alliance*' to promote 'seamless interoperability' of healthcare data [35].

3.5 Progression towards New Life Sciences Models: Pharma 3.0, Healthcare 3.0

There is a noticeable, albeit slowly, but steadily happening shift in industry from a product-centric business model to a customer/patient centric business model; and the new drivers are health outcomes. Maintaining or regaining growth will require the transition of Pharma from acquisition model to innovate partnerships and collaborative data sharing [36]. Innovation needs to focus more on business model

innovation than product innovation. Reimbursement needs to have its foundation in real market effectiveness rather than the approval of clinical trial data. What applies to Pharma, applies in equal ways to the entire healthcare industry. Efforts to reimbursement based on comparative effectiveness of clinical procedures and treatments indicate that life sciences industries are shifting to evidence-based and outcomes-focused business models – data-savvy, integrative, consumer (patient)-minded rather than product-centric. Moving quickly and following the value to progress towards a new sustainability model will be the key to success. Although there is a sense of urgency to try disruptive methods, it so far has been a 'trying the water' approach around edges of the business, not deeply embraced change. While the trends are apparent to most industry players, to think in new ways has always been uncomfortable and therefore slow in execution. If moving from 2.0 to 3.0 means, that collective impact approaches allow to move more expressively to pre-competitive sharing within the healthcare / life sciences space, the transition will not only be more rapid, it also will create new incentives for holistic approaches to this sector.

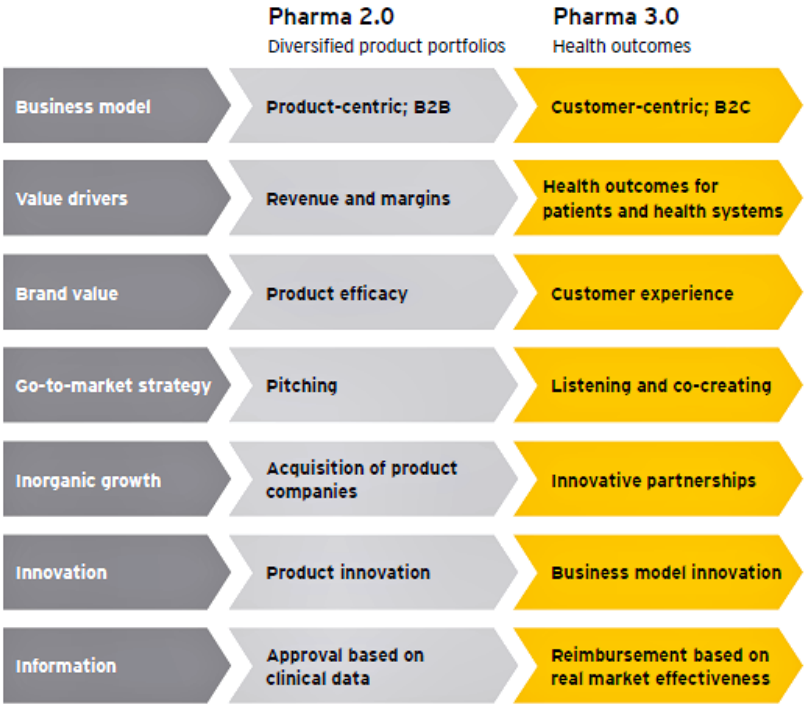

Fig. 2. Pharma on the move from 2.0 to 3.0 – Consequences for Life Sciences(Source: Ernst & Young 2012 [36])

Significant changes need to occur in business model, value drivers and innovation during the shift towards a patient-centric, outcomes-focused and innovation-driven partnership model, where any reimbursement is based on effectiveness in its application rather than the product itself.

4 Use Cases of Adoption

4.1 Pharmaceutical Industry

Impact of Excipient Choice on Formulation Stability, Purity and Drug Efficacy
Objective at a large Pharma company was the integration of disconnected chromatography data systems (CDS) and LIMS with compound and formulation databases to provide quick searches for compound purity data upon FDA inquiries. As there were no common identifiers, such inquiries required time-consuming off-line searches with ambiguous results, delaying responses to the FDA by several weeks. A system was sought to remedy this organizational problem.

The solution to build a semantic platform for compound purity and stability assessment not only was accomplished in a fraction of the allocated time (6 weeks instead of ~ 1 yr. projected project completion time in traditional data warehouse fashion), but inference could be used to quickly and unambiguously find the desired raw data in the CDS. As a pleasant side effect of the semantic data model implemented, an additional data resource was integrated to allow for queries determining the impact of the choice of excipient in a given drug formulation on active ingredient efficacy and overall drug stability.

