

Published in final edited form as:

J Intern Med. 2013 October ; 274(4): 321–328. doi:10.1111/joim.12105.

Improving data and knowledge management to better integrate health care and research

M. Cases^{1,*}, L. I. Furlong^{1,*}, J. Albanell², R. B. Altman³, R. Bellazzi⁴, S. Boyer⁵, A. Brand⁶, A. J. Brookes⁷, S. Brunak⁸, T. W. Clark⁹, J. Gea¹⁰, P. Ghazal¹¹, N. Graf¹², R. Guigó^{13,14}, T. E. Klein¹⁵, N. López-Bigas^{1,16}, V. Maojo¹⁷, B. Mons^{18,19}, M. Musen²⁰, J. L. Oliveira²¹, A. Rowe²², P. Ruch²³, A. Shabo²⁴, E. H. Shortliffe^{25,26}, A. Valencia^{14,27}, J. van der Lei²⁸, M. A. Mayer¹, and F. Sanz^{1,14}

¹Research Programme on Biomedical Informatics (GRIB), IMIM, DCEXS, Universitat Pompeu Fabra, Barcelona, Spain ²Servei d'Oncologia, Hospital del Mar, IMIM, DCEXS, Universitat Pompeu Fabra, Barcelona, Spain ³Departments of Bioengineering, Genetics and Medicine, Stanford University, Stanford, CA, USA ⁴Dipartimento di Ingegneria Industriale e dell'Informazione, Università di Pavia, Pavia, Italy ⁵Safety Assessment, AstraZeneca, Mölndal, Sweden ⁶Institute for Public Health Genomics, Maastricht University, Maastricht, The Netherlands ⁷Department of Genetics, University of Leicester, Leicester, UK ⁸Center for Protein Research, University of Copenhagen, Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark ⁹MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Cambridge, MA, USA ¹⁰Servei de Pneumologia, Hospital del Mar – IMIM, DCEXS, Universitat Pompeu Fabra, CIBERES, Barcelona, Spain ¹¹Division of Pathway Medicine, Synthesys, University of Edinburgh, Edinburgh, UK ¹²Department of Pediatric Oncology and Hematology, Saarland University, Homburg, Germany ¹³Centre de Regulació Genòmica (CRG), Barcelona, Spain ¹⁴Spanish Institute of Bioinformatics (INB), Madrid, Spain ¹⁵Department of Genetics, Stanford University, Stanford, CA, USA ¹⁶Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain ¹⁷Biomedical Informatics Group, Universidad Politécnica de Madrid, Madrid, Spain ¹⁸Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands ¹⁹Netherlands Bioinformatics Center, Nijmegen, The Netherlands ²⁰Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA ²¹DETI/IEETA, University of Aveiro, Aveiro, Portugal ²²Janssen Research & Development, Buckinghamshire, UK ²³BITEM, University of Applied Sciences, Geneva, Switzerland ²⁴IBM Research Laboratory, Haifa, Israel ²⁵Arizona State University, Phoenix, AZ, USA ²⁶Department of Biomedical Informatics, Columbia University, New York, NY,

© 2013 The Association for the Publication of the Journal of Internal Medicine

Correspondence: Ferran Sanz, C/ Dr. Aiguader 88, 08003 Barcelona, Spain. (fax: (+34) 933 160 550; ferran.sanz@upf.edu).

*These authors contributed equally to this manuscript.

Conflict of Interest

Anthony Rowe works for Janssen Research and Development, who originally developed the Tran-SMART platform prior to making it open source. Russ B. Altman is the founder of Personalis.com. No other conflict of interest was declared.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Recent initiatives related to data and knowledge management in health care and biomedical research.

USA ²⁷Spanish National Cancer Research Centre (CNIO), Madrid, Spain ²⁸Department of Medical Informatics, Erasmus University Medical Center, Rotterdam The Netherlands

Once upon a time, several engineers, biologists and clinicians realized that a lot of information in biomedicine was partitioned into ‘silos’ that do not intercommunicate. These silos were a side effect of the existence of different disciplines required to, for example, develop new drugs.

The engineers decided to dispose of the silos, and to put the information in axiomatic form to facilitate automatic reasoning over multiple data sources. They also decided to do this in a very open way so that effort was not duplicated. This seemed to be a very reasonable step and was welcomed by all.

After much axiomatization, the engineers found that there were still issues. There was a lack of agreement on many seemingly uncomplicated ‘facts’. They had to employ curators to resolve the issues, and then it was said that the curators were ‘losing the plot’. They also found that there were not only ‘discipline silos’, but also ‘intra-discipline silos’. The ‘intra-discipline silos’ were the partitions between the different evidences and between the assertions developed from the evidences and from earlier assertions, which were based on even earlier assertions, and so on. There were not only webs of disagreement, but also chains of error. And they found that connecting facts from various silos was not so uncomplicated after all, even after axiomatization. Why was this? Because the results of scientific experiments are not axioms, even if they may be treated in this way to perform isolated “bits of tasks” (T. W. Clark) (Fig. 1).

This illustrates the challenge that scientists and clinical practitioners face: the world contains a vast array of complex and diverse data, but locating and connecting the information are difficult [1–3], and deriving definitive knowledge from the data to guide research and/or for clinical practice is even harder. Many barriers that make it difficult to progress in this field were recently discussed at a scientific meeting held in Barcelona from 3 July 2012 to 4 July 2012 [4] under the general title ‘*Beyond Omics Revolutions: Integrative Knowledge Management for Empowered Healthcare and Research*’. The meeting focused on six topics: ‘dealing with biomedical knowledge explosion for better healthcare: identifying actionable knowledge items at the point of care’; ‘exploiting patient information to enrich basic biomedical research’; ‘standards for clinical-omics integration: the semantic challenge’; ‘new information technology (IT) is supporting massive biomedical data management’; ‘systems medicine: making systems biology translational’; and ‘integrative knowledge management for improving drug R&D’. The main ideas and conclusions arising from this event are presented below.

Translating research findings into ‘actionable’ knowledge in the clinical setting

New biomedical discoveries emerge at an ever-increasing rate, but their translation into health care typically occurs slowly or not at all. There is a lack of sufficient systems that can astutely identify, clarify and hand on these advances to the relevant practitioners, in usable formats. For example, thousands of biomarkers exist, comprising a few truly useful ones amongst many others that are less useful or nonactionable. Valuable new biomarkers (diagnostic, prognostic or therapeutic) are therefore not being advanced effectively into health care. The decision of which ones to progress with is simply too onerous, given the cost of modern clinical trials and a deficiency of incentives and expertise amongst researchers who would be best placed to advance markers into development. Hence, when this translation does occur, it is usually because of a major ‘pull’ from the clinical world, rather than a ‘push’ from researchers.

Clearly then, there is a need for methods and systems that can reliably and routinely identify and connect the most informative, reliable and useful information (not least biomarkers) generated by the research community. Efforts to better structure scientific knowledge, for instance, by means of nanopublications [5] or the Investigation Study Assay (ISA) commons [6], could provide key components of this solution. But the challenge is magnified by the fact that the relevant information is spread not only across research resources (e.g. literature, patents, laboratory reports, market data, medical reports and biobanks), but also in realms with less professional rigour such as social networks and patient communities (e.g. wikis, blogs and other social media platforms). Progress will therefore necessitate addressing cross-language and cross-jargon barriers, as well as all the traditional targets of interoperability such as standards for data syntax and semantics.

Beyond connecting and integrating research findings, there lies the challenge of understanding this information. Education is important here, and indeed, it has been proposed that a lack of appropriate training explains the slow uptake of companion diagnostics into clinical practice [7]. Tackling this will require robust guidelines on how to use pharmacogenomic information and also the provision accompanying pharmacokinetic, metabolic and drug interaction knowledge derived from the latest biomedical research. Thus, it is arguable that researchers have a responsibility to make their clinically relevant findings more understandable to the healthcare sector, perhaps in the form of user-friendly web portals or other software [8, 9]. Electronic health record (EHR) developers, computerized physician order entry designers and clinical decision support system (CDSS) creators and vendors likewise need to be involved in delivering additional content for such portals and in connecting such platforms to the intended end users.