Pre-clinical Toxicity Assessment and Compound Toxicity Type Classification
In a joint project (NIST/Cogenics/CLDA) to understand the impact of toxicity on biological systems, sets of known and presumed toxicants were used in large 3-year animal studies to determine biomarker classifier patterns and their applicable ranges for pre-clinical toxicity screening of compounds.

Hepatotoxicity studies consisted of a panel of hepatotoxicants at single oral dose (placebo, low, mid, high) in groups of 4 rats, at 6, 24 and 48 hrs.) and metabolic analysis of liver, serum and urine (1603 metabolic components; Bruker LC/MS-MS); gene expression microarray analysis of liver and whole blood (31096 transcript probes; Affymetrix); and statistical biomarker pre-selection at $p<0.005$, abs $fc>10$ (genes) and $p<0.005$, abs $fc>2.5$ (metabolites).

Alcohol studies were carried out at high doses t.i.d. for four days, with and without 24h withdrawal; metabolic analysis of plasma, liver and brain (1620 metabolic components),microarray analysis of liver and brain (31096 transcript probes) and statistical biomarker pre-selection at $p<0.005$, abs $fc>5$ (genes) and $p<0.005$, abs $fc>2.5$ in similar fashion.

The experimental network of statistically preselected putative genomic and metabolomic biomarkers was then enriched with public RDF resources through SPARQL queries to discover common pathway dependencies, using LOD-based systems-biological qualification of experimental pharmacogenomic correlations. As a result of semantic data integration, markers to distinguish several distinct types of

Table 2. Toxicity biomarker with biological validation: Characterization of toxicity types

BM Type	Instance	UniProt AC	Pathway Gene	Protein	Biology
gene	CYP2C40	P11510	cp2cc	Cytochrome P450 2C40	heme binding, iron ion binding, aromatase activity
gene	AKR7A3	P38918	akr7a3	Aflatoxin B1 aldehyde reductase member 3	detoxification
gene	GPX2	P83645	gpx2	Glutathione peroxidase 2	response to oxidative stress, negative regulation of inflammatory response
gene	MYC	P09418	myc	Myc proto-oncogene protein (Transcription factor p64)	regulation of gene transcription, non-specific DNA binding, activates transcription of growth-related genes
gene	MT1A	P02803, Q91ZP8	mt1a	Metallothionein-1	metal ion binding
gene	HMOX1	P06762	hmox1	Heme oxigenase 1	heme catabolic process, negative regulation of DNA binding
gene	FGF21	Q8VI80	fgf21	Fibroblast growth factor 21 (Protein Fgf21)	positive regulation of ERK1 and ERK2 cascade, MAPKKK cascade and cell proliferation
gene	AKR1B8	Q91W30	akr1b8	Aldose reductase-like protein	oxidoreductase activity
gene	TRIB3	Q9WTQ6	trib3	Tribbles homolog 3	disrupts insulin signaling by binding directly to Akt kinases, expression induced during programmed cell death
gene	YC2	P46418	gsta5	Glutathione S-transferase alpha-5 (EC 2.5.1.18)	response to drug, xenobiotic catabolic process
gene	ABCB1, RGD:619851	P43245	abcb1	Multidrug resistance protein 1 (EC 3.6.3.44)	response to organic cyclic compound, tumor necrosis factor, arsenic-containing substance or ionizing radiation
gene	RGD:1310891	Q5U2P3	Zfand2a	AN1-type zinc finger protein 2A	zinc ion binding
gene	GSTP1, GSTP2	P04908	gstp1	Glutathione S-transferase P (EC 2.5.1.18)	response to toxin, xenobiotic metabolic process, response to reactive oxygen species, response to ethanol
gene	RGD:708417	Q62789	ugt2b7	UDP-glucuronosyltransferase 2B7 (UDPGT 2B7) (EC 2.4.1.17)	major importance in conjugation and subsequent elimination of toxic xenobiotics and endogenous compounds
gene	GCLC	P19468	gclc	Glutamate--cysteine ligase catalytic subunit (EC 6.3.2.2)	response to oxidative stress
gene	TXNRD1	O89049	txnrd1	Thioredoxin reductase 1, cytoplasmic (EC 1.8.1.9)	benzene-containing compound metabolic process, cell redox homeostasis, response to drug
gene	NQO1	P05982	nqo1	NAD(P)H dehydrogenase [quinone] 1 (EC 1.6.5.2)	response to oxidative stress, response to ethanol, superoxide dismutase activity
gene	DDIT4L	Q8VDE0	ddit4l	DNA damage-inducible transcript 4-like protein	negative regulation of signal transduction, Inhibits cell growth by regulating TOR signaling pathway
metabolite	Pyroglutamic acid	Q9ER34	aco2	Aconitate hydratase, mitochondrial	citrate metabolism, isocitrate metabolism, tricarboxylic acid cycle
metabolite	Choline	Q64057	aldh7a1	Alpha-aminoadipic semialdehyde dehydrogenase (EC 1.2.1.31)	betaine biosynthesis via choline pathway, response to DNA damage stimulus