Considering all the above issues, as well as the key challenges of data interpretation, some experts concluded that the overall challenge is one of ‘knowledge engineering’ (KE), rather than simply a need for better informatics, research or medical practice. Hence, it may be difficult to make real progress with biomedical researchers and clinical practitioners alone; there is a need for a new group of multidisciplinary engineers [10]. This goes back to the tradition of KE for health, a field that stemmed from artificial intelligence research in the

1990s [11]. However, in contrast to previous KE approaches that aimed to organize all the data to reveal absolute knowledge (which is a flawed concept, as illustrated above), there is a need for a far more pragmatic approach – ‘KE 2.0’ – to identify and make directly useful the very limited set of data and knowledge items that are both reliably proven and clinically actionable. The aim would be to explicitly address the two core information problems faced by clinicians: (i) having too much existing and new data and (ii) not having time or resources to discern reliably from uncertain and erroneous information.

As shown in Supplementary Table S1, there are now many international projects that aim to integrate various types of data related to specific diseases or their pharmacological treatments. In general, however, these are not using the KE 2.0 approach, but developing new methodologies and tools for data integration and exploitation or novel strategies for massive data storage and handling. But as these types of projects make progress in consolidating and unifying the relevant data, KE 2.0 approaches can begin to be explored. However, for this to succeed, the data must be of suitable quality and breadth.

Data quantity and quality

Petabytes of potentially useful biomedical data are not captured in a structured format and/or made available for use by others in many situations. These include molecular ‘omics’ profiles (genomes, transcriptomes, proteomes, epigenomes, etc.), exposure to environmental chemicals exposomes, phenotype data (e.g. as recorded in clinical settings) and dynamic data (e.g. measurements at different points in time or space), all of which could contribute to improved research and health care. For instance, in the research world, primary data from high-throughput studies on a large number of subjects (e.g. genome-wide association studies) [12] typically never leave the laboratory in which they were generated; in the healthcare world, molecular profiles of individual patient, sometimes recorded per time period, are starting to be recorded but are then poorly exploited [13]. It is clear that simply handling this diversity and scale of data is a challenge in itself, but that should motivate focusing more effort on the problem, rather than providing a reason for allowing the data to be lost.

Many considerations relate to the quality, completeness, reliability and reproducibility of primary data and the knowledge derived from them. Relevant judgements may well be context dependent, for example, whether a biopsy from a heterogeneous tumour might be considered usefully representative of the whole tumour. Contextual metadata (data about the data) are therefore important, but such information is often not properly collected or recorded. This is directly related to current discussions about the reproducibility of research findings and the comparability of different analytical procedures. Approaches that allow consistent and repeated analysis of data sets are becoming necessary (e.g. Galaxy and GenePatterns). Questions about reproducibility concern both the data (how they were produced) and the knowledge gleaned from the data (how it was derived). Important studies of statistical and experimental design problems in contemporary scientific publications were recently reported [14, 15].

One notable problem in applying KE to biomedical research data is the nature of the knowledge being engineered. Specifically, active as opposed to consolidated scientific knowledge consists of assertions supported by evidence. What we consider knowledge is a snapshot of the consensus of the scientific community on a particular subject at a given time, but this active knowledge is subjected to continuous re-evaluation where new findings change our perspective, and ‘facts’ may be refuted after some years. Essentially, no knowledge is truly absolute. A particular complication here is that of human bias or error underlying citation distortion, not least in review articles. An example is provided by a recent review in which a role for inclusion body myositis in the aetiopathology of Alzheimer’s disease was suggested. Following the chain of assertions to the underlying evidence, it was found that in some cases, there was no such grounding evidence, and in other cases, its meaning had been distorted or the results misapplied or misconstrued [16]. These issues contribute to the existence of intradiscipline ‘silos’, which disconnect facts and assertions from the underlying evidence. In other cases, there are discrepancies between data collected from different sources. This clearly argues the need for more information accessibility and structure and less reliance on subjective human opinion. But this itself must be balanced against the risk of proposing too many hypotheses from extensive and high-throughput data, which could easily lead to spurious associations.

In this context, ongoing multiparty curation efforts from different initiatives are useful as a way to identify and organize relevant information, but they represent very costly and time-consuming tasks. Efforts on harmonization and standardization, as well as the development of software for supporting curation tasks, are therefore needed to improve and assist curators in their work.

An important point to emphasize is that very different levels of evidence are needed for CDSSs compared with what is required for research grade knowledge discovery. Medical reasoning may be represented by epistemological models, which are amenable to partial automation [17, 18], and in all cases, the data should be generated or chosen to fit a purpose. Researchers, for example, must design their experiments and simulations to record as much detailed information as possible to facilitate a comprehensive exploration of the biomedical question. By contrast, clinicians must carefully define healthcare questionnaires and register only the salient medical variables pertaining to their patients to aid in clinical decision-making. Ideally, however, to avoid silos of data, both groups should always also consider the possible or likely reuse of their data. As part of this, data provenance should be carefully recorded to make possible the retrieval of the original sources and to ensure its reliability and reproducibility, which will undoubtedly have an effect on the generation of useful predictions [19].

Time constraints at the Barcelona meeting precluded extending this discussion into areas of ethical and legal frameworks, but further information can be found elsewhere [20, 21]. It should also be noted that the European Parliament is currently discussing a data protection directive that will underpin a new legal framework [22].

Standards to facilitate translation

Increasingly, genomic information is likely to be relevant to health care, and as such, it should ideally be stored within medical records. An example of current use would be that of personalized drug dosing. Some pharmacogenomic tests are now being used in routine clinical practice; however, they are vastly underused. Key biological data on individuals should be encapsulated in its native format in clinical data structures, with ‘bubbled-up’ items being associated with phenotypic data using clinical data standards. This then generates the question as to what standards are required to allow the efficient translation of key research findings into clinical practice and what IT paradigms will be needed to support biomedical data management. Controlled vocabularies and ontologies for the integration of diverse and heterogeneous biomedical information can provide part of the answer. Fortunately, several current initiatives support the development of ontologies to describe different aspects of biology and biomedicine (e.g. the National Center for Biomedical Ontology [23], the Open Biological and Biomedical Ontologies [24] and the Ricordo project [25]). But yet more needs to be done. For instance, it is difficult to reconcile medical records with disease descriptions associated with public molecular data. This is due to the inherent complexity of diseases and the way they have been traditionally classified and described. Also, disease descriptions are heterogeneous and often dynamic, as in the case of mental illness [26].

Beyond ‘standards’ perhaps, there is actually an equal need for ‘understandings’. In other words, efforts that aim to deliver the standardization capabilities required for KE 2.0, not just standards for semantic integration irrespective of common understanding. We need to make sure that in the next generation of *in cerebro* and *in silico* reasoning strategies, it is understood what is ‘meant’ by any node and edge in a network of associations. To resolve the issue that the more expressive a standard is the less interoperable it is, constraining the standards is crucial, and also enables capturing similarities whilst preserving disparities. More specifically, health data semantics and context cannot be faithfully represented using flat structures (e.g. a list of entries), rather a compositional language is required that meaningfully connects various data entries.

Furthermore, health data standards need to accommodate unstructured data and text (e.g. clinicians’ narrative), whilst having links to structured data entries. A lifetime comprehensive recording of personal health information including omics data is certainly desirable. This arguably calls for a new model of data stewardship: the Independent Health Record Banks (IHRB) vision [27], which would support the implementation of lifelong, cross-institutional and interoperable EHRs. This would constitute an escape from fixation with legacy systems’. As long as healthcare providers are also record keepers, we will continue to have poor archives, proprietary based and isolated in silos, with most of the data semantics not represented explicitly, making it hard or impossible for CDSSs to be really effective. Instead, it is proposed that there should be a limited number of independent and regulated third parties specialized in sustaining the individual lifetime EHR, continuously curating the record and running various analyses to prepare the right info-structure for CDSSs. These tasks require unique specialization and a new kind of archive, which should

provide the most complete and coherent information framework to support the health of the individual.