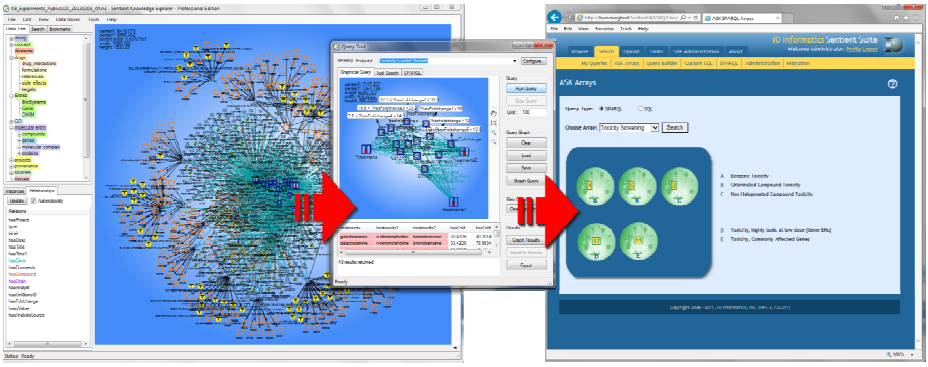

Fig. 3. Use case: Pre-clinical toxicity assessment and categorization

Network view of toxic insult from a set of compounds with affected genes and metabolites in conjunction with their associated pathways and diseases (left); visual SPARQL pattern query with ranges for each biomarker (center); published query pattern as Applied Knowledgebase (ASK) arrays, accessible on a web server for rapid toxicity type screening of compounds (right)

toxicity were established, which passed functional biology criteria for toxicity. As a result, Benzene-, Chlorinated compound- and Ethanol toxicity could be distinguished and long-term effects of Alcohol on memory functions in brain biologically recognized [37-38]. The findings from these studies have major implications for pre-clinical toxicity assessment and were reported at a FDA workshop on pharmacological mechanism-based drug safety assessment and prediction in 2011 [39]. Results for Benzene-type toxicity are depicted in Table 2

Integral Systems View on Drug Safety and Adverse Effects

A multi-national Pharma project required data integration of trial management for drug safety assessment from a large document corpus. The project scope involved ~14,000 documents (~8 gb) with ~1.2 billion spreadsheet cells containing data. Data quality was unknown and data curation requirements were not obvious at start.

During semantic integration, thesaurus-based harmonization and transformation during mapping and data quality enhancement through inference, inconsistencies in the body of data were detected and remedied, As result, the initial semantic knowledgebase contained ~ 780 M triples. Parameterizable SPARQL and a drag-&-drop query builder using a cache search index (iPool) for fast queries provide easy access through a web portal for integral adverse effect queries on compounds used in clinical trials.

4.2 Government

Cross-Species Biomarkers to Reduce Animal Studies

In an effort to reduce animal testing in accordance with world-wide trends against animal experiments, the FDA Center for Veterinarian Medicine (CVM) started a long-term project to develop species-independent biomarkers. The project involved the need to integrate genomics data, proteomics data, imaging endpoints from biopsies, assay results and animal data obtained from a variety of species to proof feasibility to determine common disease biomarkers across those species.

Objective was exploring the ability to move from large animal testing (pigs, dogs) to smaller ones (rats, mice), and further progress to human cell cultures in an effort to reduce costs and minimize the need for animal experiments.

The semantic integration of raw data and results from experimental tests, images, and animal characteristics across multiple time points and the incorporation of public resources (UniProt, KEGG, Reactome) into a comprehensive multi-species biological knowledgebase provides the basis for network analyses to discover cross-species biomarkers applicable to human adverse events and diseases.

Initial results have been very encouraging, and ongoing research integrating additional species data is under way. Ultimately, this development will lead to a significant reduction in animal testing and better drug models for human responses.

Microbial Pathogen Knowledgebase to Identify Biological Threads

Identification of biological threads or the outbreak characteristics of infectious diseases require rapid action on proper identification and characterization of the biological source. While there are several public database resources available (ICTV, MIST, PATRIC), their schemas are not built for interoperability.