Fostering literacy in health information management

The challenge of improving biomedical knowledge management goes hand in hand with the need for suitable education and training for all the relevant stakeholders: patients, clinicians, researchers, regulators and policymakers. In particular, clinicians need more support to improve their ability to interpret and use research findings, and researchers must learn how to take actionable findings closer to clinicians. Concomitantly, researchers need to better comprehend the problems raised in clinical practice that can be solved in the laboratory or by intensive use of IT. This reinforces the need for forums of interaction with the active participation of biomedical researchers, bioinformaticians and physicians with experience in clinical research. Hence, we should move from a one-size-fits-all education to that of stratified medicine and from this towards a truly individualized clinical exercise, following the paradigm shift towards the concept of predictive, preventive, personalized and participatory medicine (P4) [28]. Finally, the active participation of citizens, via blogs and other social networks, provides a way to improve the general level of health literacy and thereby to empower all individuals regarding their role in the healthcare system.

Outcomes of the debates

The experts who took part in the aforementioned debates in Barcelona also offer the following consensus statements:

1. There is an urgent need to promote communication and collaboration between experts from different disciplines in order to overcome current information silos and to set up integrated knowledge frameworks required for better managing health problems. In this regard, patients' voices also have to be considered.
2. The current rate of growth of data exceeds that of computational power (e.g. throughput of sequencing instruments will grow faster than the capacity of computers, and this can become a limitation for the spread of next-generation sequencing data use in medical practice). It should be considered malpractice to fund data generation without an adequate data exploitation and stewardship plan. Research funding must seriously consider the need for data storage and analysis, which may be comparable to the effort needed for data generation. When data are generated on human subjects, the stewardship of those data might be handled within each subject's EHR, if a cross-institutional and lifelong record is available.
3. Efforts should be made to improve the methodological and technological background to allow the integrative analysis of complex information (KE 2.0), with the aim of clarifying and delivering clinically actionable information and supporting computational predictions to facilitate the prevention and treatment of diseases.
4. Maximizing data sharing should be an imperative. Not all healthcare data need to be protected under a controlled access regime and not all research needs to be in

open access. Most current barriers for data sharing and reuse are not technical but social. In this respect, we acknowledge novel initiatives (e.g. altmetrics [29]) that seek to go beyond the classical narrative as the only source of scientific knowledge to be taken into account.

5. It is important to address language and jargon barriers to connect the worlds of traditional scientific reporting (peer-reviewed articles) and web sources (patients' blogs, twitter) as sources for knowledge discovery.
6. To facilitate the effective reuse of information, elements of provenance and context along with basic assertions have to be captured from text, databases and EHR systems.
7. The current classification of diseases is largely based on signs and symptoms and in general does not take into account current and evolving knowledge of the molecular pathways that lead to any particular illness. A disease classification based on the molecular biology or the genomics of the diseases would help in the identification of relevant therapeutic interventions.
8. Proper guidelines are needed to help clinicians understand how the results of available genetic tests should be used to optimize patient care, rather than whether tests should be ordered. Here, researchers have a role in preparing these guidelines. Disease- and/or domain-specific 'knowledge portals' could provide a key part of the overall solution, facilitating and driving analysis of data, regulating and tracking data access and providing an optimal balance and scale in terms of the centralization-federation challenge.
9. Citizens (including health professionals) must be enabled, individually and cooperatively, to access, understand, appraise and apply information that will facilitate the use of genome-based information for the benefit of individuals and their communities. In addition, citizens and patients must be consulted with regard to 'donating their data'.
10. All clinical and research data related to an individual's health should be stored in, or linked to, a single lifelong personal (electronic) health record, which would overcome current institutional borders. The development of IHRB may be a way of implementing this vision.