As the need to identify microbial pathogens quickly and precisely entails the integral access to as many resources as possible, a semantic mapping for those public resources (including a thesaurus for microorganism for synonym harmonization) was

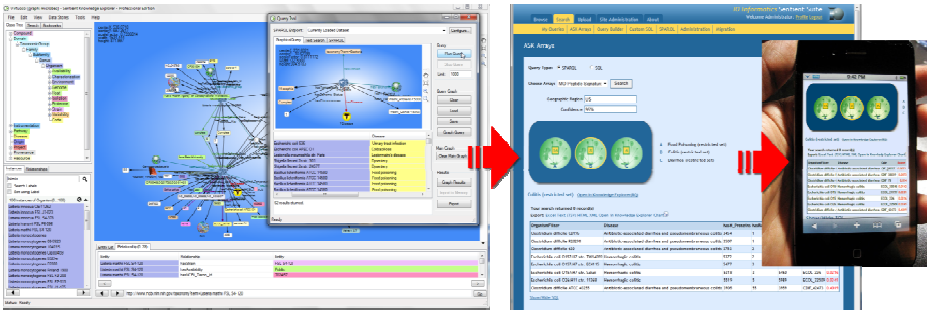

Fig. 4. Use case: Quick pathogen identification from samples using microbial knowledgebase

Visual SPARQL queries across the graph of an integrated microbial pathogen knowledgebase are used to identify threads from experimental data. Those pattern are populating a web server's ASK arrays for online and in-field screening using smart phones

established and the microbial pathogen knowledgebase enhanced with NCBI's taxonomy for group classification into family, subfamily, genus and organisms. This Knowledgebase [40-42] then was used in integrated network analysis of samples from MS-based sequencing and/or rapid microbiological assays in conjunction with historic case data to identify pathogens within the host samples.

An Applied Semantic Knowledgebase (ASK) web portal for simplified pattern queries was established to determine threads by location and confidence level for the identified pathogens.

4.3 Clinical Decision Support

Biomarkers in Transplantation: Organ Rejection Screening

Transplantation is currently the most common therapy for patients with end-stage organ failure. It involves putting a donated organ into the immunologically foreign environment of the receiving patient. While the transplantation procedure itself may go smoothly, the recipient's immune system may react to the new organ to induce rejection. White blood cells and antibodies are primarily involved in the recognition, attack and destruction of foreign tissues, yielding dysfunction of the transplanted organ.

A major challenge facing clinical caregivers in the management of organ rejection is to determine whether a transplanted organ is undergoing rejection prior to any symptoms. This typically required using highly invasive and risky procedures, such as tissue biopsies – expensive, regularly performed emotionally and physically stressful procedures which may still result in inconclusive findings. In order to prevent organ rejection, powerful therapies are used to suppress a patient's immune system. While this approach reduces the probability of rejection, it does so at a high cost. Impairment of a recipient patient's immune system leaves them susceptible to infections, malignancies and functional complications in the newly transplanted organs.

As individuals vary in their response to such therapies, understanding this variation would help physicians balancing the necessity of therapy with its possible side-effects. The ability to personalize immune suppressants for each patient not only alleviates patient discomfort and side-effects, but also reduces the enormous costs associated with over-prescription of immunosuppressive drugs and other diagnostic procedures.

The Biomarkers in Transplantation (BIT) initiative was established to identify and validate biomarkers for diagnosis of rejection of a transplanted organ via a simple blood test [43]. The program was launched in 2004 to better understand acute or chronic tissue rejection in heart, liver, and kidney transplant patients. Its application to use a web-based Applied Knowledgebase (ASK) decision support system won Bio-IT's Best Practices Award in 2010 [44]. It utilizes semantic data integration and parameterized SPARQL queries with weighing and ranges for multimodal biomarkers [45-46] to provide screening for patients at risk

Fig. 5. Use case: Score-based recommendation for immune suppression therapy at risk of organ rejection

Semantic network of a heart transplant patient with blood-test based biomarkers, disease history and organ donor information (left). Use of ASK array to screen for rejection risk and provide guidance-text based decision support on immune suppression therapy (right).

Biomarkers for COPD: Prediction of Exacerbation

Chronic obstructive pulmonary disease (COPD) is a chronic, progressive disease characterized by loss of lung function and breathlessness that reduce quality of life, productivity, and longevity. Its course is frequently complicated by acute exacerbations related to respiratory infections, ambient pollution or poor management of disease. Results are urgent visits to physicians, emergency room care, hospitalization, intensive care unit admissions, and even death. COPD is a major cause of morbidity and mortality around the world. In British Columbia (BC) and the

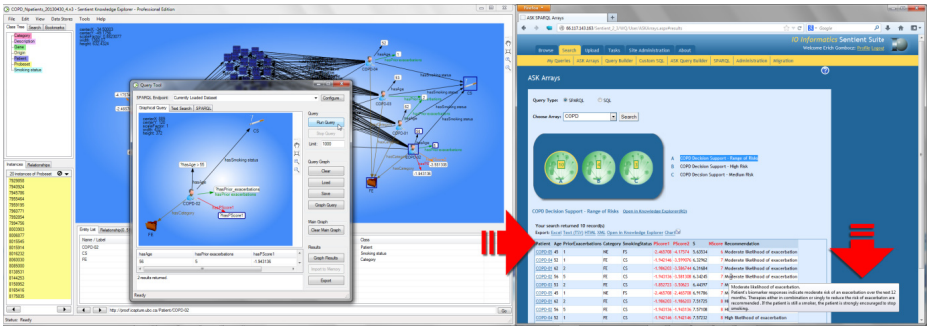

Fig. 6. Use case: Biomarker-based clinical decision support system to predict COPD exacerbations

Semantic integration of clinical data for genomic marker-based decision support predicting likelihood of exacerbation. Network view of 5 patients with different smoking history and prior exacerbations (left). ASK array physician guidance based on normalized algorithmic scoring of weighed biomarker expression profiles (right)

rest of Canada, COPD exacerbations are leading cause of hospitalization, and prevalence has continued to rise, increasing 41% since 1982. It now affects 10-14% of Canadians 40 years of age and older and 600 million people globally [47-48]. As the 4th leading cause of mortality in Canada and the US [49] and the only major cause of mortality for which death rates continue to rise [48-49], the economic costs of COPD management for society are estimated to be over a billion dollars annually in Canada with $736 million of that directly attributable to exacerbations [51]. Most of the morbidity occurs during exacerbations, and their direct costs are predicted to surpass $1 billion by 2015 [52].

A biomarker-based decision support system to predict likelihood of exacerbation and advise treating physician via web-based access to screen patients alleviates those risks and provides a tremendous improvement for patient care in reducing emergency care and hospitalization.

5 Discussion, Future Outlook

5.1 Applied Knowledge as Cost Saver

A 2012 released OECD Health Data Report [19] provides among others statistics on health expenditures, healthcare utilization, demographic references, healthcare quality indicators, pharmaceutical market and long-term care resources. In 2010, in the US public expenditure on health were 48.2% from the total expenditures compared to 87.7% in the Netherland, 85.1% in Denmark and 83.2% in the UK [19]. Applied knowledge from semantic integration of experimental, clinical and public proteomics, genomics, metabolomics and pathway resources led to the development and qualification of multi-modal biomarker pattern applicable for rejection risk assessment with enormous cost savings.

Taking the examples from 4.3 using biomarker blood test for clinical decision support can be used replacing monthly biopsies for up to a year after transplantation at average costs of $4000.-/each. As of 2007, the average price of a kidney-only transplant was $246,000 in the first year. A single lung transplant totaled $399,000. A heart transplant patient's first-year total medical costs were $658,000. In 2011, in the US 1,760 patients on the heart wait list received heart transplants. This represents a decrease from 2,333 hearts transplanted in 2010 and 2,211 in 2009 (Source: UNOS/OPDN). The Heart and Stroke Foundation in Canada reports that heart disease and stroke costs the Canadian economy more than $20.9 billion every year in physician services, hospital costs, lost wages and decreased productivity (Conference Board of Canada, 2010). In 2010, there were 167 heart transplants in Canada, with 135 patients on the waiting list for organ donors. One can imagine the cost savings and quality of life enhancement for patients obtainable through widespread use of preventive non-invasive screening methods Similarly, the effects of being able to predict COPD exacerbations which cause permanent lung tissue damage, are impressive indicators how far reaching integral patient-centric procedures based on semantic knowledgebases have influenced the socio-economics of healthcare.

Actionable knowledge and near-real time alerting of physicians about patients at risk in life-threatening conditions is a testimonial to the real value of interoperable, agile data in life sciences.

5.2 Socio-economics - Higher Quality of Life

As exemplified by examples provided, the use of semantic data integration technologies and integrated, harmonized network approaches utilizing both, internal experimental, clinical, observational and demographic data and public resources provided as RDF/OWL via SPARQL endpoints to enrich, qualify, validate – even plan additional new experiments – has moved from exploratory projects to mainstream acceptability.

Biomarker-based screening for kidney disease to avoid biweekly dialysis or transplantation, heart organ transplant monitoring with biomarkers instead of costly and unpleasant monthly biopsies, and prediction of exacerbations in COPD are just the beginning of a new era of patient-centric, date-driven improvement in health outcomes where everyone involved in applying Pharma 3.0 and Healthcare 3.0 principles [36] will win back sustainability based on reimbursement of real, not perceived effectiveness at the reward of huge socio-economic benefits and improved prevention, care and quality of life.

5.3 Actions Today and Tomorrow

We can see already today the adaption towards more open-minded strategic approaches to build integrated, interoperable (and open?) life science knowledge system capable of remarkable results at significantly lower costs [53] – but there still remains a lot to do.