In summary, medicine is an increasingly data-intensive discipline, with a growing need to link individual patient health records to rapidly changing research knowledge for better differential diagnosis, prognosis and prediction of treatment response. Equally, biomedical research will gain enormously from the integrative analysis of clinical and multi-omics information (Fig. 2). Capitalization on these opportunities must be guided by a precise understanding of the many complex issues related to the integration of large amounts of diverse information. Partly this involves overcoming barriers between different disciplines, such as biology, medicine and computer sciences, implying a key role for 'knowledge engineers'.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This review is based on the debates held in Barcelona from 3 July 2012 to 4 July 2012 with the active participation of all authors. The debates were organized by B-Debate (an initiative of Biocat and Obra Social 'La Caixa') and Universitat Pompeu Fabra (Barcelona). The event was held within the framework of the European INBIO-MEDvision project (funded by the EU Seventh Framework Programme for Research and Technological Development (FP7) under grant agreement no. 270107). In addition, we received support from EU FP7 project no. 200754 (GEN2PHEN) and the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115002 (eTOX) and no. 115191 (Open PHACTS), resources of which are composed of financial contribution from the EU FP7 and in kind contributions from companies of the European Federation of Pharmaceutical Industries and Associations. L.I.F received support from Instituto de Salud Carlos III Fondo Europeo de Desarrollo Regional (CP10/00524).

References

1. Ruttenberg A, Clark T, Bug W, et al. Advancing translational research with the Semantic Web. *BMC Bioinformatics*. 2007; 8(Suppl 3):S2. [PubMed: 17493285]
2. Antezana E, Kuiper M, Mironov V. Biological knowledge management: the emerging role of the Semantic Web technologies. *Brief Bioinform*. 2009; 10:392–407. [PubMed: 19457869]
3. Butte AJ. Translational Bioinformatics: coming of age. *J Am Med Inform Assoc*. 2008; 15:709–714. [PubMed: 18755990]
4. B-Debates. [Accessed July 10, 2013] <http://www.bdebate.org/debat/beyond-omics-revolutions-integrative-knowledge-management>.
5. Mons B, van Haagen H, Chichester C, et al. The value of data. *Nat Genet*. 2011; 43:281–283. [PubMed: 21445068]
6. Sansone SA, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. *Nat Genet*. 2012; 44:121–126. [PubMed: 22281772]
7. Wilffert B, Swen J, Mulder H, Touw D, Maitland-Van der Zee AH, Deneer V. From evidence based medicine to mechanism based medicine. Reviewing the role of pharmacogenetics. *Int J Clin Pharm*. 2011; 33:3–9. [PubMed: 21365387]
8. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol*. 2012; 8:592. [PubMed: 22806140]
9. Preissner S, Kroll K, Dunkel M, et al. Super CYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res*. 2010; 38:D237–D243. [PubMed: 19934256]
10. Beck T, Gollapudi S, Brunak S, et al. Knowledge engineering for health: a new discipline required to bridge the “ICT gap” between research and healthcare. *Hum Mutat*. 2012; 33:797–802. [PubMed: 22392843]
11. Warner, HR.; Sorenson, DK.; Bouhaddou, O. *Knowledge Engineering in Health Informatics*. New York: Springer-Verlag; 1997.
12. Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med*. 2009; 360:1759–1768. [PubMed: 19369657]
13. Chen R, Mias GI, Li-Pook-Than J, et al. Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell*. 2012; 148:1293–1307. [PubMed: 22424236]
14. Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AHS. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes*. 2011; 4:304. [PubMed: 21854631]
15. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets. *Nat Rev Drug Discov*. 2011; 10:712.

16. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*. 2009; 339:b2680. [PubMed: 19622839]
17. Sparkes A, Aubrey W, Byrne E, et al. Towards Robot Scientists for autonomous scientific discovery. *Autom Exp*. 2010; 2:1. [PubMed: 20119518]
18. Riva A, Nuzzo A, Stefanelli M, Bellazzi R. An automated reasoning framework for translational research. *J Biomed Inform*. 2010; 43:419–427. [PubMed: 19931420]
19. Ekins S, Waller CL, Bradley MP, Clark AM, Williams AJ. Four disruptive strategies for removing drug discovery bottlenecks. *Drug Discov Today*. 2013; 18:265–271. [PubMed: 23098820]
20. Hudson KL. Genomics, Health Care, and Society. *N Engl J Med*. 2011; 365:1033–1041. [PubMed: 21916641]
21. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Genet*. 2012; 13:395–405.
22. European Parliament - Data protection. [Accessed July 10, 2013] http://ec.europa.eu/justice/data-protection/index_en.htm.
23. National Center for Biomedical Ontology. [Accessed July 10, 2013] <http://www.bioontology.org/>.
24. Open Biological and Biomedical Ontologies. [Accessed July 10, 2013] <http://www.obofoundry.org/>.
25. Ricordo project. [Accessed July 10, 2013] <http://www.ricordo.eu/>.
26. Tabares-Seisdedos R, Dumont N, Baudot A, et al. No paradox, no progress: inverse cancer comorbidity in people with other complex diseases. *Lancet Oncol*. 2011; 12:604–608. [PubMed: 21498115]
27. Independent Health Record Banks. [Accessed July 10, 2013] <http://www.independenthealthrecordbanks.blogspot.co.il/>.
28. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol*. 2011; 8:184–187. [PubMed: 21364692]
29. Almetrics. [Accessed July 10, 2013] <http://altmetrics.org/manifesto/>.