We need to do more to promote and proliferate these efforts among wider communities to ensure that the life sciences industries are sustainable, effective and applied to help through early intervention, better prognosis and integrated patient-centric, knowledge-based treatment to improve outcomes, increase life expectancy and the quality of life for all. I would urge you to join me in my assessment, that we cannot afford to wait any longer, and that the phase of hesitation on early adaptation to implement innovate solutions and business processes in our quest for comprehensive, integrative systems approaches to better understand biology is over.

We know, what is necessary to change the model, and we have examples leading the way to a bright future – but knowing is not enough; it's time to act, and more than any time before, the time is now.

"*Knowing is not enough; we must apply. Willing is not enough; we must do.*"
- Johann Wolfgang von Goethe (1782)

Acknowledgements. The following groups and researchers have been contributing to the success of the projects described in 4. Use Cases of Adoption:

Toxicity Project: Pat Hurban, Alan J. Higgins, Imran Shah, Hongkang Mei, Ed K. Lobenhofer (Cogenics, Morrisville, NC), Fulton T. Crews (Bowles Center for Alcohol Studies / UNC, Chapel Hill, NC)
Microbial Pathogen Project: Sherry Ayers (FDA NARMS, Silver Spring, MD)
Species-independent Biomarkers: Haile F. Yancy, Michael J. Myers, Rudell Screven (FDA VET / CVM, Laurel, MD)
Biomakers inTransplantation and COPD: Bruce Mc Manus, Raymond T. Ng, Scott Tebbutt (Centre for the Prevention of Organ Failures / PROOF, Vancouver, BC, Canada)
RDF / OWL Database Resources and Ontologies: Jerven T. Bolleman (Swiss Institute Bioinformatics / SIB / UniProt Consortium, Geneva, Switzerland), Michel Dumontier (Bio2RDF II, Carleton University, Ottawa, Canada), Mark A. Musen, Patricia L. Whetzel (BMIR / NCBO Stanford, CA)
W3C HCLS LLD / Pharmacogenomics SIG: Scott Marshall, Michel Dumontier
IO Informatics: Andrea Splendiani, Jason A. Eshleman, Robert A. Stanley
Working Groups: Best Practices in Data Sharing, Informatics for Personalized Medicine
Grant Support for Toxicity Studies: NIST ATP #70NANB2H3009, NIAAA #HHSN281200510008C

References

1. Technology survey on LIMS and ELN in Life Sciences. Project Share Biotech, University of Nantes, France (2011), http://www.biogenouest.org/sites/default/files/Biogenouest/Fichiers/qualite/lims_eln_sharebiotech_study_report.pdf
2. OECD health systems 2012 (2012), http://www.oecd.org/health/health-systems/oecdhealthdata2012.htm
3. Richter, B.G., Sexton, D.P.: Managing and Analyzing Next-Generation Sequence Data. PLoS Comput. Biol. 5(6), e1000369 (2009), doi:10.1371/journal.pcbi.1000369, http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000369#s5
4. Wearable wireless health sensor for remote bio-monitoring. Gizmag (2010)
5. Sensors facilitate health monitoring. Sensor Magazine (2011), http://www.sensorsmag.com/specialty-markets/medical/sensors-facilitate-health-monitoring-8365
6. 10 wearable health tech devices to watch. Information Week (2012), http://www.informationweek.com/healthcare/mobile-wireless/10-wearable-health-tech-devices-to-watch/240012613
7. Mobile ECG monitor via iPhone. AliveCor (2011), http://www.alivecor.com/?gclid=CPP9yqyL-LUCFchaMgodg3QANA
8. fitbit one – Wireless activity monitor. Fitbit (2011), http://www.fitbit.com/one/gallery
9. Sanjeev, A., Boaz, B.: Computational Complexity – A Modern Approach, Cambridge (2009) ISBN 978-0-521-42426-4
10. Zax, D.: MIT Technology Review, November 2011: Is Personal Data the New Currency? (2011), http://www.technologyreview.com/view/426235/is-personal-data-the-new-currency/