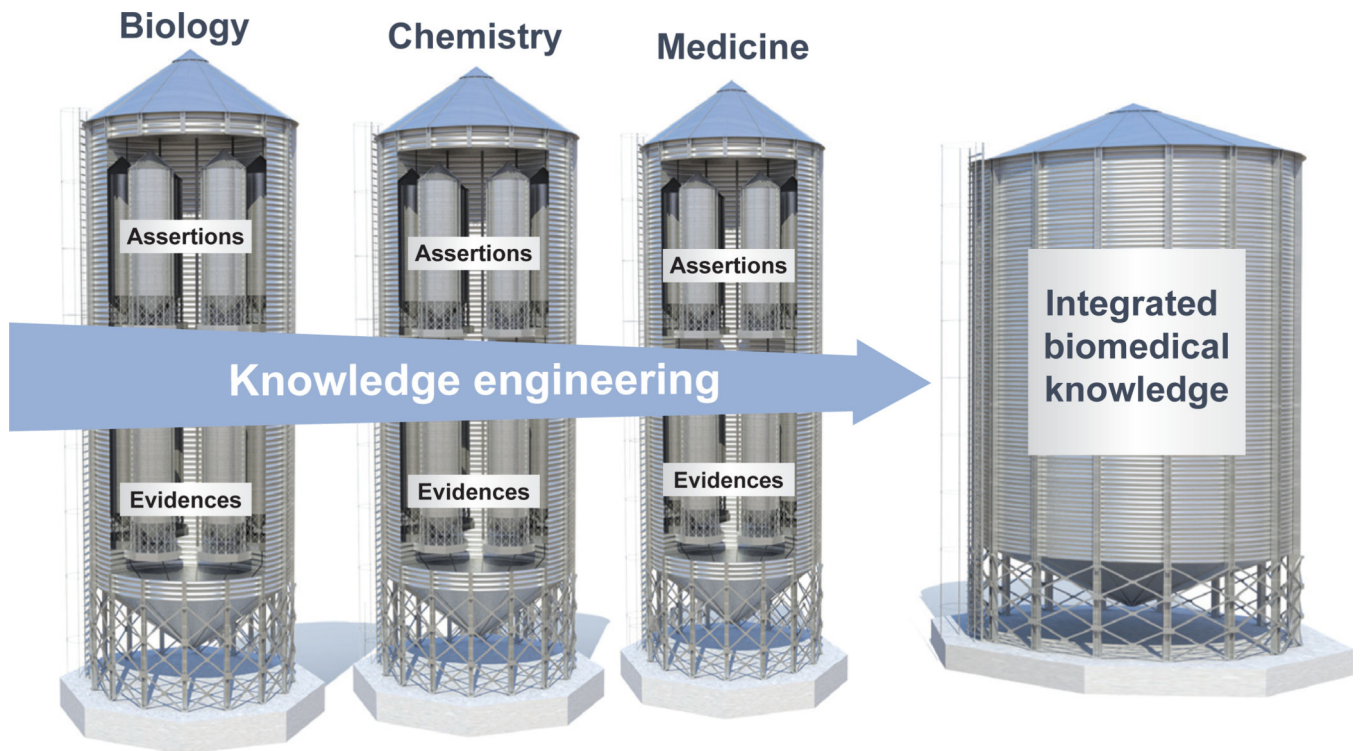


Fig. 1. Interdiscipline and intradiscipline knowledge ‘silos’ to be overcome. The ‘intradiscipline silos’ are the partitions between the different evidences and between the assertions developed from the evidences and from earlier assertions.