11. Ohlhorst, F.J.: Big Data Analytics: Turning Big Data into Big Money. Wiley (2012) ISBN: 978-1-118-14759-7, http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118147596.html
12. Dean, M.: Crowdsourcing as Healthcare Policy Pros&Cons. The Nation (2013), http://www.thenation.com/blog/174059/crowdsourcing-health-care-policy-cant-we-do-better
13. Engelen, L.: Crowdsource your health. Video on TED.com (2012), http://www.ted.com/talks/lucien_engelen_crowdsource_your_health.html
14. Davis, K.: The $1,000 Genome. The Revolution in DNA Sequencing and the New Era of Personalized Medicine. Free Press (2010) ISBN 9781416569596
15. Mardis, E.: The $1,000 genome, the $100,000 analysis? Genome Medicine 2, 84 (2010)
16. Herper, M.: The truly staggering costs of inventing new drugs. Forbes (2012), http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/
17. PBS, 2012: Healthcare costs: How the US compares with other countries (2012), http://www.pbs.org/newshour/rundown/2012/10/health-costs-how-the-us-compares-with-other-countries.html
18. Evangelista, E.: Per patient clinical trial costs rise 70% in three years. Marketwire (2011), http://www.marketwire.com/press-release/per-patient-clinical-trial-costs-rise-70-in-three-years-1538269.htm
19. OECD Health Statistics: Data Dissemination and Results (2012), http://www.oecd.org/els/health-systems/Item%202_OECD%20Health%20Data_Dissemination%20and%20Results_MCC_Final.pdf
20. Harvard Journal of Law and Technology: 25(1) 2011: Much Ado about Data Ownership (2011), http://jolt.law.harvard.edu/articles/pdf/v25/25HarvJLTech69.pdf
21. Resource Description Framework (RDF): Concepts and Abstract Syntax (2004) W3C Recommendation (February 10, 2004), http://www.w3.org/TR/rdf-concepts/
22. OWL 2 Web Ontology Language - Document Overview (2nd edn.) (2012) W3C Recommendation (December 11, 2012), http://www.w3.org/TR/owl2-overview/
23. Cyganiak, R., Zhao, J., Alexander, K., Hausenblas, M.: VoID Vocabulary of Interlinked Datasets. DERI, W3C note (2011)
24. PROV-O: The PROV Ontology. W3C Candidate Recommendation (2012)
25. Musen, M.A., Noy, N.F., Shah, N.H., Whetzel, P.L., Chute, C.G., Story, M.A., Smith, B.: The National Center for Biomedical Ontology. J. Am. Med. Inform. Assoc. 19(2), 190–195 (2012)
26. Salvadores, M., Horridge, M., Alexander, P.R., Fergerson, R.W., Musen, M.A., Noy, N.F.: Using SPARQL to Query BioPortal Ontologies and Metadata. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012, Part II. LNCS, vol. 7650, pp. 180–195. Springer, Heidelberg (2012)
27. Luciano, J.S., Andersson, B., Batchelor, C., Bodenreider, O., Clark, T., Denney, C.K., Domarew, C., Gambet, T., Harland, L., Jentzsch, A., Kashyap, V., Kos, P., Kozlovsky, J., Lebo, T., Marshall, S.M., McCusker, J.P., McGuinness, D.L., Ogbuji, C., Pichler, E., Powers, R.L., Prud'hommeaux, E., Samwald, M., Schriml, L., Tonellato, P.J., Whetzel, P.L., Zhao, J., Stephens, S., Dumontier, M.: The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. J. Biomed. Semantics 2(suppl. 2), S1 (2011)

28. Connecting medical informatics and bioinformatics: Advances in knowledge acquisition and management. Mayo Clinic's Enterprise Data Trust. JAMIA (2010), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000789/
29. Podgorelec, V., Grašič, B., Pavlič, L.: Medical diagnostic process optimization through the semantic integration of data resources: New opportunities for automated reasoning at patient-data level: Medical diagnostic process optimization through the semantic integration of data resources. Computer Methods and Programs in Biomedicine (2009), http://www.sciencedirect.com/science/article/pii/S0169260709000832
30. Linked Life Data http://linkedlifedata.com/sources LLD: Instance mapping, Predicate mapping; 1.1: no of statements: 8,740,201,002;EU IST project
31. Linked Open Data http://linkeddata.org/ LOD Cloud Diagram: http://lod-cloud.net/ LDOW2012 Linked Data on the Web. Bizer, C., Heath, T., Berners-Lee, T., Hausenblas, M.: WWW Workshop on Linked Data on the Web, Lyon, France (April 16, 2012)
32. Callahan, A., Cruz-Toledo, J., Ansell, P., Klassen, D., Tumarello, G., Dumontier, M.: Improved dataset coverage and interoperability with Bio2RDF Release 2. In: SWAT4LS Workshop, Paris, France, November 30 (2012)
33. 13 healthcare IT trends in 2013, http://www.cio.com/slideshow/detail/83055
34. Plasterer, T.N., Stanley, R., Gombocz, E.: Correlation Network Analysis and Knowledge Integration. In: Dehmer, M., Emmert-Streib, F., Graber, A., Salvador, A. (eds.) Applied Statistics for Network Biology: Methods in Systems Biology. Wiley-VCH, Weinheim (2011) ISBN: 978-3-527-32750-8
35. Is healthcare IT interoperability (almost) here? HIMSS13, http://www.cio.com/article/731757/Is_Healthcare_IT_Interoperability_Almost_Here_
36. Ernst & Young: The third place: healthcare everywhere – Global Life Sciences Report (2012), http://www.ey.com/Publication/vwLUAssets/Progressions_The_third_place:_health_care_everywhere_-_Global_Life_Sciences_Report_2012/$FILE/Progressions_Global_Life_Sciences_Report_2012_The_third_place_health_care_everywhere.PDF
37. Gombocz, E.A., Higgins, A.J., Hurban, P., Lobenhofer, E.K., Crews, F.T., Stanley, R.A., Rockey, C., Nishimura, T.: Does network analysis of integrated data help understanding how alcohol affects biological functions? - Results of a semantic approach to biomarker discovery. 2008 Biomarker Discovery Summit 2008, Philadelphia, PA (2008)
38. Higgins, A.J., Gombocz, E.A., Stanley, R.A.: From Correlation to Biological Understanding: Multi-modal semantic networks for biomarker discovery and qualification. In: Bio-IT World 2008, Boston, MA (2008), http://www.io-informatics.com/news/pdfs/CHI_BioIT2008_Talk.pdf
39. Gombocz, E., Stanley, R.: Predictive Toxicology: Applied Semantics with a major impact on drug safety. In: FDA Drug Safety Workshop: Pharmacological Mechanism-Based Drug Safety Assessment and Prediction FDA, White Oak Campus, Silver Spring, MD (2011)
40. Gombocz, E., Candlin, J.: A Novel Approach to Recognize Peptide Functions in Microorganisms: Establishing Systems Biology-based Relationship Networks to Better Understand Disease Causes and Prevention. In: 8th Annual Conference US Human Proteome Organization: The Future of Proteomics (HUPO 2012), San Francisco, CA (2012)