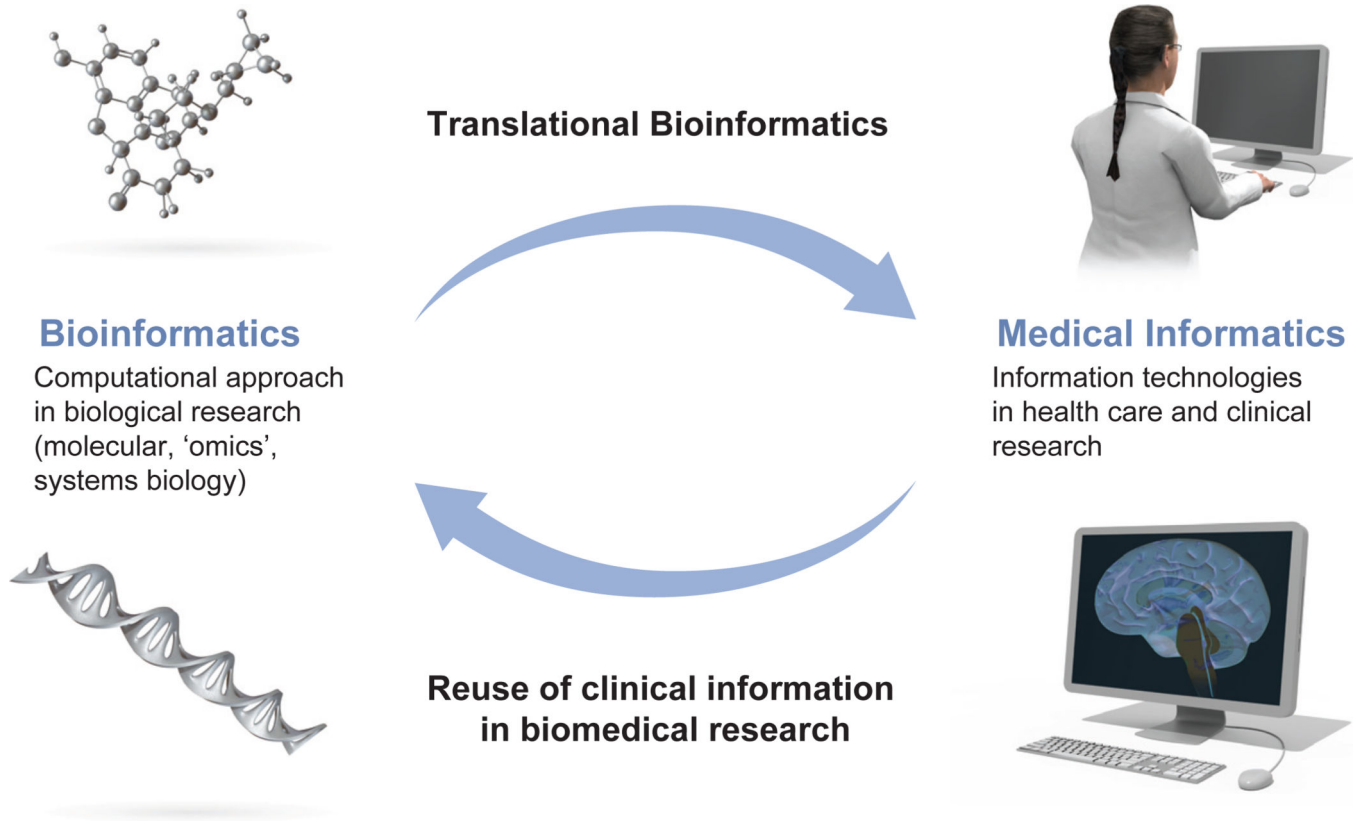


Fig. 2. Translational and integrative approaches in biomedical informatics.