41. Gombocz, E., Candlin, J.: How semantic technology helps fighting infectious diseases: Biological systems approach to understand microbial pathogens. In: Semantic Technology & Businesss Conference (SemTech 2012), San Francisco (2012), http://www.io-informatics.com/news/pdfs/SemTech2012_EGombocz_Talk20120606_r3.pdf
42. Gombocz, E., Candlin, J., Stanley, R., Chiang, D.: Semantically Enhancing Protein Identification: Systems Biology Knowledgebase for Infectious Disease Screening. In: Bio-IT World 2012, Boston, MA (2012), http://www.io-informatics.com/news/pdfs/BioIT2012_Poster.pdf
43. A new way to predict and diagnose organ rejection. Transplant Foundation Research of British Columbia (2010), http://www.trfbc.org/site/PageServer?pagename=News_Biomarkers
44. Bio-IT World Best Practices Award 2010 in Personalized & Translational Medicine: PROOF / iCAPTURE Centre of Excellence; Semantic Data Integration, Knowledge Building and Sharing Applied to Biomarker Discovery and Patient Screening for Pre-symptomatic Heart, Lung or Kidney Failure in Transplantation Medicine (2010), http://www.prweb.com/releases/2010/04/prweb3917414.htm
45. Ng, R.T., Gombocz, E.: Biomarker Development to Improve Decision Support for the Treatment of Organ Failures: How Far Are We Today? In: ADAPT 2010, Arlington, VA (2010), http://www.io-informatics.com/news/pdfs/ADAPT2010_Talk.pdf
46. Stanley, R., McManus, B., Ng, R., Gombocz, E., Eshleman, J., Rockey, C.: W3C Semantic Web Use Cases and Case Studies Case Study: Applied Semantic Knowledgebase for Detection of Patients at Risk of Organ Failure through Immune Rejection. Joint Case Study of IO Informatics and University British Columbia (UBC), NCE CECR PROOF Centre of Excellence, James Hogg iCAPTURE Centre, Vancouver, BC, Canada (2011)
47. Global Initiative for Chronic Obstructive Lung Disease, Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (2009) (updated), http://www.goldcopd.com/Guidelineitem.asp?l1=2&l2=1&intId=2003
48. Murray, C.J., Lopez, A.D.: Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet. 349(9064), 1498–1504 (1997)
49. Jemal, A., Ward, E., Hao, Y., Thun, M.: Trends in the leading causes of death in the United States, 1970-2002. JAMA 294(10), 1255–1259 (2005)
50. Buist, A.S., McBurnie, M.A., Vollmer, W.M.: International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. Lancet. 370(9589), 741–750 (2007)
51. Mathers, C.D., Loncar, D.: Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3(11), e442 (2006)
52. Mittmann, N., Kuramoto, L., Seung, S.J., Haddon, J.M., Bradley-Kennedy, C., Fitzgerald, J.M.: The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. Respir. Med. 102(3), 413–421 (2008)
53. Gombocz, E.A.: On the road to production: Semantic integration cases indicate successful adoption to improve knowledge-based decisions in Pharma and healthcare NCBO Webinar Series, Stanford, CA. Recording (2013)